

ORIGINAL ARTICLE

Machine Learning–Enabled Multimodal Fusion of Intra-Atrial and Body Surface Signals in Prediction of Atrial Fibrillation Ablation Outcomes

Siyi Tang¹, MS; Orod Razeghi², PhD; Ridhima Kapoor³, MD; Mahmood I. Alhusseini, MS; Muhammad Fazal, MD; Albert J. Rogers⁴, MD, MBA; Miguel Rodrigo Bort, PhD; Paul Clopton⁵, MS; Paul J. Wang⁶, MD; Daniel L. Rubin⁷, MD; Sanjiv M. Narayan⁸, MD, PhD; Tina Baykaner⁹, MD, MPH

BACKGROUND: Machine learning is a promising approach to personalize atrial fibrillation management strategies for patients after catheter ablation. Prior atrial fibrillation ablation outcome prediction studies applied classical machine learning methods to hand-crafted clinical scores, and none have leveraged intracardiac electrograms or 12-lead surface electrocardiograms for outcome prediction. We hypothesized that (1) machine learning models trained on electrograms or electrocardiogram (ECG) signals can perform better at predicting patient outcomes after atrial fibrillation ablation than existing clinical scores and (2) multimodal fusion of electrogram, ECG, and clinical features can further improve the prediction of patient outcomes.

METHODS: Consecutive patients who underwent catheter ablation between 2015 and 2017 with panoramic left atrial electrogram before ablation and clinical follow-up for at least 1 year following ablation were included. Convolutional neural network and a novel multimodal fusion framework were developed for predicting 1-year atrial fibrillation recurrence after catheter ablation from electrogram, ECG signals, and clinical features. The models were trained and validated using 10-fold cross-validation on patient-level splits.

RESULTS: One hundred fifty-six patients (64.5±10.5 years, 74% male, 42% paroxysmal) were analyzed. Using electrogram signals alone, the convolutional neural network achieved an area under the receiver operating characteristics curve (AUROC) of 0.731, outperforming the existing APPLE scores (AUROC=0.644) and CHA2DS2-VASc scores (AUROC=0.650). Similarly using 12-lead ECG alone, the convolutional neural network achieved an AUROC of 0.767. Combining electrogram, ECG, and clinical features, the fusion model achieved an AUROC of 0.859, outperforming single and dual modality models.

CONCLUSIONS: Deep neural networks trained on electrogram or ECG signals improved the prediction of catheter ablation outcome compared with existing clinical scores, and fusion of electrogram, ECG, and clinical features further improved the prediction. This suggests the promise of using machine learning to help treatment planning for patients after catheter ablation.

GRAPHIC ABSTRACT: A graphic abstract is available for this article.

Key Words: Machine learning ■ atrial fibrillation ■ cardiac electrophysiology ■ catheter ablation

Atrial fibrillation (AF) ablation is the cornerstone of therapy for symptomatic AF, and it helps improve quality of life and prolongs survival in several populations.^{1,2} Improved tools for predicting the success

of AF catheter ablation are needed to guide clinicians in better patient selection for this procedure, as well as setting realistic patient expectations following the procedure.

Correspondence to: Tina Baykaner, MD, MPH, Stanford University, 453 Quarry Rd, 334C, Palo Alto, CA 94304. Email tina4@stanford.edu
The abstract was presented at Heart Rhythm Society (HRS) Scientific Sessions. The abstract embargo for this meeting will be lifted on April 28th, 2022.
This article was sent to N.A. Mark Estes III, MD, Guest Editor, for review by expert referees, editorial decision, and final disposition.
Supplemental Material is available at <https://www.ahajournals.org/doi/suppl/10.1161/CIRCEP.122.010850>.
For Sources of Funding and Disclosures, see page 508.
© 2022 American Heart Association, Inc.

Circulation: Arrhythmia and Electrophysiology is available at www.ahajournals.org/journal/circep

WHAT IS KNOWN?

- Atrial fibrillation ablation is the cornerstone of therapy for symptomatic atrial fibrillation, with increasing evidence on its safety and efficacy.
- Clinical scores have been developed to predict success of catheter ablation, to guide better patient selection, with most clinical scores reaching an area under the receiver operating characteristics curve (AUROC) of 0.55 to 0.65 in accurately predicting atrial fibrillation ablation success.

WHAT THE STUDY ADDS

- Deep neural networks trained on intracardiac signals and 12-lead electrocardiogram signals, in addition to clinical features, can improve the prediction accuracy of catheter ablation outcomes compared with existing clinical scores.
- A convolutional neural network using intracardiac signals in atrial fibrillation achieves an AUROC of 0.731, similarly a convolutional neural network using 12-lead electrocardiogram alone achieves an AUROC of 0.767. Fusion of electrogram, electrocardiogram, and clinical features further improves the prediction (AUROC=0.859) compared with models with a single modality.
- Machine learning models can help treatment planning for patients after catheter ablation of atrial fibrillation through more accurate prediction of treatment outcomes.

Nonstandard Abbreviations and Acronyms

AAD	antiarrhythmic drug
AF	atrial fibrillation
AUROC	area under the receiver operating characteristics curve
CAD	coronary artery disease
CatBoost	categorical boosting classifier
CKD	chronic kidney disease
CNN	convolutional neural network
CT	computerized tomography
LA	left atrium
Lad	left atrial diameter
ML	machine learning
PVI	pulmonary vein isolation

Clinical scores have been developed to predict success after catheter ablation of AF with area under the receiver operating characteristics curve (AUROC) of 0.55 to 0.65 for majority of the models, with rare models reaching an AUROC of 0.75.³⁻⁵ However, none of these previous predictive scores have incorporated electrophysiological data, which may place specific AF mechanisms within the clinical context to improve predictive accuracy.

We hypothesized that (1) machine learning (ML) models trained on intracardiac electrograms or surface electrocardiograms (ECG) signals can perform better at predicting patient outcomes after AF ablation (ie, 1-year AF recurrence) compared with existing clinical scores and (2) multimodal fusion of electrogram, ECG, and clinical features can further improve the prediction of patient outcomes.

Although there are no prior ML-based studies that directly take signals as inputs to predict AF ablation outcomes, recent advances in the use of ML in signal analysis of human rhythm disorders have led to promising preliminary results. For example, ML models were able to predict future ventricular arrhythmia from ventricular signals.⁶ Prior works using ML to predict success of AF ablation includes estimation of recurrence by predicting shape descriptors directly from magnetic resonance imaging⁷ and combining imaging and clinical biomarkers to predict cryoballoon pulmonary vein isolation (PVI) outcomes.⁸ ML methods and personalized computational modeling have also been used together to predict recurrence following PVI.⁹ In addition, handcrafted features derived from computerized tomography (CT) scans have been shown to be associated with likelihood of postablation AF recurrence.¹⁰

Deep neural networks are the state-of-the-art ML models that are able to learn complex features directly from large amounts of data without the need of feature engineering.¹¹ Deep neural networks have shown promising empirical successes across a wide variety of medical domains.¹² Unlike previous works using classical ML models,⁸⁻¹⁰ we aim to develop and validate (1) a deep neural network for post-ablation AF recurrence prediction from signals (electrogram and ECG) and (2) a multimodal fusion framework that leverages the three modalities—electrogram, ECG, and patients' clinical features—to further improve the model performance (Figure 1A).

METHODS

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Subject Recruitment

This is a retrospective analysis of consecutive adult patients with paroxysmal or persistent AF who underwent catheter ablation between 2015 and 2017 at a tertiary referral center by 5 providers. To be included, patients were required to have panoramic left atrial electrograms recorded before ablation and clinical follow-up for at least 12 months following ablation for accurate assessment of their AF ablation procedure outcomes. All patients had pulmonary vein isolation as a part of the AF ablation procedure; additional ablation lesions per the operating physicians' discretion were allowed. This comprised of ablation of localized AF sources via focal impulse and rotor mapping (FIRM, in 100% of patients), ablation of left atrial linear lesions (in 24% of patients) and cavotricuspid isthmus ablation (in 27% of patients).

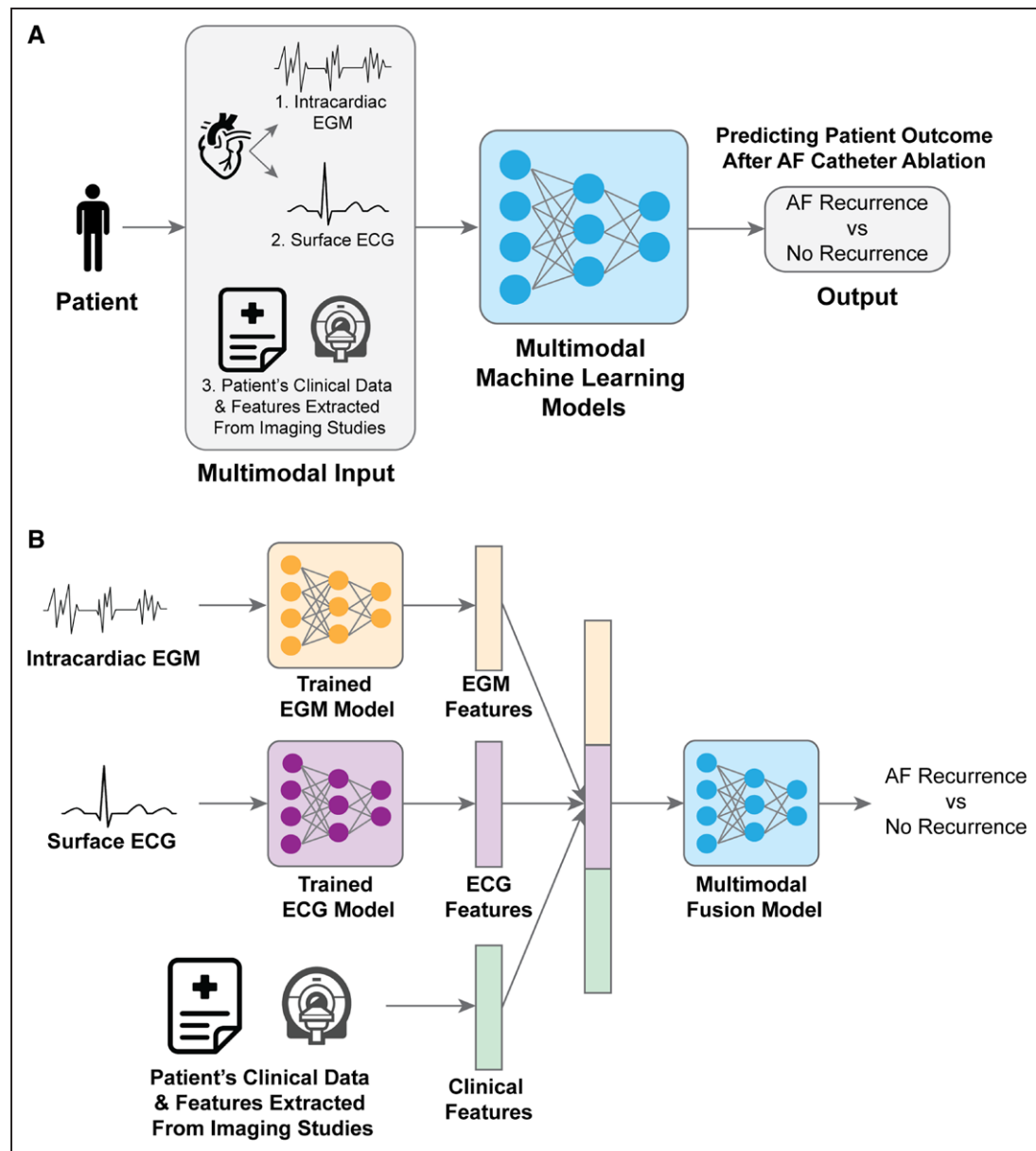


Figure 1. Overview of our methods and multimodal fusion framework.

A, Overview of our methods. The inputs come from 3 modalities: patient electrogram (EGM) signals, electrocardiogram (ECG) signals, and clinical features. A multimodal machine learning model fuses the inputs from the 3 modalities and outputs prediction of atrial fibrillation (AF) recurrence. **B**, Details of our multimodal fusion framework. We first trained a model on EGM signals only for AF recurrence prediction, and a separate model on ECG signals only for AF recurrence prediction. We then extracted EGM and ECG features from the respective trained models. Finally, the EGM and ECG features were concatenated with the clinical features and were subsequently passed to a multimodal fusion model to predict AF recurrence.

Clinical and demographic data were obtained from electronic medical records. Twelve-lead ECGs in sinus rhythm obtained within 1 year of the ablation procedure were included. Patients with no 12-lead ECG available ($n=3$) were excluded from the ECG-only model and were imputed with the means of the other patients' ECG features in the fusion models. This study protocol was approved by the Institutional Review Board of Stanford University. Due to the retrospective nature of the study, no informed consent was required. The corresponding author had full access to all the data in the study and took responsibility for its integrity and the data analysis.

Ablation Procedure and Clinical Follow-Up

All procedures were performed under general anesthesia. Various ablation catheters were used to achieve PVI, which included point-by-point radiofrequency ablation with a contact force sensing 3.5 mm tip irrigated catheter (Biosense Webster; Abbott) or cryoballoon (Arctic Front, Medtronic). Unipolar panoramic intracardiac signals used for ML analysis were obtained before any ablation with a 64-pole basket catheter (FIRMap catheter, Abbott) during AF. If patients presented to the electrophysiology laboratory in normal sinus rhythm, AF was induced with burst pacing.

Patients were followed up routinely in the outpatient setting, and all had 3-month evaluations for at least 1 year, which included rhythm assessment with 12-lead ECGs at 3 and 6 months and a 14-day event monitor at 1 year. AF recurrence was defined as >30 second duration episodes on ECG monitoring, or >1% AF burden on device interrogation for the patients with implantable monitors. In this study, we focus on the outcome of whether a patient has recurrent AF within 1 year after catheter ablation.

Demographic and Clinical Features

The demographic variables extracted from electronic health records included patients' age at the time of ablation, sex, height, weight, body mass index, race, and ethnicity. Clinical comorbidities such as presence of hypertension, hyperlipidemia, transient ischemic attack, stroke (CVA), coronary artery disease, diabetes (DM), chronic kidney disease, congestive heart failure, and obstructive sleep apnea were collected. Arrhythmia characteristics such as type of AF (paroxysmal, persistent or long standing persistent), and history of prior AF ablation were recorded. Structural features extracted from imaging studies included left ventricular ejection fraction and left atrial diameter from transthoracic echocardiograms; and left atrial volume, surface area and sphericity index from CT scans that were routinely obtained within 1 year before AF ablation. These variables were selected based on the literature on known factors which could impact AF ablation outcomes.^{3–5,8,13–16} A complete list of clinical features and number of missing values is shown in [Table S1](#). Missing values were imputed with the most frequent value of the feature. [Table S2](#) shows model performance with different missing value imputation techniques, and [Table S3](#) shows model performance in patients without missing values.

Modeling Clinical Features for AF Recurrence Prediction

As a baseline method, we built a classifier for predicting 1-year AF recurrence from demographic and clinical features. For each patient, a multi-dimensional feature vector was constructed from the clinical and demographic features, where continuous variables were normalized to have zero mean and unit variance and categorical variables were one-hot encoded. We used the categorical boosting (CatBoost) classifier,¹⁷ a state-of-the-art, gradient boosted decision tree-based ML algorithm, for AF recurrence prediction. Briefly, CatBoost sequentially builds many weak learners (ie, decision trees) and creates a strong predictive model by greedy search and ensembling. We chose CatBoost because it has been shown to outperform other gradient boosted decision tree-based algorithms and naturally handles both continuous and categorical variables.¹⁷

Preprocessing of Electrogram and ECG Signals

In each patient, unipolar left atrial intracardiac electrograms were recorded during AF. Unipolar signals were recorded from a 64-pole basket catheter positioned in the mid left atrium (LA) before any ablation were exported. Preprocessing of electrogram signals included QRS subtraction and resampling to 200 Hz. See [Supplemental Methods](#) for details.

Preprocessing of ECG signals included a bandpass filtering of 0.05 to 100 Hz and resampling to 200 Hz. Eight

independent ECG channels were used (channels I, II, and V1–6) as any linear dependency can be naturally learned by deep neural networks (ie, channel III can be derived vectorially from channels I and II).

Each electrogram and ECG signal was augmented by dividing into 5-sec windows with a 4-sec overlap between consecutive windows, resulting in a 1000×64 matrix for each input electrogram data point and a 1000×8 matrix for each input ECG data point.

Modeling Electrogram and ECG Signals for AF Recurrence Prediction

We developed a convolutional neural network (CNN) for predicting 1-year AF recurrence from electrogram or ECG signals.

Similar to Attia et al,¹⁴ our CNN consisted of several layers of bottleneck blocks with 1-dimensional (1D) convolutions operating on the time dimension, followed by a 1D convolutional layer operating on the channel dimension. Intuitively, the time-dimension convolutional layers capture the temporal dependency in the signal by extracting features from signals within one channel, whereas the final channel-dimension convolutional layer aggregates the features across channels to obtain a spatial representation of the signal. Details of the CNN can be found in [Supplemental Methods](#) and [Figure S1](#).

Fusion Model for AF Recurrence Prediction

Finally, we developed a multimodal fusion framework that leverages more than one modality to improve the prediction of AF recurrence ([Figure 1B](#)).

First, electrogram features were extracted from the CNN that was trained on electrogram signals only. All the features from the electrogram signals from the same patient were averaged to obtain a single electrogram feature representation for each patient. ECG features for each patient were extracted in a similar way. Next, for each patient, the feature vectors of the fused modalities (ie, electrogram features, ECG features, and clinical features) were concatenated to form a multimodal feature vector. Last, a classifier was trained on the patients' multimodal feature vectors for predicting 1-year AF recurrence. As a fair comparison to clinical feature-based models, we also applied the CatBoost¹⁷ classifier in the fusion framework.

As ablation experiments, we also validated fusion of 2 modalities (ie, electrogram and clinical features, ECG and clinical features, or electrogram and ECG features) and compared the results to fusion of 3 modalities (electrogram, ECG, and clinical features).

Model Training and Validation

Stratified 10-fold cross-validation (patient-wise split) was used to train and validate each of the models described above. Specifically, all patients were randomly divided into 10 groups (ie, folds) with the same proportion of AF recurrence in each fold (ie, stratified 10-fold). At the *i*-th cross-validation step, the *i*-th fold was used to test the model and the remaining 9 folds were used to train the model. This above process was repeated 10 times, such that each patient only appeared in one of the test folds.

To mitigate overfitting, data augmentation was applied during training. We designed 5 data augmentation methods using electrophysiology domain knowledge: (1) randomly shift (forward

or backward in time) each 5-sec window by up to 2.5-sec, (2) randomly scale the raw signal by a factor within range 0.5 to 2, (3) randomly shift the DC value within range -10 to 10 microvolts, (4) randomly masking with zeros for up to 25% of the 5-sec window, (5) randomly add gaussian noise with zero mean and a $SD < 0.2$. Importantly, these data augmentations did not result in invalid signals but naturally increased the variability of the training data, which could mitigate overfitting of deep neural networks.

Training for the CNNs on electrogram and ECG signals was accomplished using the Adam optimizer¹⁸ in PyTorch on a single NVIDIA P100 GPU. For CNNs, we followed the same model architecture configuration as that in Attia et al¹⁴ (except for reducing the number of bottleneck blocks from 9 to 6 in ECG-based CNN) and did not tune the model hyperparameters. Training for CatBoost was done using the CatBoost Python package,¹⁷ and CatBoost hyperparameters were tuned using grid search (see [Supplemental Methods](#) for details). All models were trained to optimize AUROC. We assessed the model's ability to predict 1-year AF recurrence using AUROC, sensitivity, specificity, accuracy, and F1-scores. To derive sensitivity, specificity, accuracy, and F1-scores, a probability threshold was selected based on the highest F1-score on the 10 fold test sets.

Statistical Analysis

For population characteristics, continuous data are reported as mean \pm SD, unless otherwise stated, and are tested for normality using the Shapiro-Wilk test ($P > 0.05$). Independent samples *t* test and Mann-Whitney *U* test were run to determine if there were differences in mean values between cohorts for analysis of continuous data. Categorical variables were compared using the Pearson χ^2 test or Fisher exact test where expected frequencies were < 5 . For model evaluation, we report the mean and SD of AUROC, sensitivity, specificity, accuracy, and F1-scores of the 10-fold test results. In addition, we measure the calibration of the models using Brier score¹⁹ and expected calibration error (ECE).²⁰ Briefly, the Brier score measures the mean squared difference between the predicted probability assigned to the possible label and the actual label. The ECE approximates the expectation between model confidence and accuracy by binning the predictions into equally-spaced bins and taking a weighted average of the bin's accuracy and confidence difference. For both Brier score and ECE, lower values indicate better calibrated models. A statistical significance threshold (α) of 0.05 was used for all the reported tests.

RESULTS

Overall Summary

Between 2015 and 2017, 226 consecutive AF ablations were done using a 64-pole basket catheter that recorded simultaneous panoramic unipolar electrograms from the left and the right atria. Of these, 161 had left atrial signals recorded before any ablation. Five were excluded due to poor signal quality, leaving 156 patients to be analyzed for this study. Baseline characteristics of these patients are shown in Table 1. PVI was done using radiofrequency in 118 patients (76%), cryoballoon in 38 patients (24%). Thirty-four patients (21.8%) were on an antiarrhythmic

drugs (AADs) at the time of follow up (10.2% on class IC agents, 3.9% on class III agents [sotalol or dofetilide], 8.3% on amiodarone and 1.9% on dronedarone). Additional ablation lesions beyond PVI and ablation of localized sources are presented in Table 1.

Catheter Ablation Outcomes

On follow-up at 1 year, 112 (72%) patients remained free of AF. Patients with and without recurrence had a similar age, body mass index and comorbidities (Table 1). AAD use was not different among groups. 28% of the patients had a prior history of AF ablation. Presence of hyperlipidemia and diabetes correlated with AF recurrence ($P = 0.04$) in univariate analysis. Ablation of additional left atrial lines did not correlate with AF ablation outcomes.

Validation of Existing AF Ablation Outcome Prediction Scores: APPLE and CHA2DS2-VaSC

First, we validated 2 existing clinical feature-based prediction scores, APPLE³ and CHA2DS2-VaSC,⁴ for 1-year AF recurrence prediction using CatBoost.¹⁷ Detailed formulation of APPLE and CHA2DS2-VaSC scores can be found in [Supplemental Methods](#).

The CatBoost classifier achieved an AUROC of 0.644 ($SD = 0.129$) on APPLE scores and an AUROC of 0.650 ($SD = 0.133$) on CHA2DS2-VaSC scores (Table 2, first and second rows).

ML-Based AF Recurrence Prediction From Clinical Features

Using clinical features, the CatBoost classifier achieved an AUROC of 0.755 ($SD = 0.093$; Table 2, third row), outperforming the performance of the CatBoost classifier trained on APPLE and CHA2DS2-VaSC scores. This performance improvement is expected given that multiple clinical features were used, whereas APPLE and CHA2DS2-VaSC scores only accounted for 5 and 7 clinical features, respectively.

Figure 2 shows the model interpretation of the clinical features that contribute the most to AF recurrence prediction in our clinical feature-based model, where the 5 most important features are left ventricular ejection fraction, height, body mass index, weight, left atria volume from CT, and left atria surface area; which have previously been reported to correlate with development of incident AF^{15,21} or poorer outcomes following AF ablation.^{16,22}

ML-Based AF Recurrence Prediction From Electrogram or ECG

Using electrogram signals only, the CNN achieved an AUROC of 0.731 ($SD = 0.105$) for AF recurrence

Table 1. Baseline Characteristics of Population

	All subjects (n=156)	Free from AF (n=112)	Recurrent AF (n=44)	P Value
Demographics				
Age, y (mean±SD)	64.5±10.5	64.5±9.9	64.5±11.9	0.988
Male sex, n (%)	115 (74%)	87 (78%)	28 (64%)	0.073
Height (m, mean±SD)	1.77±0.1	1.77±0.1	1.77±0.1	0.298
Weight (kg, mean±SD)	96.6±24.4	98.1±24.3	92.6±24.4	0.205
BMI (kg/m ² , mean±SD)	30.6±6.8	31.2±7.1	29.3±5.8	0.117
Comorbidities				
CAD, n (%)	30 (19%)	25 (22%)	5 (11%)	0.118
CHF, n (%)	32 (21%)	25 (22%)	7 (16%)	0.359
Hypertension, n (%)	104 (67%)	76 (68%)	28 (64%)	0.615
Hyperlipidemia, n (%)	88 (56%)	69 (62%)	19 (43%)	0.037
TIA or CVA, n (%)	13 (8%)	11 (10%)	2 (5%)	0.352
Diabetes, n (%)	30 (19%)	26 (23%)	4 (9%)	0.037
OSA, n (%)	59 (38%)	43 (38%)	16 (36%)	0.784
CKD, n (%)	24 (15%)	17 (15%)	7 (16%)	0.872
Prior AF ablation, n (%)	43 (28%)	26 (23%)	17 (39%)	0.052
Type of AF				0.210
Paroxysmal AF, n (%)	67 (43%)	47 (42%)	20 (46%)	
Persistent AF, n (%)	66 (42%)	45 (40%)	21 (48%)	
Long-standing persistent AF, n (%)	23 (15%)	20 (19%)	3 (7%)	
AF ablation type*				0.248
Left atrial linear ablation	38 (24%)	30 (34%)	8 (18%)	
CTI	42 (27%)	35 (31%)	7 (16%)	
Antiarrhythmic drug use	34 (22%)	26 (23%)	8 (18%)	0.667

Values are n, mean±SD, or median (interquartile range). Categorical variables are compared using Fisher exact test; continuous variables using the *t* test or Mann-Whitney *U* test if data are not normally distributed. AF indicates atrial fibrillation; BMI, body mass index; CAD, coronary artery disease; CHF, congestive heart failure; CKD, chronic kidney disease; CTI, cavotricuspid isthmus ablation; OSA, obstructive sleep apnea; and TIA, transient ischemic attack.

*In addition to pulmonary vein isolation and ablation of localized rotational and focal sources by FIRM mapping.

prediction (Table 2, 4th row); using ECG signals only, the CNN achieved an AUROC of 0.767 (SD=0.122; Table 2, 5th row), both of which outperform APPLE and CHA2DS2-VASc scores.

In addition, we visualize examples of electrogram and ECG learned by the CNNs using the Uniform Manifold

Approximation and Projection²³ (UMAP) dimensionality reduction technique. As shown in Figure S2, the same patient's electrogram features are clustered together, whereas different patients' electrogram features are further apart. Moreover, electrogram/ECG features of patients with AF recurrence are further away from

Table 2. Results of 1-Year AF Recurrence Prediction

	AUROC	Sensitivity	Specificity	Accuracy	F1-score
APPLE score	0.644±0.129	0.915±0.138*	0.350±0.329	0.504±0.213	0.533±0.111
CHA2DS2-VASc score	0.650±0.133	0.905±0.162	0.427±0.355	0.560±0.226	0.568±0.124
Clinical Feature	0.755±0.093	0.875±0.137	0.680±0.198	0.728±0.121	0.656±0.102
EGM	0.731±0.105	0.885±0.116	0.627±0.131	0.701±0.098	0.630±0.092
ECG	0.767±0.122	0.812±0.176	0.770±0.183	0.781±0.112	0.682±0.108
Fusion of EGM and clinical data	0.788±0.110	0.905±0.117	0.706±0.144	0.764±0.107	0.691±0.117
Fusion of ECG and clinical data	0.836±0.063	0.865±0.112	0.812±0.124	0.827±0.070	0.747±0.075
Fusion of EGM and ECG	0.833±0.084	0.915±0.138*	0.793±0.124	0.826±0.083	0.753±0.096
Fusion of EGM, ECG and clinical feature	0.859±0.082*	0.870±0.200	0.867±0.121*	0.866±0.076*	0.784±0.106*

Values are mean±SD across 10-folds.

ECG indicates electrocardiogram; and EGM, electrogram.

*Best mean results for each metric.

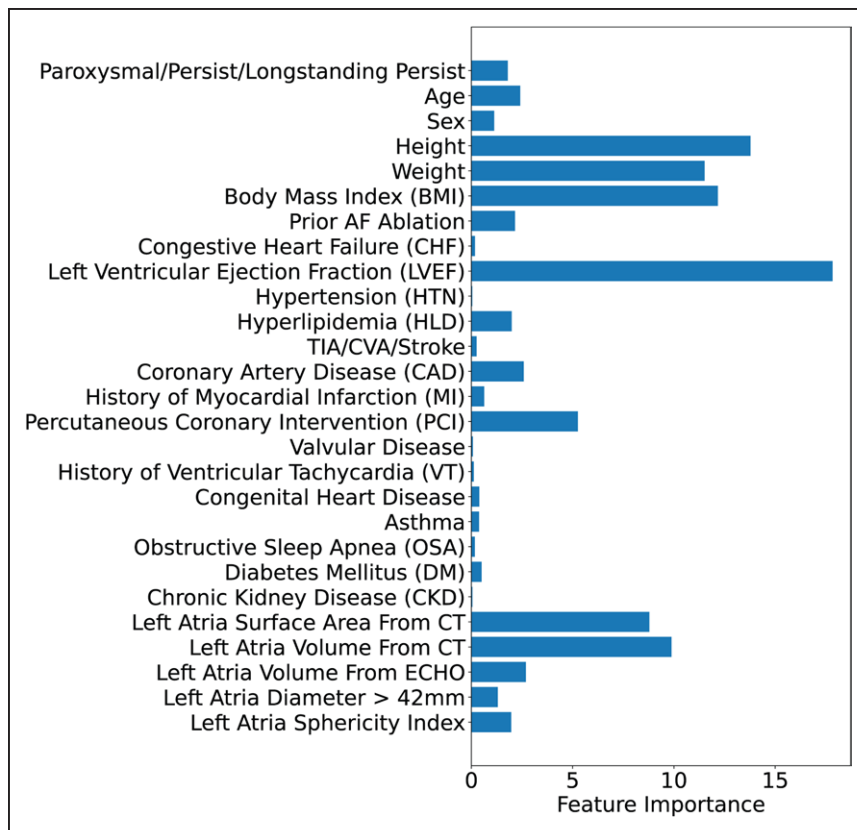


Figure 2. Clinical feature-based model interpretation.

Importance of clinical features in predicting atrial fibrillation (AF) recurrence using the CatBoost classifier (averaged across 10 folds). The 5 most important features are: left ventricular ejection fraction (LVEF), height, body mass index (BMI), weight, left atria volume from computed tomography (CT), and left atria surface area from CT.

features of patients without AF recurrence, suggesting that the CNNs are able to learn distinct patterns in patients with different outcomes.

ML-Based AF Recurrence Prediction From Fusion of Electrogram, ECG, and Clinical Features

Our final fusion model that combines electrogram, ECG, and clinical features achieved an AUROC of 0.859 (SD=0.082; Table 2, last row), outperforming the APPLE scores, CHA2DS2-VASc scores, and ECG or electrogram signals alone, suggesting the effectiveness of our fusion framework.

Figure 3 shows the ROC curves of the clinical feature-based models, the signal-based CNN models, and the fusion model. At a low false positive rate, such as 20% false positive rate, our fusion model had a true positive rate (TPR) of 80%, which translates clinically to missing 20% recurrent AF patients with 20% of the predicted recurrent AF being false positives. In contrast, the CHA2DS2-VASc score-based classifier and the clinical feature-based classifier only achieved a TPR of 40% and 58%, respectively, which translates to missing 60% and 42% recurrent AF patients, respectively, with the same number of false positives.

Moreover, combining 2 modalities performed better than single modalities (Table 2, 6th–8th rows), which is intuitive given that 2 modalities encode additional

features than a single modality. Model performance in various subgroups are provided in Supplemental Results and Tables S4 through S6.

In addition to discriminative measures (eg, AUROC, sensitivity, and specificity), we evaluate the calibration of the models using Brier score¹⁹ and expected calibration error.²⁰ See Supplemental Results and Table S7 for details.

DISCUSSION

In this study, we developed a deep convolutional neural network that encodes the spatiotemporal dependencies in electrogram and ECG signals, as well as a multimodal fusion framework that leverages clinical features, electrogram, and ECG for predicting 1-year AF recurrence after catheter ablation. Our study was based on a cohort of 156 patients.

To our knowledge, compared with the existing AF recurrence prediction scores to date, this provides the highest performance in predicting which patients would be free from AF 1 year following ablation.

Other studies evaluating prediction of AF ablation outcomes using machine learning include Shade et al⁹ that utilized ML and personalized computational modeling in 32 patients to predict AF recurrence following PVI with either cryoballoon or radiofrequency approach. In their machine learning model, their sources of information (imaging, clinical data) were combined equally.⁹ Late

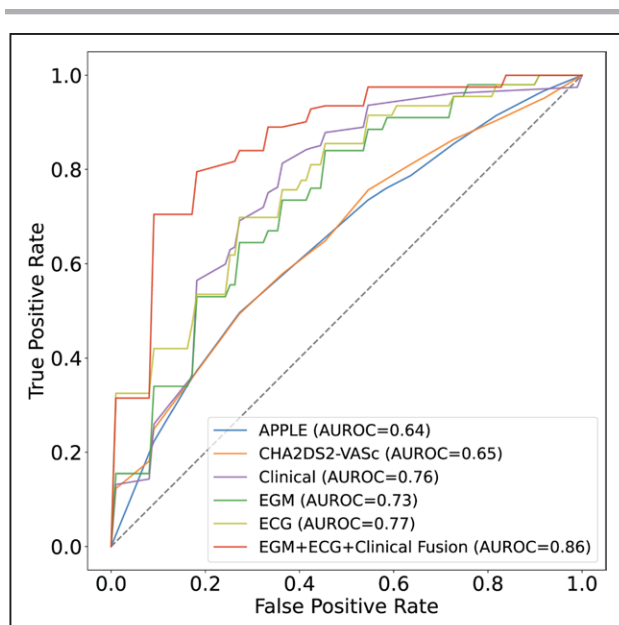


Figure 3. Receiver operating characteristics (ROC) curves of the clinical feature-based models, signal-based models, and the fusion model.

The x-axis shows the false positive rate averaged across 10 folds for each model, and the y-axis shows the true positive rate averaged across 10 folds for each model. AUROC indicates area under the receiver operating characteristics curve.

gadolinium enhanced magnetic resonance imaging scans were used for imaging data. AUROC of 0.82 was reported when clinical variables were included in the model. Firouznia et al¹⁰ extracted data from chest CT scans to establish their association with likelihood of postablation AF recurrence in 203 patients using a random forest classifier. Certain derived imaging features such as left atrial surface area, volume, and sphericity index used in their study were also included in our model as a part of clinical features.¹⁰ PVI in this study was completed with either cryoballoon or radiofrequency catheters. Moreover, posterior wall, septal, superior vena cava, and cavotricuspid isthmus ablation were performed according to operator choice, although further details of extra-PVI ablation were not discussed in the study or included in the models.

In our study, all patients underwent PVI with cryoballoon or radiofrequency approach. Similar to Firouznia et al,¹⁰ patients undergoing various ablation strategies were included, including ablation of localized sources detected by FIRM mapping strategy in 100% of patients, left atrial linear lesions in 24%, and cavotricuspid isthmus ablation in 27% of patients. FIRM strategy was used in all patients as it allowed simultaneous recording of unipolar signals in the left atrium before any ablation in this cohort, which was a prerequisite in our analysis. Our models were able to predict long-term (1 year) freedom from arrhythmias independent of the ablation strategy. Clinical benefit of lesions beyond PVI in patients with persistent AF has been a subject of debate, with multiple studies showing no additional benefit of extra PV lesions in long term freedom from

AF,^{24,25} with some demonstrating incremental benefit,^{26,27} and larger multicenter studies underway to evaluate this further.²⁸ Furthermore, incorporation of intracardiac electrograms indeed improved prediction of AF ablation outcomes, suggesting that an AF mechanism might be at play that could be delineated further by feature interpretation of these signals. Given the wide variety of ablation approaches used in the training and testing cohorts for our machine learning model, and limited representation of subgroups such as women, generalizability of our findings to the broader population could be limited.

Limitations

This study was performed at a single center, involves a small cohort with underrepresentation of women, and results have not been validated externally. Heterogeneity in ablation approaches may limit generalizability of the findings to specific ablation strategies. Despite this limitation, all the patients underwent PVI, and evidence of benefit of further ablation beyond PVI, including linear ablation and ablation of sites of organized rotational or focal activation, has not been proven consistently in multicenter randomized studies.^{24,29,30} All patients in this study underwent FIRM mapping and ablation that formed the basis of the unipolar EGMs used in the model. The necessity of the use of FIRM mapping is a limitation to this study, as this is not a widely used catheter or mapping strategy in the community.

Freedom from AF appears higher for a mixed cohort of patients but is consistent with other studies that used intermittent monitoring rather than implanted loop recorders. Intermittent monitoring of AF recurrence with 12-lead ECGs and 14-day event monitors likely underrepresents true AF recurrence, which could affect the accuracy of our predictive model. The retrospective nature of the data limited strict guidelines over AAD use in follow up, that is, for certain patients, preprocedure AADs were continued postablation due to patient or provider preference regardless of procedure outcome. Twenty-eight percent of patients had prior AF ablation, which may have impacted intracardiac signal characteristics. Twelve-lead ECGs in sinus rhythm before ablation were not available in all patients. When a patient's 12-lead ECGs in sinus rhythm before ablation was not available, a 12-lead ECG in sinus rhythm immediately after ablation was used for analysis, which could result in bias in analyses. While we show that when evaluating the trained models on patients whose preablation 12-lead ECGs are available, the model performance did not differ significantly from our original analysis (ie, post-ablation 12-lead ECGs are used for patients whose preablation ECGs are not available), we did not re-train the models on preablation ECGs only due to the limited size of our cohort (n=107 patients with preablation ECGs). Majority of these patients had a 12-lead ECG in sinus rhythm before ablation that was not performed at our center in

an electronic format that could be exported for analysis, due to the tertiary referral center status where the study was conducted. Some of the data that were used in the models to predict ablation success, including intracardiac signals, are obtained at the time of ablation, and may not help in patient selection for ablation procedure, but can rather guide medical management and expectations following the procedure. Furthermore, while we show that most of the trained models perform similarly on patient subgroups (patients with paroxysmal versus nonparoxysmal AF; patients with cryoablation versus radiofrequency ablation), future study with a larger cohort that trains models on these subgroups independently is needed to further compare these subgroups. Last, while we show that our CNNs and fusion model are better calibrated than the existing APPLE and CHADS2-VASc scores (Table S7), the Brier scores and expected calibration errors are still relatively high; advanced calibration techniques³¹ for deep neural networks need to be incorporated in the future to produce better calibrated models.

Conclusions

Our machine learning approach provides an automatic technique to predict freedom from atrial arrhythmias in patients undergoing AF ablation, outperforming traditional scoring systems. Larger datasets are needed in the future to train and validate this approach even further to help develop personalized ablation strategies for patients with AF.

ARTICLE INFORMATION

Received January 12, 2022; accepted June 29, 2022.

Affiliations

Department of Electrical Engineering (S.T.); Department of Medicine, Division of Cardiovascular Medicine (R.K., M.I.A., M.F., A.J.R., M.R.B., P.C., P.J.W., S.M.N., T.B.); Department of Radiology and Biomedical Data Science (D.L.R.), Stanford University, CA. University College London, Centre for Advanced Research Computing, United Kingdom (O.R.). CoMMLab, Universitat de Valencia, VA, Spain (M.R.B.).

Sources of Funding

This work was funded by National Institutes of Health (K23 HL145017) grant to Dr Baykaner.

Disclosures

M.I. Alhusseini reports intellectual property rights from Stanford University. Dr Fazal reports no disclosures. Dr Rubin reports grants from NIH (1U01CA190214, 1U01CA187947, U01CA242879, U24CA226110) and consulting fees from Roche-Genentech, Dr Narayan reports research grants from NIH (HL70529, HL103800, HL83359), consulting from LifeSignals.ai Inc, TDK Inc., Up to Date, Abbott Laboratories, and American College of Cardiology Foundation (all modest); Intellectual Property Rights from University of California Regents and Stanford University. Dr Baykaner reports funding from NIH (K23 HL145017) and consulting fees from Medtronic, BIOTRONIK and PaceMate.

Supplemental Material

Supplemental Methods
Supplemental Results
Tables S1–S7
Figures S1 and S2
References^{32–35}

REFERENCES

- Packer DL, Mark DB, Robb RA, Monahan KH, Bahnson TD, Poole JE, Noseworthy PA, Rosenberg YD, Jeffries N, Mitchell LB, et al; CABANA Investigators. Effect of catheter ablation vs antiarrhythmic drug therapy on mortality, stroke, bleeding, and cardiac arrest among patients with atrial fibrillation: the CABANA Randomized Clinical Trial. *JAMA*. 2019;321:1261–1274. doi: 10.1001/jama.2019.0693
- Marrouche NF, Brachmann J, Andresen D, Siebels J, Boersma L, Jordaens L, Merkely B, Pokushalov E, Sanders P, Proff J, et al; CASTLE-AF Investigators. Catheter ablation for atrial fibrillation with heart failure. *N Engl J Med*. 2018;378:417–427. doi: 10.1056/NEJMoa1707855
- Kornej J, Hindricks G, Arya A, Sommer P, Husser D, Bollmann A. The APPLE score - a novel score for the prediction of rhythm outcomes after repeat catheter ablation of atrial fibrillation. *PLoS One*. 2017;12:e0169933. doi: 10.1371/journal.pone.0169933
- Jacobs V, May HT, Bair TL, Crandall BG, Cutler M, Day JD, Weiss JP, Osborn JS, Muhlestein JB, Anderson JL, et al. The impact of risk score (CHADS2 versus CHA2DS2-VASc) on long-term outcomes after atrial fibrillation ablation. *Heart Rhythm*. 2015;12:681–686. doi: 10.1016/j.hrthm.2014.12.034
- Kosich F, Schumacher K, Potpara T, Lip GY, Hindricks G, Kornej J. Clinical scores used for the prediction of negative events in patients undergoing catheter ablation for atrial fibrillation. *Clin Cardiol*. 2019;42:320–329. doi: 10.1002/clc.23139
- Rogers AJ, Selvalingam A, Alhusseini MI, Krummen DE, Corrado C, Abuzaid F, Baykaner T, Meyer C, Clopton P, Giles W, et al. Machine learned cellular phenotypes in cardiomyopathy predict sudden death. *Circ Res*. 2021;128:172–184. doi: 10.1161/CIRCRESAHA.120.317345
- Bhalodia R, Goparaju A, Sodergren T, Morris A, Kholmovski E, Marrouche N, Cates J, Whitaker R, Elhabian S. Deep Learning for End-to-End Atrial Fibrillation Recurrence Estimation. In: *2018 Computing in Cardiology Conference (CinC)*. IEEE; 2018. p. 1–4. DOI: 10.22489/CinC.2018.160
- Budzianowski J, Hiczkiewicz J, Burchard P, Pleszko K, Rzeźniczak J, Budzianowski P, Korybalska K. Predictors of atrial fibrillation early recurrence following cryoballoon ablation of pulmonary veins using statistical assessment and machine learning algorithms. *Heart Vessels*. 2019;34:352–359. doi: 10.1007/s00380-018-1244-z
- Shade JK, Ali RL, Basile D, Popescu D, Akhtar T, Marine JE, Spragg DD, Calkins H, Trayanova NA. Preprocedure application of machine learning and mechanistic simulations predicts likelihood of paroxysmal atrial fibrillation recurrence following pulmonary vein isolation. *Circ Arrhythm Electrophysiol*. 2020;13:e008213. doi: 10.1161/CIRCEP.119.008213
- Firouznia M, Feeny AK, LaBarbera MA, McHale M, Cantlay C, Kalfas N, Schoenhagen P, Saliba W, Tchou P, Barnard J, et al. Machine learning-derived fractal features of shape and texture of the left atrium and pulmonary veins from cardiac computed tomography scans are associated with risk of recurrence of atrial fibrillation postablation. *Circ Arrhythm Electrophysiol*. 2021;14:e009265. doi: 10.1161/CIRCEP.120.009265
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444. doi: 10.1038/nature14539
- Esteve A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, Cui C, Corrado G, Thrun S, Dean J. A guide to deep learning in healthcare. *Nat Med*. 2019;25:24–29. doi: 10.1038/s41591-018-0316-z
- Firouznia M, Feeny AK, LaBarbera MA, McHale M, Cantlay C, Kalfas N, Schoenhagen P, Saliba W, Tchou P, Barnard J, et al. Machine learning-derived fractal features of shape and texture of the left atrium and pulmonary veins from cardiac computed tomography scans are associated with risk of recurrence of atrial fibrillation postablation. *Circulation: Arrhythmia and Electrophysiology*. 2021;14:e009265. doi: 10.1161/CIRCEP.120.009265
- Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, Carter RE, Yao X, Rabinstein AA, Erickson BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet*. 2019;394:861–867. doi: 10.1016/S0140-6736(19)31721-0
- Lavie CJ, Pandey A, Lau DH, Alpert MA, Sanders P. Obesity and atrial fibrillation prevalence, pathogenesis, and prognosis: effects of weight loss and exercise. *J Am Coll Cardiol*. 2017;70:2022–2035. doi: 10.1016/j.jacc.2017.09.002
- Khaykin Y, Oosthuizen R, Zarnett L, Essebag V, Parkash R, Seabrook C, Beardsall M, Tsang B, Wulffhart Z, Verma A. Clinical predictors of arrhythmia recurrences following pulmonary vein antrum isolation for atrial fibrillation: predicting arrhythmia recurrence post-PVAI. *J Cardiovasc Electrophysiol*. 2011;22:1206–1214. doi: 10.1111/j.1540-8167.2011.02108.x

17. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. In: *Advances in Neural Information Processing Systems*. 2018.
18. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. 2015.
19. Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev*. 1950;78:1-3. Accessible at: <https://books.google.com/books?hl=en&lr=&id=jnbpAAAAMAAJ&oi=fnd&pg=PA11-1A6&ots=OX13-JoHzM&sig=66qS3nq3yYAE0mAcAPh2bVkgGcHE#v=onepage&q&f=false>
20. Naeini MP, Cooper GF, Hauskrecht M. Obtaining well calibrated probabilities using bayesian binning. *Proc Conf AAAI Artif Intell*. 2015;2015:2901-2907.
21. Santhanakrishnan R, Wang N, Larson MG, Magnani JW, McManus DD, Lubitz SA, Ellinor PT, Cheng S, Vasan RS, Lee DS, et al. Atrial fibrillation begets heart failure and vice versa: temporal associations and differences in preserved versus reduced ejection fraction. *Circulation*. 2016;133:484-492. doi: 10.1161/CIRCULATIONAHA.115.018614
22. Pathak RK, Middeldorp ME, Lau DH, Mehta AB, Mahajan R, Twomey D, Alasady M, Hanley L, Antic NA, McEvoy RD, et al. Aggressive risk factor reduction study for atrial fibrillation and implications for the outcome of ablation: the ARREST-AF cohort study. *J Am Coll Cardiol*. 2014;64:2222-2231. doi: 10.1016/j.jacc.2014.09.028
23. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv [stat.ML]*. 2018. doi: <https://doi.org/10.48550/arXiv.1802.03426>
24. Verma A, Jiang CY, Betts TR, Chen J, Deisenhofer I, Mantovan R, Macle L, Morillo CA, Haverkamp W, Weerasooriya R, et al; STAR AF II Investigators. Approaches to catheter ablation for persistent atrial fibrillation. *N Engl J Med*. 2015;372:1812-1822. doi: 10.1056/NEJMoa1408288
25. Brachmann J, Hummel JD, Wilber DJ, Sarver AE, Rapkin J, Shpun S, Szili-Torok T. Prospective randomized comparison of rotor ablation vs conventional ablation for treatment of persistent atrial fibrillation—The REAF-FIRM trial. *Heart Rhythm*. 2019;16:963-965.
26. Clarke JD, Piccini JP, Friedman DJ. The role of posterior wall isolation in catheter ablation of persistent atrial fibrillation. *J Cardiovasc Electrophysiol*. 2021;32:2567-2576. doi: 10.1111/jce.15164
27. Narayan SM, Baykaner T, Clopton P, Schricker A, Lalani GG, Krummen DE, Shivkumar K, Miller JM. Ablation of rotor and focal sources reduces late recurrence of atrial fibrillation compared with trigger ablation alone: extended follow-up of the CONFIRM trial (Conventional Ablation for Atrial Fibrillation With or Without Focal Impulse and Rotor Modulation). *J Am Coll Cardiol*. 2014;63:1761-1768. doi: 10.1016/j.jacc.2014.02.543
28. Terricabras M, Piccini JP, Verma A. Ablation of persistent atrial fibrillation: Challenges and solutions. *J Cardiovasc Electrophysiol*. 2020;31:1809-1821. doi: 10.1111/jce.14311
29. Thiagarajah A, Kadhim K, Lau DH, Emami M, Linz D, Khokhar K, Munawar DA, Mishima R, Malik V, O'Shea C, et al. Feasibility, safety, and efficacy of posterior wall isolation during atrial fibrillation ablation: a systematic review and meta-analysis. *Circ Arrhythm Electrophysiol*. 2019;12:e007005. doi: 10.1161/CIRCEP.118.007005
30. Kirzner JM, Raelson CA, Liu CF, Thomas G, Ip JE, Lerman BB, Markowitz SM, Cheung JW. Effects of focal impulse and rotor modulation-guided ablation on atrial arrhythmia termination and inducibility: Impact on outcomes after treatment of persistent atrial fibrillation. *J Cardiovasc Electrophysiol*. 2019;30:2773-2781. doi: 10.1111/jce.14240
31. Guo C, Pleiss G, Sun Y, Weinberger KQ. On Calibration of Modern Neural Networks. In: Precup D, Teh YW, editors. *Proceedings of the 34th International Conference on Machine Learning*. PMLR; 2017. p. 1321-1330.
32. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: Bach F, Blei D, editors. *Proceedings of the 32nd International Conference on Machine Learning*. Lille, France: PMLR; 2015. p. 448-456.
33. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2016. pp. 770-778.
34. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15:1929-1958. doi: 10.1109/TCYB.2020.3035282
35. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837-845. doi: 10.2307/2531595