

<https://doi.org/10.1038/s43856-025-01058-4>

Predicting arrhythmia recurrence post-ablation in atrial fibrillation using explainable machine learning

Check for updates

Savannah F. Bifulco^{1,5}, Matthew J. Magoon^{1,5}, Yaacoub Chahine², Issac Kim¹, Fima Macheret², Nazem Akoum^{1,2,6} ✉ & Patrick M. Boyle^{1,3,4,6} ✉

Abstract

Background Following atrial fibrillation ablation, it is challenging to distinguish patients who will remain arrhythmia-free from those at risk for recurrence. New explainable machine learning (xML) techniques allow for systematic assessment of arrhythmia recurrence risk following catheter ablation. We aim to develop an xML algorithm that predicts recurrence and reveals key risk factors to facilitate better follow-up strategy after an ablation procedure. **Methods** We reconstructed pre- and post-ablation models of the left atrium (LA) from late gadolinium enhanced magnetic resonance (LGE-MRI) for 67 patients. Patient-specific features (LGE-based measurements of pre/post-ablation arrhythmogenic substrate, LA geometry metrics, computational simulation results, and clinical risk factors) trained a random forest classifier to predict recurrent arrhythmia. We calculated each risk factor's marginal contribution to model decision making via SHapley Additive exPlanations (SHAP). **Results** The classifier accurately predicts post-ablation arrhythmia recurrence (mean receiver operating characteristic [ROC] area under the curve [AUC]: 0.80 ± 0.04 ; mean precision-recall [PR] AUC: 0.82 ± 0.08). SHAP analysis reveals that of 89 features tested, the key population risk factors for recurrence are: large left atrium, low LGE-quantified post-ablation scar in the atrial floor region, and previous attempts at direct current cardioversion. We also examine patient-specific recurrence predictions, since xML allows us to understand why a particular individual can have large prediction weights for some categories without tipping the balance towards an incorrect prediction. Finally, we validate our model in a completely new, 15-patient retrospective holdout cohort (80% correct). **Conclusion** Our SHAP-based explainable machine learning approach is a proof-of-concept clinical tool to explain arrhythmia recurrence risk in patients who underwent ablation by combining patient-specific clinical profiles and LGE-derived data.

Plain language summary

Atrial fibrillation (AFib) is a common heart rhythm problem. It is treated by catheter ablation, in which a thin flexible tube is inserted into the heart and a treatment administered that will destroy the part of the heart from which the abnormal heart rhythms originate. We used a computational method to predict whether AFib would come back after ablation. We trained our model on detailed heart scans, clinical data, and computer simulations from 67 patients. Our method accurately predicted which patients would have a recurrence and highlighted important risk factors, such as large heart size, specific scar distributions after ablation, and people having had previous electrical shock therapy. We confirmed our model worked well in a separate group of 15 patients. Our approach could help doctors better understand individual patient risks and plan more effective follow-up care after ablation.

Atrial fibrillation (AFib) is the most common cardiac arrhythmia, affecting 1–2% of the world's population and significantly contributing to morbidity and mortality¹. Pulmonary vein isolation (PVI) is an established rhythm control strategy and is the cornerstone of catheter ablation for AFib treatment, but it results in recurrent atrial arrhythmia (AA) in ~20–40% of patients^{2,3}. Differentiating patients at risk for post-ablation AA recurrence from those who will remain arrhythmia-free is challenging. Developing a

method to predict recurrent arrhythmia following ablation via explainable machine learning (xML) could provide valuable insights for ablation planning and decision-making, leading to improved outcomes in AFib patients.

Prior work has investigated mechanisms and risk factors associated with AFib recurrence. These studies have considered clinical features such as hypertension⁴, obesity⁵, diabetes⁶, cardiomyopathy⁷, and smoking status⁸.

¹Department of Bioengineering, University of Washington, Seattle, WA, USA. ²Division of Cardiology, University of Washington, Seattle, WA, USA. ³Institute for Stem Cell & Regenerative Medicine, University of Washington, Seattle, WA, USA. ⁴Center for Cardiovascular Biology, University of Washington, Seattle, WA, USA.

⁵These authors contributed equally: Savannah F. Bifulco, Matthew J. Magoon. ⁶These authors jointly supervised this work: Nazem Akoum, Patrick M. Boyle.

✉ e-mail: nakoum@uw.edu; pmjboyle@uw.edu

Left atrial (LA) models derived from late gadolinium enhanced magnetic resonance imaging (LGE-MRI) also offer a means to characterize potential arrhythmogenic substrate⁹. These models have been leveraged to investigate fibrosis¹⁰, ablation-delivered scar^{11–13}, and LA shape characteristics as risk factors for recurrent arrhythmia^{14,15}. 12 lead electrocardiograms (ECGs) also enable spectral analysis of atrial fibrillatory waves (f-waves), and others report the amplitude and dominant frequency of pre-ablation f-waves correlate with durable ablation success^{16–20}. Integrating this rich multi-modal risk factor data for prediction of post-ablation recurrence is a potential avenue for generating a robust classifier and furthering our scientific understanding of AFib.

Existing machine learning algorithms have leveraged various features such as pre-ablation LGE-MRI imaging data, patient-specific simulations, and electronic health records (EHRs) to predict recurrent AFib^{21–24}, with varying degrees of success; area under the receiver operating characteristic curve (ROC AUCs) values ranged from 0.61 to 0.85. Notably, these algorithms lack robust explainability metrics to elucidate how input features influence the final decision, which has hindered clinical implementation²⁵. Moreover, quantitative characterization of the extent and location of ablation-induced scar (e.g., from post-ablation LGE-MRI) has not yet been included in these algorithms despite its substantial impact on procedural outcomes^{26,27}.

In this proof-of-concept study, we develop an xML-based recurrent arrhythmia prediction model that combines patient-specific fibrotic tissue, ablation-delivered scar (assessed post-ablation), LA geometry metrics, simulations conducted in computational models reconstructed from pre- and post-ablation patient MRI, and clinically relevant EHR data. This method accurately predicts the likelihood of recurrence (mean receiver operating characteristic [ROC] area under the curve [AUC]: 0.80 ± 0.04 ; mean precision-recall [PR] AUC: 0.82 ± 0.08). This performance is validated (80% correct) in a 15-patient retrospective holdout cohort comprising data never seen in during model training or validation. The algorithm's output points to risk factors that are most influential in the algorithm's decision, specifically large left atrium, low LGE-quantified post-ablation scar in the atrial floor region, and previous attempts at direct current cardioversion. These risk factors can be analyzed on a cohort-wide or patient-specific scale, offering important contextualization to xML findings that can be further tested in randomized trials.

Methods

Patient cohort and image acquisition

This study retrospectively included patients from University of Washington (UW) Medical Center with documented persistent AFib or paroxysmal AFib who had already received both pre- and post-procedural LGE-MRI scans and underwent either cryoballoon or radiofrequency (RF) ablation. Paroxysmal AFib was defined by AFib episodes at least 30 s in duration that terminated within 7 days spontaneously or in response to intervention²⁸. Persistent AFib was defined by AFib episodes that persisted for a minimum of 7 days²⁸. Cardiac LGE-MRIs were obtained using previously described protocols for all participants within 90 days prior to their ablation procedure and again 3–6 months post-ablation to quantify the extent of LA fibrosis and scar, respectively²⁹. Exclusion criteria for AFib patients included those who had a prior catheter ablation, patients with cardiac implantable electronic devices, severe claustrophobia, renal dysfunction, and contraindications to MRI or gadolinium-based contrast. Scans were performed on the Philips Ingenia system, 15–25 min after contrast injection, using a three-dimensional inversion-recovery, respiration-navigated, ECG-gated, gradient echo pulse sequence. Acquisition parameters included transverse imaging volume with a voxel size of $1.25 \times 1.25 \times 2.5$ mm (reconstructed to $0.625 \times 0.625 \times 1.25$ mm). Scan time was 5–10 min dependent on respiration and heart rate.

Patients had clinical assessment and catheter ablation in the UW AFib program. All patients underwent PVI, and some had additional substrate modification at the operator's discretion. Patients' clinical features were determined at time of initial visit and are tabulated in Supplementary Data 1,

including persistent vs. paroxysmal AFib status³⁰, comorbidities, and medications. The symptom severity feature was determined by assessing the burden of self-reported AFib symptoms on a scale of 1 to 4³¹. Presence of cardiomyopathy was defined by myocardial dysfunction with or without heart failure, including genetic cardiomyopathies, valvular cardiomyopathy, and cardiac sarcoidosis. Following ablation, patients were followed longitudinally at UW with 7-day ambulatory electrocardiogram monitoring at 3, 6, and 12 months after ablation. Recurrence was defined by at least 30 s of documented AA after a 90-day blanking period³⁰. The recurrence rhythm was classified as either AFib or atrial flutter by expert ECG interpretation. Loss-to-follow-up bias was limited as all patients completed at least 2 years of prespecified follow-up. There was no missing data for any patient.

Anatomical model reconstruction

Geometric models were reconstructed from both pre- and post-ablation LGE-MRI scans by Merisight Inc. (Salt Lake City, UT) to assess LA volume and surface area. Geometric models were reconstructed from pre-ablation scans, and the relative extent of fibrosis in the LA was quantified via an adaptive histogram thresholding algorithm to determine pre-ablation LGE-MRI derived fibrosis³². For post-ablation models, ablation scar was quantified on post-ablation LGE-MRI using previously established methods^{31,33}. Non-rigid registration was used to map LGE-derived post-ablation scar patterns onto existing LA pre-ablation fibrotic models. Hyper-enhancement on post-ablation scans was assumed to be ablation-induced scar; this accounts for the fact that hyperenhancement from ablation scar is at a higher absolute level than that of native fibrosis. Consequently, regions labeled as fibrotic pre-ablation fall below the hyperenhancement threshold in post-procedure scans.

Extraction of LGE-MRI derived fibrosis and ablation-delivered scar features

In pre- and post-ablation models, we characterized fibrosis and ablation-delivered scar area in five regions defined with respect to LA landmarks, as in our prior work: left pulmonary veins (LPVs), right pulmonary veins (RPVs), posterior wall, anterior wall, and atrial floor³⁴. First, the LA was subdivided into three broad anatomical areas LA floor, posterior wall, and anterior wall including left atrial appendage using standardized cutoff values in the UAC space³⁵. Then, LPV and RPV areas were established using a region-growing approach such that each accounted for 15% of the total LA surface area.

Average fibrosis entropy and density were also calculated for pre- and post-ablation models. Prior computational and clinical work has shown that regions tending to harbor reentrant driver (RD) activity are characterized by fibrotic boundary zones with high fibrosis entropy (FE) and high fibrosis density (FD)^{10,36}. Thus, we calculated the extent of such regions in each patient-specific model using the same equation derived via machine learning, as in prior studies^{10,37}:

$$0.4096(\text{FD})^2 + 3.28(\text{FD})(\text{FE}) - 0.1036(\text{FE})^2 - 0.7112(\text{FD}) - (\text{FE}) + 0.0429.$$

Prior computational work suggests ablation-induced scar and certain non-conductive anatomical landmarks (i.e., mitral valve annuli, pulmonary vein ostia) contribute to recurrent AA¹³. Any region of hyper-enhancement in the post-ablation LGE-MRI, compared to the pre-ablation LGE-MRI, was considered ablation-delivered scar. We counted the total number of scar regions with area >1 cm². Within this set we also counted the number of LGE-derived scar areas and non-conductive anatomical landmarks of specific size (area from 2 to 20 cm²; perimeter from 15 to 60 cm) and in proximity to fibrosis ($>30\%$ fibrosis in the surrounding 1 cm area), as these indicate potentially arrhythmogenic substrate¹³.

Design and evaluation of random forest machine learning classifier

Figure 1 provides a simple flowchart of the model development and explanation workflow. The Least Absolute Shrinkage and Selection Operator (LASSO) approach was used to reduce the overall number of features³⁸. This algorithm performs well when the number of observations is

low and the number of features is high. LASSO attempts to eliminate variables that are irrelevant (i.e., unrelated to the outcome) or highly collinear. Five-fold cross-validation, considering only data from the original cohort, identified the optimal parameterization (Supplementary Fig. 1). LASSO regression was applied to the full original cohort with the optimal hyperparameter (Fig. 1A). The resulting set of features was then used to train a random forest machine learning classifier to recognize risk factors associated with recurrent AA (Fig. 1B)^{39,40}. We used five-fold stratified, 80:20-split cross-validation to mitigate the risk of overfitting the model; this approach ensured an equal distribution of recurrent and non-recurrent AFib patients in training and test sets. At no point were data from the holdout cohort incorporated in LASSO regression or random forest model development.

The Shapley Additive explanations (SHAP) framework facilitates algorithmic model interpretations via calculation of marginal contributions (SHAP values) of each risk factor for each prediction. Briefly, SHAP values are calculated by evaluating the model's response to perturbation of each

feature, revealing the relative influence on the model's final prediction⁴¹. We applied SHAP to probe feature importance in the trained random forest classifier, providing insight into the overall influence of each risk factor^{42,43}.

Finally, we applied the fully trained and validated random forest classifier to a completely new holdout cohort of 15 patients never seen by the classifier in any prior training or test set. Only the features used in the random forest classifier were extracted for this cohort. As for the original model, we used SHAP to examine each feature's importance for each patient in the holdout set.

Computational simulations of patient-specific electrophysiology

A detailed description of computational simulation methods can be found in the "Supplementary Methods" section of the Supplementary Information. Briefly, fibrotic LA models were constructed and parameterized as in previous work^{13,35,44}. Our methodology for computational modeling at the cell⁴⁵ and tissue scales⁴⁶ can also be found in previously published papers^{10,47,48}. Simulations were performed on the Hyak supercomputer

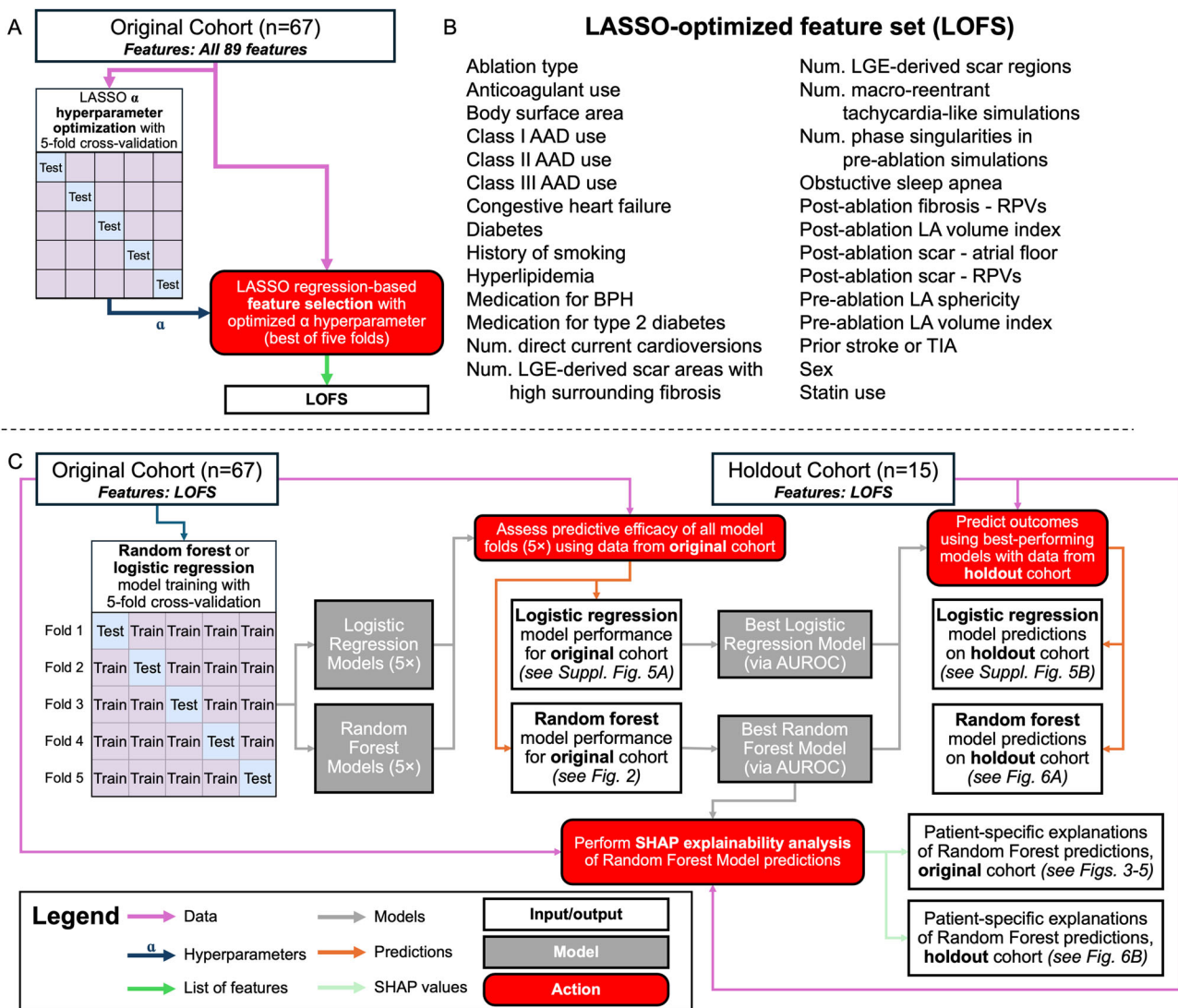


Fig. 1 | Flowchart outlining the complete model development process. A To address multicollinearity between the 89 risk factors compiled for this study, LASSO regression removed the least important and collinear features. B The product of this first stage was the 27-element LASSO-optimized feature set (LOFS). C Subsequently, the LOFS were used to train and test either random forest machine learning (ML) or logistic regression models using five-fold, 80:20-split cross-validation. The random forest and logistic regression models were then tested on data from a never-before-seen 15-patient holdout cohort. To assess model explainability, the marginal

contributions of individual LOFS values on overall random forest model predictions in the original and holdout cohorts were evaluated by SHAP analysis. SHAP analysis was not needed for the logistic regression model, as each feature had a coefficient explicitly describing its impact on model predictions. Holdout and explainability tests were always performed on the single best logistic regression or random forest model, as assessed during the predictive efficacy stage via the area under the receiver operating characteristic curve (AUROC) metric. Links to relevant figures later in the study in which specific results are presented are provided in Output panels.

Table 1 | Patient characteristics in combined training and testing cohort

	Recurrence	No Recurrence	p-value
N (%)	40 (59.7%)	27 (40.3%)	n/a
Age (years)	65 (53–77)	59 (46–72)	0.4896
BMI	31.2 (20.4–42.0)	29.9(22.5–37.5)	0.9747
Sex (male)	60.0%	55.6%	0.7251
Hypertension	57.5%	59.3%	0.8929
Hyperlipidemia	37.5%	40.7%	0.7968
Coronary artery disease	20%	14.8%	0.5966
OSA	32.5%	33.3%	0.7091
Stroke/TIA	10.0%	3.7%	0.3469
Diabetes	10.0%	7.4%	0.7272
COPD	10.0%	0%	0.09566
Cigarette use	17.5%	18.5%	0.9233
Thyroid disease	15.0%	7.4%	0.6570
CHF	32.5%	22.2%	0.3679
Pre-Ablation left atrial LGE (%)	15.6% (11.48–21.9%)	17.6% (8.95–20.25%)	0.9033
AFib ablation type:			
Cryo PVI only	35.0%	44.4%	0.4444
Cryo PVI + RF SM	5.0%	3.7%	0.8166
RF PVI only	17.5%	11.1%	0.4814
RF PVI + SM	42.5%	40.7%	0.8929

Variables are reported as median (IQR) and compared with Mann-Whitney U tests as appropriate. For binary variables, proportions were reported and compared with χ^2 tests. BMI body mass index, CHF congestive heart failure, COPD chronic obstructive pulmonary disease, LGE late gadolinium enhancement, OSA obstructive sleep apnea, PVI pulmonary vein isolation, RF radiofrequency, SM substrate modification, TIA transient ischemic attack.

system at UW using the openCARP simulation environment for cardiac electrophysiology⁴⁹. In each pre- and post-ablation model, virtual burst pacing to attempt reentry initiation was applied at 15 LA sites at locations corresponding to common AFib trigger origins⁵⁰. The presence of RDs was characterized by detecting persistent phase singularities (i.e., organizing centers of reentry)^{51,52}. We also counted the number of macro-reentrant tachycardia morphologies in pre- and post-ablation simulations. For each patient, we aggregated simulation features tabulated in Supplementary Data 1.

Electrocardiographic f-wave analysis

Others have reported the value of f-wave analysis for AFib recurrence risk stratification^{16–20}. Raw signal data for each pre-ablation 12-lead electrocardiogram (ECG) interpreted by an expert clinician as AA were analyzed for f-wave characteristics. Patients were included in this analysis if we had access to raw electrical recordings from a 10 s, resting, 12-lead ECG performed at a UW-affiliated hospital or clinic; individuals whose AFib diagnosis was confirmed via ECG at unaffiliated institutions were thus excluded from this part of the analysis for lack of access to raw ECG recordings. If multiple pre-ablation ECGs were available showing AA, the most recent recording was used. The signal was decomposed into atrial and ventricular components by spatiotemporal QRST cancellation⁵³. The atrial component was analyzed to find the dominant frequency in lead I between 4.83 and 11.67 Hz (290–700 bpm). A sliding window was used to approximate the mean amplitude across the signal.

Statistics and reproducibility

Random forest and logistic regression models (coded using Python v3.10.11, developed with scikit-learn⁵⁴) were evaluated with receiver operator

characteristic (ROC) and precision-recall (PR) curves. The area under the curve (AUC) was reported in either case, along with F1-scores. SHAP analysis indicated relationships between patient features and random forest model classifications^{41–43}. Reference the Data Availability and Code Availability statements for more information about accessing the data and code underlying this work^{55,56}. This study was approved by the UW institutional review board (IRB), and all participants provided written informed consent.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

In the cohort used for training and testing (Table 1), 40 of 67 patients (59.7%) experienced recurrence of either AFib (29/40; 72.5%) or atrial flutter (11/40; 27.5%) in the two-year follow-up period. 36 of 67 patients (53.7%) underwent circumferential wide area RF or segmental cryoballoon ablation around the PVs without additional substrate modification. The remaining 31 of 67 patients (46.3%) had additional substrate modification via RF ablation (e.g., posterior wall isolation). The median time to reported recurrence was 153 days. Follow-up time was limited to two years for every patient. No patients were lost to follow-up.

A table of all 89 features with sub-classifications of patient-specific (1) clinical attributes, (2) LA geometry, (3) fibrosis patterns, (4) ablation-induced scar patterns, and (5) biophysically detailed simulation results is provided in Supplementary Data 1. Feature selection by LASSO yielded the LASSO-optimized feature set (LOFS), with the total number of risk factors reduced to 27 (Table 2, Supplementary Fig. 1).

The LOFS from the original cohort was input into a random forest algorithm with 80:20 split five-fold cross validation. This algorithm was exposed to 40 recurrent AFib or atrial flutter cases and 27 non-recurrent cases. Our classifier was successful in retrospectively predicting recurrent arrhythmia (Fig. 2A; mean ROC AUC: 0.80 ± 0.04). AUC values in each data partition (i.e., ROC fold) ranged from 0.73 to 0.85. Our classifier had a balanced precision-recall score (Fig. 2B; mean PR AUC: 0.82 ± 0.08), with fold values ranging from 0.78 to 0.92. This fold of the model with the best ROC AUC is henceforth referred to as the *optimal classifier*.

Figure 3 shows the LOFS and marginal contributions of each individual feature to the resulting random forest model, extracted by SHAP analysis. The features are arranged in decreasing order of importance. Within feature rows, individual patients are represented by each data point, and color coding indicates the patient's respective feature value. The same data are shown with more detail by dependence plots (Supplementary Fig. 2), presented in the same order as Fig. 3. Post-ablation LA volume index (LAVI) was the most important feature. Patients with a high post-ablation LAVI were at risk for recurrent arrhythmia, while patients with low post-ablation LA volume indices were more likely to be arrhythmia free. Similar trends were observed for pre-ablation LA sphericity and LAVI. Fibrosis in the RPV region was associated with risk of recurrence, as was a smaller number of LGE-derived scar clusters with high surrounding fibrosis. Low burden (or absence) of post-ablation scar in the atrial floor and fewer total LGE-derived ablation regions also steered the model towards predicting recurrence. High or low scar near the RPVs also corresponded to recurrence risk, but intermediate levels of scar had the opposite relationship. In pre-ablation patient-specific simulations, the presence of two to four macro-reentrant tachycardias steered the model away from predicting recurrence, while the number of RD organizing centers had no clear effect. Lastly, many clinical attributes contributed to model predictions. Factors contributing to the model predicting AA recurrence included more direct current (DC) cardioversions, and low body surface area (BSA); RF as opposed to cryoballoon ablation also had a weak effect. Histories of congestive heart failure (CHF), stroke/transient ischemic attack (TIA), and statin use all steered the model toward predicting recurrence, whereas hyperlipidemia and use of class I/III antiarrhythmic drugs (AADs) or anticoagulants steered the model away from predicting recurrence. No clear effect on model outcome was observed

for sex, obstructive sleep apnea (OSA), diabetes diagnosis or medication use, smoking history, class II AAD use, or benign prostatic hyperplasia (BPH) medication use. All features related to medication use refer to pre-ablation patient treatment.

Relationships between feature value and impact on model output (i.e., SHAP value) are shown in Fig. 4 (most impactful cases with distinct behaviors) and Supplementary Fig. 2. Figure 4A highlights an S-shaped SHAP relationship between for post-ablation LAVI SHAP values. The maximal slope occurs at ~51 mL/m², with patients above this threshold having higher risk of recurrence. We fit an exponential decay model for the SHAP relationship with post-ablation scar on the atrial floor, indicating that patients with very little atrial floor scar often recurred, but above a certain threshold the effect of this feature had diminishing returns (Fig. 4B). We identified a positive, linear SHAP relationship with increasing number of pre-ablation DC cardioversions pressuring our model to predict recurrence (Fig. 4C). Finally, we used a moving average filter to characterize the SHAP relationship for post-ablation scar in the RPV. Patients with between 5 and 20 cm² of scar in the RPV region had a marginally decreased model-predicted recurrence risk (Fig. 4D).

We summed positive and negative SHAP values for all features to derive patient-specific arrhythmia recurrence prediction profiles. Figure 5 shows example profiles for three patients for whom the model correctly predicted AA recurrence (or the lack thereof). Each risk factor is represented by an arrow that indicates the direction in which the feature forces the model (rightward for features that favor recurrence, and vice-versa); arrow length encodes the strength of each feature’s influence. Figure 5A shows a recurrent AA prediction with model confidence for a patient who experienced post-ablation atrial fibrillation. The most important factors contributing to this

decision were elevated post-ablation LAVI and posterior wall fibrosis in the pre-ablation scan. The use of class III anti-arrhythmic drugs prior to ablation was one of a few attributes that led the model to suspect this patient would not recur, but in aggregate, the SHAP values of these features was greatly outweighed by those that suggested recurrence. Figure 5B represents the model’s prediction for a patient who was arrhythmic free for the entire 2-year follow-up; our xML algorithm correctly classified this patient based on low LAVI, no prior cardioversions, and several other features. Figure 5C highlights a case where our model correctly predicted AA recurrence despite the mitigating influence of elevated LAVI, which was outweighed by several substrate features: CHF, 10 LGE-derived scar regions, statin use, 25.5 cm² of post-ablation scar in the region of the RPVs, etc. This emphasizes that even in cases where the leading predictor of adverse outcome appears favorable, other features combine in aggregate to steer the model to the correct prediction.

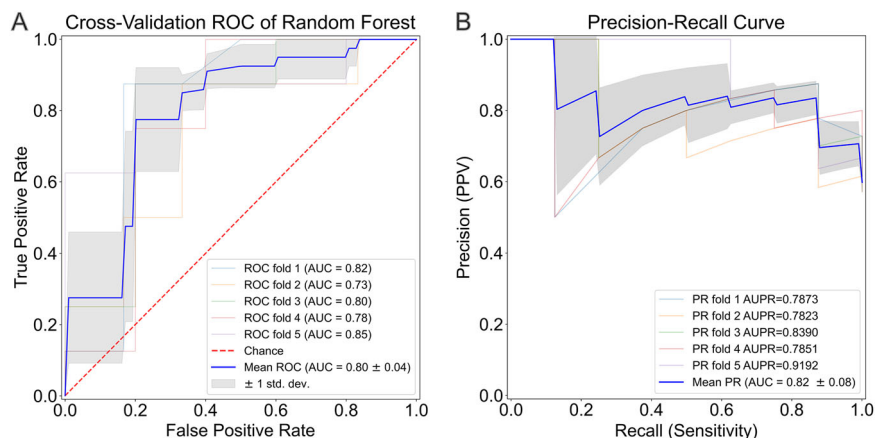
We performed a retrospective internal validation study in a 15-patient holdout cohort to quantify model performance in the most unbiased manner possible with the data available. None of these data were previously seen by the model in any iteration of training or testing. Patient characteristics of the holdout cohort did not significantly differ from the training and validation cohorts except in BMI (Table 3). The recurrence rate between the two groups was similar (59.7% vs 53.3%) and the proportion of patients receiving each ablation type was similar. The median time to reported recurrence was 305 days; follow-up was limited to two years. Only the LOFS were used for ablation outcome prediction in these cases; these data were exposed to the *optimal classifier* with no further changes to the model or parameter tuning whatsoever. Figure 6A shows the breakdown of model performance for each of the 15 patients. 12/15 (80%) patients were correctly

Table 2 | List of the 27 features comprising the LASSO-optimized feature set (LOFS)

LOFS		
Ablation type	Hyperlipidemia	Post-ablation fibrosis - RPVs
Anticoagulant use	Medication for BPH	Post-ablation LAVI
BSA	Medication for type 2 diabetes	Post-ablation scar - atrial floor
Class I AAD use	Number of direct current cardioversions	Post-ablation scar - RPVs
Class II AAD use	Number of LGE-derived scar areas with high surrounding fibrosis	Pre-ablation LA sphericity
Class III AAD use	Number of LGE-derived scar regions	Pre-ablation LAVI
CHF	Number of macro-reentrant tachycardia-like simulations	Prior stroke or TIA
Diabetes	Number of phase singularities in pre-ablation simulations	Sex
History of smoking	OSA	Statin use

AAD antiarrhythmic drug, BPH benign prostatic hyperplasia, BSA body surface area, CHF congestive heart failure, LA left atrium, LAVI left atrium volume index, LGE late gadolinium enhancement, OSA obstructive sleep apnea, RPV right pulmonary vein, TIA transient ischemic attack.

Fig. 2 | Random forest prediction algorithm efficacy in testing sets across five folds. **A** ROC curves for testing set (AUC: 0.80 ± 0.04). **B** Precision-Recall curves for testing set. (AUC: 0.82 ± 0.08).



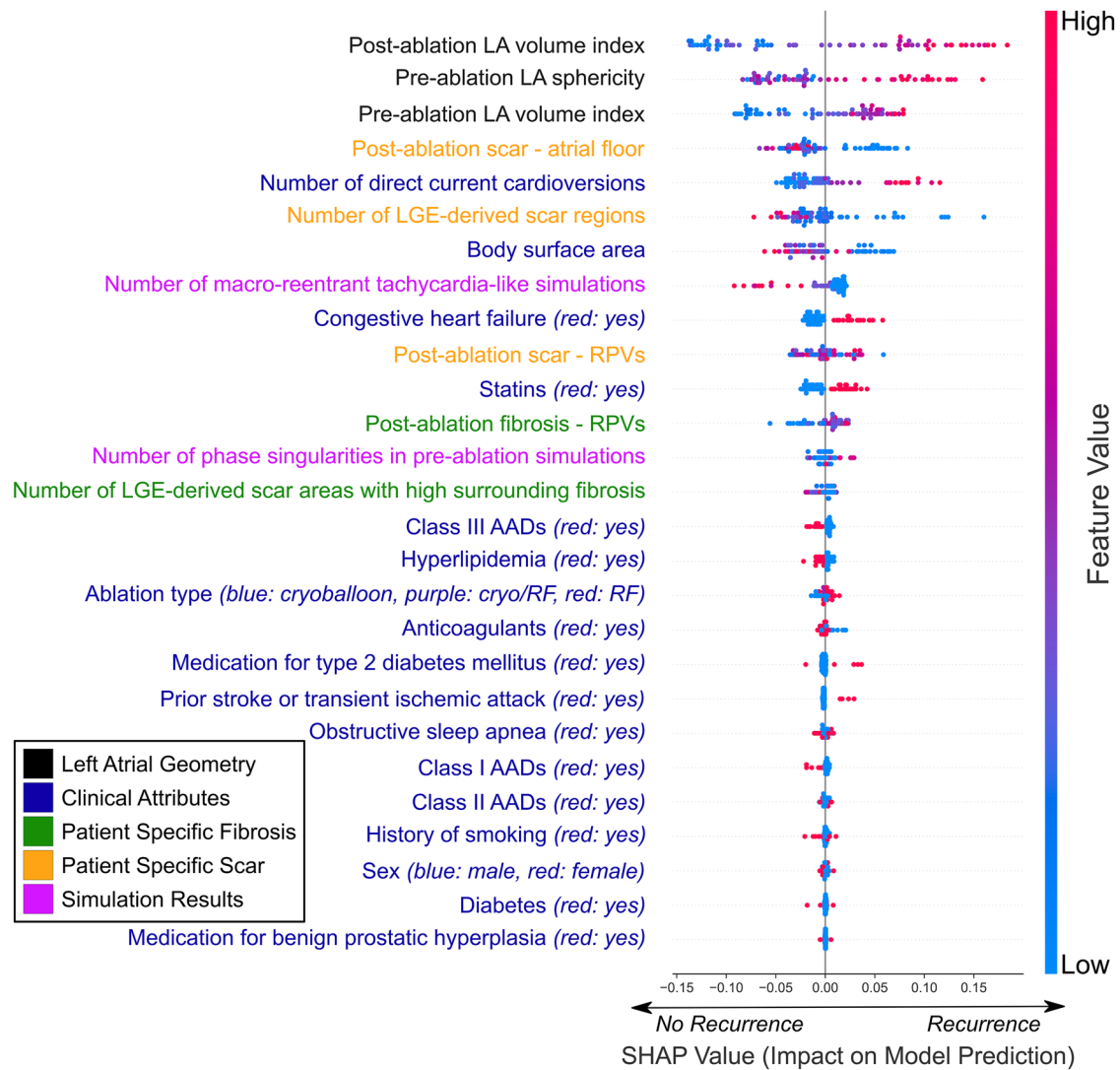


Fig. 3 | Summary of feature importance. Features are sorted in descending order of importance. Features are classified as either LA geometry (black), clinical attributes (blue), fibrosis (green), ablation-delivered scar (orange), or results from biophysically detailed simulations (magenta). Red dots indicate patients with a high feature

value for continuous variables, or presence of variable in binary cases. The vertical center line represents no impact on model outcome, while right-shifted data suggests association with recurrence and left-shifted data indicates non-recurrence. AAD antiarrhythmic drug, RPV right pulmonary veins.

classified as recurrent or non-recurrent with three false positives (20%) and no false negatives (0%). SHAP breakdowns for representative correct and incorrect predictions are shown in Fig. 6B, C, respectively. For the case where recurrence was correctly predicted, the model decision was driven by high post-ablation LAVI and LA sphericity. For the example where the model predicted recurrence but the patient remained arrhythmia-free, two features steered the model dramatically towards the incorrect conclusion: high LA sphericity and prior statin use. The model was notably near equipoise in this case, with other features forcing it nearer to the correct prediction (e.g., low post-ablation LAVI, low residual fibrosis near the RPVs).

We created complementary *alternate models* to address scientific and clinical questions. To address potential co-linearity between post- and pre-ablation LAVI (first- and third-most important features in the final model, respectively) we trained a new random forest model with all LOFS except pre-ablation LAVI (i.e., 26 features). Removing pre-ablation LAVI had no meaningful impact on performance (ROC AUC: 0.81 ± 0.07 , 80% accuracy in holdout cohort; Supplementary Fig. 3A). Second, we repeated the entire machine learning workflow (LASSO + random forest) with the change in LAVI ($\Delta\text{LAVI} = \text{post- minus pre-ablation LAVI}$) added as a distinct feature. The resulting models performed poorly. Across 200 LASSO attempts, model

accuracy on the holdout cohort ranged from 40 to 80% (median: 60%). ΔLAVI was selected as a feature in only one single LASSO attempt; the resulting random forest had a favorable ROC AUC (0.87 ± 0.10), but reduced holdout accuracy (67%), suggesting model overfitting (Supplementary Fig. 3B). Finally, replacing pre-ablation LAVI in the LOFS with ΔLAVI resulted in slightly improved model performance (ROC AUC: 0.85 ± 0.03 , 80% holdout accuracy; Fig. 7A). In the context of this alternate model, only optimal LAVI reduction (from 0 to -20 mL/m^2) steered the algorithm away from predicting recurrence (Fig. 7B). In all alternate models considered, LA geometric features remained important drivers of model predictions.

Clinically oriented alternate models were also trained and tested. Considering the potential utility of predicting AA recurrence prior to ablation, three post-ablation features were dropped from the LOFS, and post-ablation fibrosis in the region of the RPVs was replaced with its pre-ablation equivalent. This model had reduced performance (ROC AUC: 0.75 ± 0.07 , 67% holdout accuracy; Supplementary Fig. 3C). Due to limited access to the computational facilities required for biophysically detailed LA electrophysiology modeling, we created an alternate model in which the two simulation-derived features were removed at the random forest training and testing phase. ROC AUC remained high, 0.81 ± 0.09 , but holdout accuracy

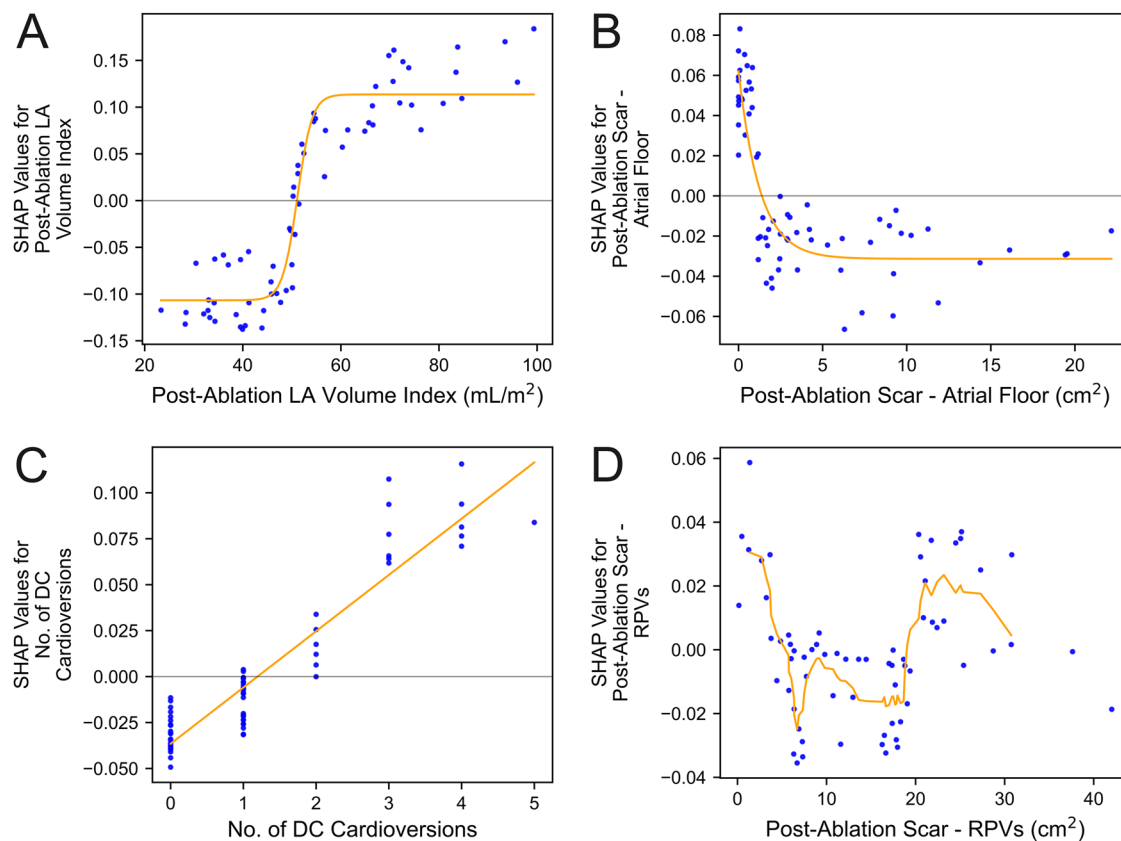


Fig. 4 | Dependence plots of SHAP importance vs. raw feature values. SHAP values greater than zero correlate to increased risk of recurrence, and those less than 0 correlate with reduced recurrence risk. **A** Dependence plot for post-ablation LA volume index (LAVI), demonstrating increased recurrence in patients with a post-ablation LAVI above 51.121 mL/m². $y = 0.220/(1+e^{-(0.668*(x-51.121))})-0.107$. **B** Dependence plot for post-ablation scar in the atrial floor region, showing less scar on the atrial floor was associated with increased recurrence.

$y = 0.0934*e^{(-0.788x)}-0.031$. **C** Dependence plot for number of direct current (DC) cardioversions, indicating recurrence was more likely in patients who received more DC cardioversions. $y = 0.0306x-0.0366$; $R^2 = 0.8699$. **D** Dependence plot for post-ablation scar in the region of the RVPs, fit with a 7-period moving average trendline, suggesting the random forest model identified a complex relationship between scar in the region of the RVPs and recurrence risk.

decreased to 73% (one additional patient misclassified; Supplementary Fig. 3D).

Lastly, we considered incorporating features derived from f-wave analysis. Since fewer patients were eligible for inclusion (see Methods), we performed a completely new model creation process. Alongside the LOFS, we supplied f-wave amplitude and dominant frequency in lead I for 41 of 67 patients from the train/test cohort and 6 of 15 patients from the holdout cohort (Supplementary Fig. 4A). The ROC AUC of the new model was 0.76 ± 0.14 and holdout accuracy was 4 of 6 (66%) (Supplementary Fig. 4B). Direct comparison of this model to those discussed in prior sections is ill-advised due to the difference in cohort sizes. SHAP analysis for this version of the model including f-wave features suggested lower amplitude may correspond to increased recurrence risk; the influence of dominant frequency was unclear (Supplementary Fig. 4C, D).

To compare our random forest model against a linear model, we fit a logistic regression model to the LOFS. The mean ROC AUC of the logistic regression model across five-fold cross-validation was slightly lower at 0.77 ± 0.11 (Supplementary Fig. 5A). The logistic regression model also performed slightly worse in the holdout cohort, misclassifying 4/15 patients (Supplementary Fig. 5B). Accordingly, the logistic regression model's F1-score of 0.78 was worse than the F1-score for the random forest model.

Discussion

We designed an xML algorithm to integrate multi-modal data for prediction of recurrent arrhythmias following catheter ablation of AFib. Compared to

prior studies^{21,22,24,57}, we achieved similar performance (ROC AUC of 0.80 ± 0.04 versus 0.61-0.85) in predicting recurrent arrhythmia risk, with the added benefit of interpretable explanations. When tested on a previously unseen holdout cohort, the model maintained 80% accuracy and an F1-score of 0.84 (Fig. 6). This exceeded the performance of a comparable logistic regression model, which was 73% accurate with an F1-score of 0.78 when tested on the same cohort (Supplementary Fig. 5B). The combined accuracy and interpretability of models like ours will allow electrophysiologists to receive optimized prediction scores while also gaining scientific insight into why those predictions were made. We present three distinct data perspectives: population-based (Fig. 3), risk factor-based (Fig. 4), and patient-specific (Fig. 5). We also explore potential clinical application of our work and assess for overfitting by testing our model on a holdout cohort, previously unseen at any stage of training or testing (Fig. 6).

To minimize the influence of confounders on our algorithm, we used rigorous feature selection. Definitionally, observed confounders are highly correlated features, while unobserved confounders are features that are unmeasurable or unaccounted for. To mitigate effects from observed confounders and reduce the dimensionality of the data supplied to our random forest model, we used LASSO regression (Supplementary Fig. 1). This ensured independence of feature information, which improves generalizability by eliminating redundant variables. Unobserved confounders are a central barrier to drawing causal inferences from observational data and introduce bias that can be difficult to avoid. To address this, we prioritized features that have been previously examined in the literature for their impact on procedure outcome (Supplementary Data 1).

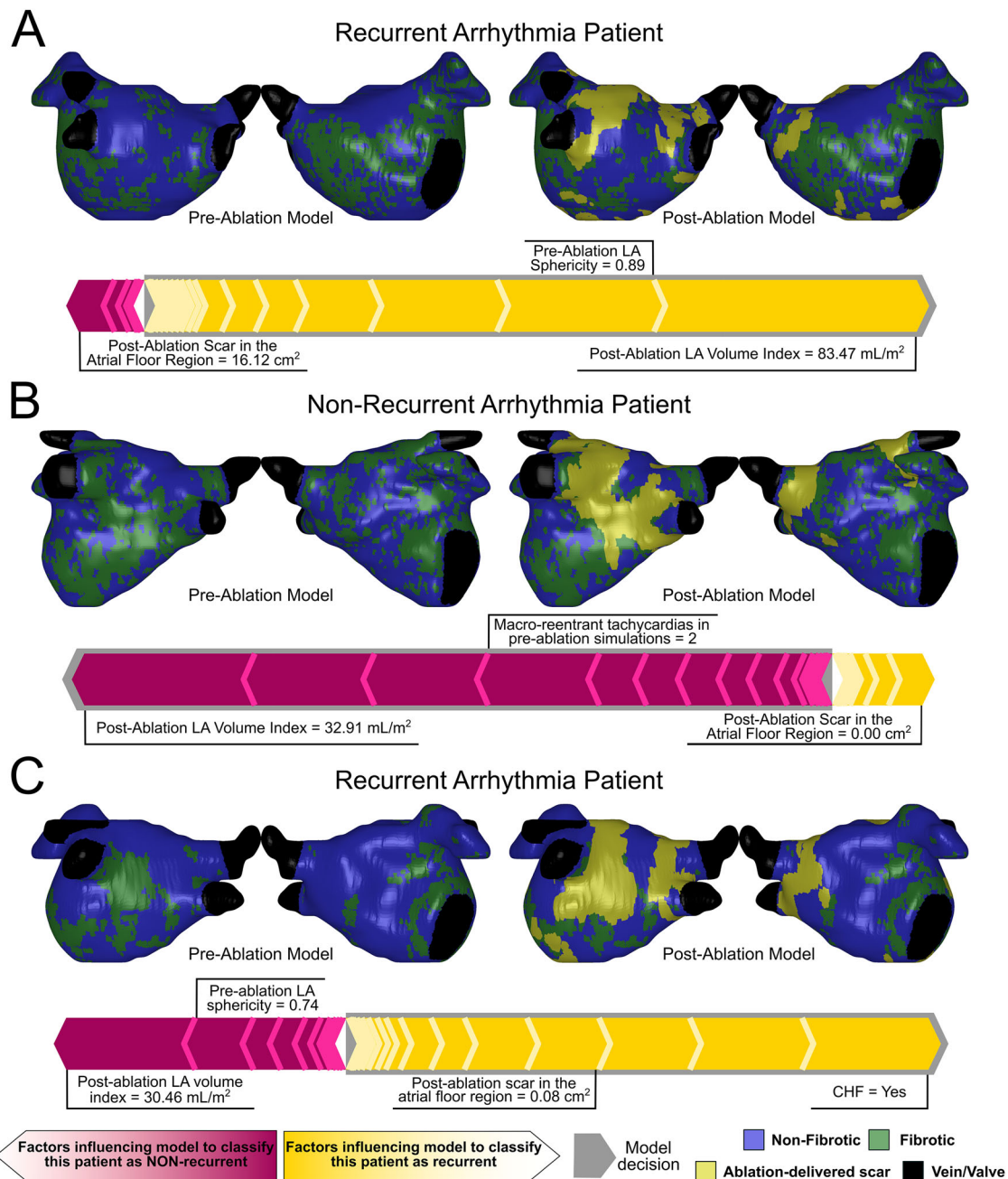


Fig. 5 | Explanation of individual patient prediction scores for representative patients in the original dataset. Yellow and magenta arrows indicate the marginal contribution of each feature, and the grey outline indicates the model’s prediction. Black indicator lines provide more detailed information about the features corresponding to the arrows they intersect. **A** Visualization of pre- and post-ablation models and explanation of a correct recurrent arrhythmia prediction driven by

geometry and substrate features. This patient recurred with atrial fibrillation (AFib). **B** Pre- and post-ablation models of patient and explanation of a correct non-recurrent arrhythmia prediction driven by geometry and clinical features. **C** Visualization of pre- and post-ablation patient-specific models and explanation of a correct non-recurrent arrhythmia prediction driven by pre- and post-ablation substrate features. This patient also recurred with AFib.

Random forest learning is a versatile algorithm and was specifically chosen for this problem because feature values are not altered during the learning process. This approach can outperform linear models by identifying nonlinear relationships between features (Supplementary Fig. 5). However, if we had used other machine learning methods like support vector machines, we would have lost the ability to observe how specific feature values impact model outcomes in a reliable and confident manner. This is a key aspect of our xML approach because risk factor quantifications like those presented in Fig. 4 would be difficult to obtain from other machine learning methods.

When applying SHAP analysis to understand feature importance from a population perspective, we identified important influences from LA geometry and changes in LA geometry on AA recurrence. Our model gained important insight from the indexed post-ablation LA volume, its change after the ablation procedure, and LA sphericity. This is consistent with prior work indicating that mild reverse remodeling of the LA associated with ablation may predict long-term ablation success, especially in patients with lower baseline fibrosis⁵⁸. While the work presented here does not (and cannot) prove causal relationships between features and outcomes, LA dilation is known to play a role in recurrence⁵⁹ and is independently

Table 3 | Patient characteristics in original (training and testing) and holdout (internal validation) cohorts

	Original Cohort	Holdout Cohort	p-value
N	67	15	
Recurrence	59.7%	53.3%	0.658
Age (years)	62 (56–70)	70 (62–76)	0.1215
BMI	30.6 (26.4–35.8)	26.3 (23.1–31.7)	0.0353
Sex (male)	58.2%	40.0%	0.2054
Hypertension	58.2%	46.6%	0.4224
Hyperlipidemia	38.8%	60.0%	0.1378
Coronary artery disease	17.9%	13.3%	0.6788
OSA	34.3%	40.0%	0.6851
Stroke/TIA	7.5%	6.7%	0.9259
Diabetes	9.0%	6.7%	0.7852
COPD	6.0%	0%	0.343
Cigarette use	17.9%	6.7%	0.2884
Thyroid disease	13.4%	0%	0.1378
CHF	28.4%	6.7%	0.08021
Pre-Ablation LA LGE (%)	16.7% (10.95–21.35%)	15.9% (10.6–20.6%)	0.6229
AFib ablation type:			
Cryo PVI only	38.8%	40.0%	0.9378
Cryo PVI + RF SM	4.5%	6.7%	0.7357
RF PVI only	14.9%	26.7%	0.2817
RF PVI + SM	41.8%	26.7%	0.2838

Variables are reported as median (IQR) and compared between cohorts with Mann-Whitney U tests as appropriate. For binary variables, proportions were reported and compared with χ^2 tests.

associated with AA-free survival⁶⁰. Previous machine learning models achieved an ROC AUC of 0.67 when predicting AFib recurrence using LA shape metrics from pre-ablation CT⁶¹. Post-ablation atrial enlargement is also associated with adverse long-term ablation outcomes independent of left ventricular function⁶². Mechanistically, LA enlargement promotes AFib directly (larger physical area for rotor perpetuation⁶³) and indirectly (via properties like atrial stretch⁶⁴). In our study, higher LA volume indices (>51 mL/m²) or a more spherical pre-ablation LA (>0.81) were associated with recurrence.

Our model suggests the creation of fewer distinct ablation-induced scar areas and less scar in the LA floor (Fig. 4B) is associated with elevated AA recurrence risk. Poor scar formation during ablation is indeed associated with AFib recurrence²⁶. Notably, surface area of ablation-delivered scar and percentage of scar with respect to total LA size were included in our study but eliminated during feature selection. This suggests the number of ablation-induced scar areas is a more robust predictor. We also note the complicated relationship between scar in the RPV region and AA recurrence. While more data are needed, there may be an optimal ablation extent between ~5.0 cm² and 20.0 cm² in this area, reducing risk of AA recurrence (Fig. 4D). This likely approximates the extent of RPV tissue typically ablated during PVI. Future work could investigate if the creation of many independent scar areas during ablation might yield favorable outcomes; follow-up studies could be used to validate the number of LGE-derived scar areas created, especially in the LA floor and RPV regions, as a predictor of AA recurrence.

Many studies have investigated the influence of patient characteristics like sex and age on risk of AA recurrence after ablation. DECAAF I found slightly higher (not statistically significant) incidence of AA recurrence in females than males⁶⁵; other work suggests higher recurrence rates in females is attributable to extra-PV triggers in females⁶⁶. CABANA indicated this may be confounded by lower referral rates for women, compounded by the fact that that women tend to be referred later in AFib progression^{67,68}. Like more recent studies⁶⁹, our model did not indicate recurrence was more likely for males or females (Fig. 3, Supplementary Fig. 2Y). Similarly, the relationship between age and AA recurrence remains unclear. CABANA showed no age-related variations in ablation effectiveness⁷⁰. In our dataset,

A Internal Testing (Holdout) Cohort

Patient Number	Recurrence Rhythm	Model Prediction (Confidence)
11	Atrial Fibrillation	Arrhythmia (0.35)
4	Atrial Flutter	Arrhythmia (0.35)
15 *	None	Arrhythmia (0.31)
5	Atrial Fibrillation	Arrhythmia (0.25)
14	Atrial Flutter	Arrhythmia (0.21)
8	Atrial Fibrillation	Arrhythmia (0.20)
12	Atrial Fibrillation	Arrhythmia (0.18)
13 *	None	Arrhythmia (0.13)
7	Atrial Fibrillation	Arrhythmia (0.08)
9	Atrial Fibrillation	Arrhythmia (0.06)
2 *	None	Arrhythmia (0.02)
10	None	No Arrhythmia (-0.11)
6	None	No Arrhythmia (-0.13)
1	None	No Arrhythmia (-0.20)
3	None	No Arrhythmia (-0.21)

-0.4 (No Recurrence) Confidence (Recurrence) 0.4
* Incorrect Model Prediction

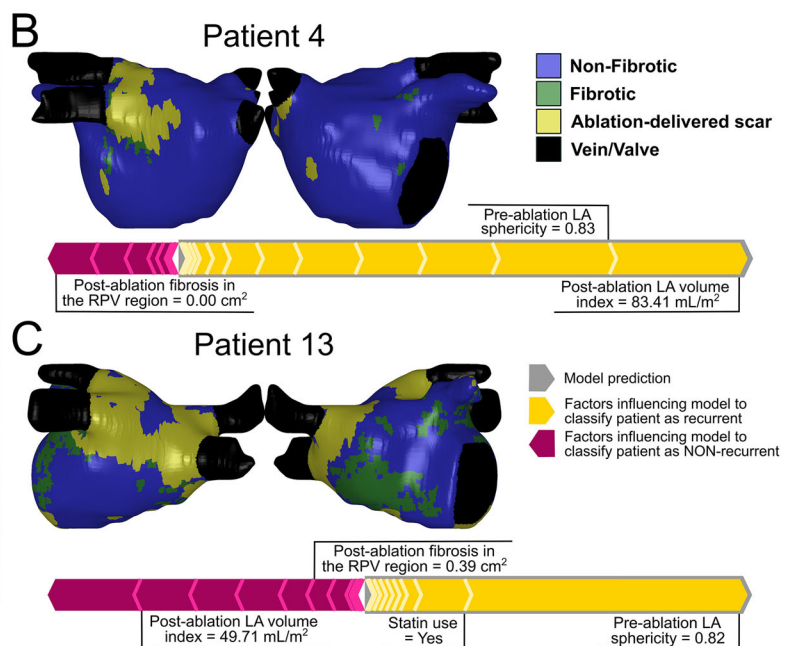
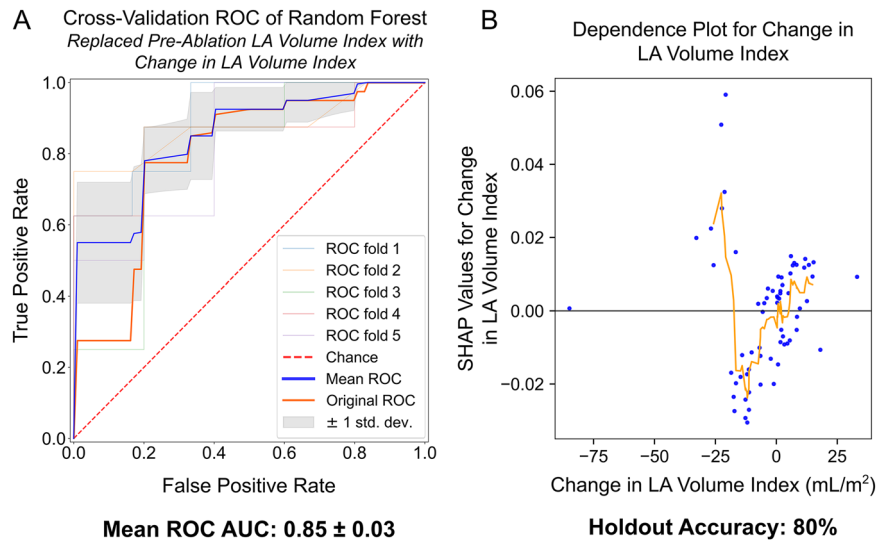


Fig. 6 | Summary of model performance on a previously unseen internal validation (holdout) cohort. **A** Chart of each patient outcome and model prediction in the 15-patient cohort. Gradient indicates relative confidence (summed SHAP values offset by the model’s decision threshold of 0.10) of non-recurrence and recurrence,

respectively, calculated as the sum of SHAP values. **B** Example for correct prediction with post-ablation patient-specific anatomical model and prediction breakdown via SHAP analysis. **C** Example for incorrect positive prediction.

Fig. 7 | Performance of a model trained with the change in LA volume index (LAVI) in place of the pre-ablation LAVI. **A** ROC curve from five-fold training and cross-validation of the model, showing a mean ROC AUC of 0.85 ± 0.03 , with individual fold ROC AUCs ranging from 0.80 to 0.90. The model retained its 80% accuracy on the holdout set. **B** Dependence plot showing the relationship between the change in LAVI and the marginal contribution to model outcome.



LASSO regression did not select age as a potentially valuable feature. These results support the notion that factors derived from imaging studies are superior to demographic metrics for evaluating recurrence risk.

The ability of xML models to incorporate and explain effects of imaging data is appealing. DECAAF-II found no difference in AA recurrence for patients randomized to receive MRI-guided ablation compared to those receiving PVI alone⁷¹. It is thus noteworthy that our model had reduced performance when supplied with pre-ablation data only (Supplementary Fig. 3C). The application of explainability tools like SHAP offers valuable insights on the clinical challenge of predicting AA recurrence. In our proof-of-concept study, we reveal LA changes corresponding to durable ablation success. Specifically, we identify the optimal extent of reverse remodeling associated with freedom from AA (Fig. 7) and we characterize the extent and spatial distribution of durable ablation-delivered scar associated with successful procedures (Fig. 3). While Δ LAVI itself was not considered useful by LASSO regression, we suspect this was due to LASSO regression’s bias favoring features that linearly relate to outcome, whereas the relationship between Δ LAVI and outcome was highly nonlinear (Fig. 7B). Future studies with larger sample sizes may opt to prospectively include Δ LAVI to reinforce these findings. We also show the utility of pre-ablation computational simulations built from LGE-MRI imaging data (Fig. 3). Future work in larger, prospective studies may use these findings to determine ablation strategies that reduce overall population-level AA recurrence.

An advantage of xML is that it facilitates post-hoc model tuning to remove unintended sources of overfitting, which is especially relevant when working with small cohorts. For instance, in our model the SHAP value corresponding to patients who smoke corresponds to a slight lean towards predicting non-recurrence. In contrast, research consistently indicates that smoking worsens ablation outcomes^{8,72}. We can thus conclude that the SHAP value for the few patients in our original cohort with a smoking history (17.9%) is likely due to overfitting. Fault diagnosis and identification of improperly weighted variables is a key xML feature compared to conventional “black box” ML⁷³. The example described above emphasizes how specific features can be identified for removal to improve model accuracy in future work with minimal time spent problem-solving. We caution that feature importance assessed by SHAP analysis does not translate to clinical risk. Due to the individual and collective contributions of each feature to the model it would not be sufficient to derive a simplified model considering only the most important features indicated by SHAP analysis without repeating the model development and validation process.

Precision medicine is an emerging approach that integrates multimodal data to customize treatment and further disease understanding. Using SHAP analysis, we can generate recurrence predictions and feature importance breakdowns to create personalized, data-rich risk snapshots. In

some cases, features combine synergistically to produce a high confidence prediction; more often, the model identifies multiple key features that contribute for and against a prediction, leading to a multi-faceted decision-making process. Since feature relationships shown in this study are not causal, it must be clarified that modifying these features (e.g., avoiding statins for a patient whose statin use coaxes the model towards predicting recurrence) may not improve health outcomes. Nonetheless, xML provides a platform for further research and validation that could help guide the care and advising of post-ablation patients.

Interestingly, key decision drivers for specific individuals often diverge from the most important population-based features. For example, if the low post-ablation LAVI for the patient in Fig. 5C was considered in isolation, it might lead to an incorrect conclusion that this individual was at low risk for recurrence. In effect, xML enables holistic assessment of each patient’s risk factors, informed by but not adhering strictly to population trends. The prediction “breakdowns” shown in our study are a promising tool for medical professionals and potentially the patients themselves, allowing them to assess relative risk factor impacts and make more data-informed decisions, agnostic to population trends that may not apply.

The role of biophysically detailed simulation-derived features in the model was interesting. LASSO regression only selected two such features to include in the model: the number of macro-reentrant tachycardia circuits in pre-ablation simulations, and the number of phase singularities (i.e., organizing centers) in pre-ablation simulations (Fig. 3, Supplementary Fig. 2H, M). Patients with more macro-reentrant tachycardia circuits in pre-ablation simulations generally had lower risk of recurrence. This is mechanistically interesting as it suggests patients with pre-ablation fibrotic substrate susceptible to anatomic reentry (as opposed to functional reentrant spiral waves) more commonly experienced durable benefit from AFib ablation. The effect of the number of phase singularities was unclear, suggesting the influence of this feature was highly dependent on other feature values.

We also incorporated f-wave analysis in a variant of our model, but limited data availability prevents a meaningful interpretation of the resulting model. Consistent with others’ findings^{17,19}, our recurrence was more common in patients with low f-wave amplitude (Supplementary Fig. 4C); low power likely prevented the model from finding a clear relationship between recurrence and dominant frequency (Supplementary Fig. 4D). This sub-analysis supports the incorporation of ECG-derived features in future efforts to stratify recurrence risk.

Our use of a holdout cohort aims to simulate how our model could work in a clinical environment and assess this method’s generalizability in a broader population. The model was blind to these data throughout its training and cross-validation. Differences existed between these two cohorts (e.g., BMI; Table 3), as we would expect if the model were applied to a new

patient population. Figure 6C highlights a false positive prediction in which we could visually troubleshoot the model's incorrect decision. In this case, the model may have over-weighted the relevance of LA sphericity and the history of statin use or under-weighted the influence of the post-ablation LAVI fibrosis near the RPVs. Additionally, a healthcare provider could consider the broader clinical context, such as fibrosis in non-RPV areas, and confidently weigh the model's prediction against their own clinical judgement. As the field evolves and we learn more about independent associations between features like these and adverse outcomes, our xML approach provides a ready-made scaffolding for incorporating that new knowledge in clinical decision-making. Overall, we expect model-agnostic explanations of feature weights from xML algorithms will increase physician understanding and confidence in predictive models²⁵.

Due to the nature of "black box" ML algorithms, it has been difficult to understand how their predictions relate to physiological mechanisms of AA recurrence. There is risk associated with clinical use of these complex models since predictions are difficult to interpret and might thus lead to physician frustration or distrust. It is challenging to decide which predictions should be seen as actionable and which are safe to ignore. Our work shows how xML could pave the way towards solutions that reduce obscurity by prioritizing interpretability, transforming the current paradigm by creating models that can coherently explain why a particular prediction is made.

Our algorithm shows proof-of-concept for several clinical applications: (1) repeat ablation planning, (2) post-ablation care, (3) patient-clinician communication, and (4) future hypothesis-driven research. Based on key fibrosis and ablation-delivered scar features identified for each patient, selection of ideal candidates for redo procedures and identification of sites for optimal targeted re-ablation sites could be feasible. Following an ablation procedure and follow-up MRI scan, healthcare teams could use xML-based predictions to make informed decisions about monitoring for recurrence or adjust medications. We also believe xML has great potential to foster patient-clinician communication, facilitating better understanding of risk factors and their associations with outcomes in a digestible, visual format. Finally, while we have emphasized that associations between feature values and outcomes in this study do not imply causality, we have identified potential drivers of recurrence that could be validated in hypothesis-driven research.

Our work should be seen as a promising preliminary exploration of xML-based prediction of catheter ablation recurrence. Our sample size was modest, and all patients were from a single center. While measures were taken to ensure proper feature selection and model design to limit overfitting, a larger cohort is needed to confirm these findings. Notably, it is essential that any future study seeking to expand upon our work in a different or larger cohort must repeat our entire methodology, including both feature selection and ML model development stages – in other words, it would be inappropriate to use the LOFS from this study as the starting point for new work in a distinct dataset. Potential future studies with larger cohorts could explore more extensive model optimization techniques to further enhance model performance. It is also important that our approach be subjected to further testing using an external holdout cohort (i.e., using data from non-UW patients) to ensure its generalizability to data from other institutions. Additionally, patients presenting to their pre-ablation MRI in AFib in our study were not cardioverted into sinus rhythm before their pre-ablation MRI, nor were these patients assigned to a different cohort, introducing a potential source of confounding. A larger, more diverse, and prospective cohort would give us the flexibility of designing a model that can handle more features, opening the door to using other clinical data like ECGs or electroanatomic maps. We anticipate that this would enrich our algorithm, potentially improving accuracy. Like other clinical trials, distinct models could also be considered for patients with persistent or paroxysmal AFib, given the differing recurrence rates between these populations^{74,75}. Prospective trials of this classifier and randomized trials assessing ML-identified features to verify causal relationships would also be valuable steps prior to clinical deployment.

Other research has suggested that simulations meaningfully improve the ability of machine learning models to predict recurrent arrhythmias^{21,22}. In our study, LASSO regression selected features from pre-ablation simulations to supply to the random forest model, but did not select features from post-ablation simulations. Additionally, when simulation results were removed from the model, its performance only slightly decreased (Supplementary Fig. 3D). We attribute this to our extensive analysis of LGE-MRI, which added post-ablation scans and assessed a larger gamut of fibrosis-derived and ablation-scar derived features. A complementary reason is that prior studies performed more extensive simulations, with changes in fibrosis representation and features selected by deductive algorithms. Simulations are a valuable platform for mechanistic inquiry and custom-tailored ablation planning^{76,77}, but calculation of the associated features is computationally complex and requires infrastructure inaccessible in many clinical settings. As such, since our goal was proof-of-concept for using xML to predict recurrence in a clinically feasible way, we are pleased our model performed well with or without simulation features.

Notably, the recurrence rate in the cohort studied here was high (59.7%) compared to other contemporary studies examining mixed groups of paroxysmal and persistent AFib patients (e.g., 49.9% AFib recurrence rate in ablation arm of CABANA trial⁶⁸). This may be a result of selection bias, since patients who have already recurred or who are deemed by clinicians to be at higher risk for recurrence are more likely to be scheduled for post-ablation LGE-MRI scan.

Conclusions

We developed an xML classifier to accurately predict arrhythmia recurrence following catheter ablation from EHR data alongside pre- and post-ablation LGE-MRI scans. Critically, our classifier only uses clinical data obtained non-invasively, with features guided by existing knowledge of arrhythmia recurrence mechanisms. We envision the coupling of our ML model with an explainability technique as a framework for using ML-enabled clinical tools to strengthen the engagement of clinicians and stakeholders in informed and shared decision-making. In addition to presenting a predictive solution that avoids obscurity by favoring model interpretation, we present novel mechanistic hypotheses generated from critically evaluating the influence of each feature on our ML model, which was developed without enforcing a priori hypotheses about the nature of relationships between these features and risk of AF recurrence. Future work should explore larger datasets, including external holdout cohorts, to confirm our findings, further optimize feature selection, improve model accuracy, and investigate potential mechanistic relationships uncovered by our analysis.

Data availability

All 164 left atrial anatomical models used in this study have been made available to the public via the following permanent link <https://doi.org/10.5061/dryad.kkwh70sg0>⁵⁵. This dataset comprises two models per patient (pre-ablation with patterns of fibrotic remodeling, post-ablation with scar created by the procedure) for 82 individuals. Supplementary Data 2 contains the source data underlying Figs. 3–4 and Supplementary Fig. 2. Supplementary Data 3 contains the source data underlying Supplementary Figs. 4C–D. Supplementary Data 4 contains the source data underlying Fig. 7B. To avoid the possibility of patient identification, all source data have been disaggregated such that adjacent feature values and SHAP values are correctly linked but individual rows of tabular data do not necessarily correspond to features from the same individual; instead, each pair of feature/SHAP value columns is sorted by feature value. Other data related to the article will be shared with interested parties for non-commercial reuse on reasonable request to the co-corresponding authors and approval by the UW IRB.

Code availability

Code used for ML model training, validation, and explanation is available at <https://doi.org/10.5061/dryad.4tmpg4fp9>⁵⁶. See the associated README for details on using this code.

Abbreviations

AA	atrial arrhythmia
AAD	antiarrhythmic drug
AFib	atrial fibrillation
AUC	area under the curve
BPH	benign prostatic hyperplasia
BSA	body surface area
CHF	congestive heart failure
Cryo	cryoballoon
DC	direct current
ECG	electrocardiogram
EHR	electronic health record
f-wave	fibrillatory wave
FD	fibrosis density
FE	fibrosis entropy
IRB	institutional review board
LA	left atrium
LASSO	least absolute shrinkage and selection operator
LAVI	LA volume index
LGE	late gadolinium enhanced
LOFS	LASSO-optimized feature set
LPVs	left pulmonary veins
MRI	magnetic resonance imaging
OSA	obstructive sleep apnea
PR	precision-recall
PVI	pulmonary vein isolation
RD	reentrant driver
RF	radio frequency
ROC	receiver operating characteristic
RPVs	right pulmonary veins
SHAP	SHapley Additive exPlanations
TIA	transient ischemic attack
UW	University of Washington
xML	explainable machine learning

Received: 10 November 2023; Accepted: 21 July 2025;
Published online: 14 October 2025

References

- Andrade, J., Khairy, P., Dobrev, D. & Nattel, S. The clinical profile and pathophysiology of atrial fibrillation: relationships among clinical features, epidemiology, and mechanisms. *Circ. Res.* **114**, 1453–1468 (2014).
- Kobza, R. et al. Late recurrent arrhythmias after ablation of atrial fibrillation: incidence, mechanisms, and treatment. *Heart Rhythm* **1**, 676–683 (2004).
- Vizzardi, E. et al. Risk factors for atrial fibrillation recurrence: a literature review. *J. Cardiovasc. Med.* **15**, 235–253 (2014).
- Santoro, F. et al. Impact of uncontrolled hypertension on atrial fibrillation ablation outcome. *JACC Clin. Electrophysiol.* **1**, 164–173 (2015).
- Wang, T. J. et al. Obesity and the risk of new-onset atrial fibrillation. *JAMA* **292**, 2471–2477 (2004).
- Guckel, D. et al. The effect of diabetes mellitus on the recurrence of atrial fibrillation after ablation. *J. Clin. Med.* **10** <https://doi.org/10.3390/jcm10214863> (2021).
- Buckley, B. J. R. et al. Atrial fibrillation in patients with cardiomyopathy: prevalence and clinical outcomes from real-world data. *J. Am. Heart Assoc.* **10**, e021970 (2021).
- Cheng, W. H. et al. Cigarette smoking causes a worse long-term outcome in persistent atrial fibrillation following catheter ablation. *J. Cardiovasc. Electrophysiol.* **29**, 699–706 (2018).
- Macheret, F. et al. Comparing inducibility of re-entrant arrhythmia in patient-specific computational models to clinical atrial fibrillation phenotypes. *JACC Clin. Electrophysiol.* **9**, 2149–2162 (2023).
- Zahid, S. et al. Patient-derived models link re-entrant driver localization in atrial fibrillation to fibrosis spatial pattern. *Cardiovasc. Res.* **110**, 443–454 (2016).
- Ali, R. L. et al. Arrhythmogenic propensity of the fibrotic substrate after atrial fibrillation ablation: a longitudinal study using magnetic resonance imaging-based atrial models. *Cardiovasc. Res.* **115**, 1757–1765 (2019).
- Hakim, J. B., Murphy, M. J., Trayanova, N. A. & Boyle, P. M. Arrhythmia dynamics in computational models of the atria following virtual ablation of re-entrant drivers. *Europace* **20**, iii45–iii54 (2018).
- Bifulco, S. F., Macheret, F., Scott, G. D., Akoum, N. & Boyle, P. M. Explainable machine learning to predict anchored reentry substrate created by persistent atrial fibrillation ablation in computational models. *J. Am. Heart Assoc.* **12**, e030500 (2023).
- Bieging, E. T. et al. Left atrial shape predicts recurrence after atrial fibrillation catheter ablation. *J. Cardiovasc. Electrophysiol.* **29**, 966–972 (2018).
- Jia, S. et al. Left atrial shape is independent predictor of arrhythmia recurrence after catheter ablation for atrial fibrillation: a shape statistics study. *Heart Rhythm O2* **2**, 622–632 (2021).
- Haissaguerre, M. et al. Atrial fibrillatory cycle length: computer simulation and potential clinical importance. *Europace* **9**, vi64–vi70 (2007).
- Nault, I. et al. Clinical value of fibrillatory wave amplitude on surface ECG in patients with persistent atrial fibrillation. *J. Inter. Card. Electrophysiol.* **26**, 11–19 (2009).
- Matsuo, S. et al. Clinical predictors of termination and clinical outcome of catheter ablation for persistent atrial fibrillation. *J. Am. Coll. Cardiol.* **54**, 788–795 (2009).
- Lankveld, T. et al. Atrial fibrillation complexity parameters derived from surface ECGs predict procedural outcome and long-term follow-up of stepwise catheter ablation for atrial fibrillation. *Circ. Arrhythm. Electrophysiol.* **9**, e003354 (2016).
- Di Marco, L. Y., Raine, D., Bourke, J. P. & Langley, P. Characteristics of atrial fibrillation cycle length predict restoration of sinus rhythm by catheter ablation. *Heart Rhythm* **10**, 1303–1310 (2013).
- Shade, J. K. et al. Preprocedure application of machine learning and mechanistic simulations predicts likelihood of paroxysmal atrial fibrillation recurrence following pulmonary vein isolation. *Circ. Arrhythm. Electrophysiol.* **13**, e008213 (2020).
- Roney, C. H. et al. Predicting atrial fibrillation recurrence by combining population data and virtual cohorts of patient-specific left atrial models. *Circ. Arrhythm. Electrophysiol.* **15**, e010253 (2022).
- Kim, J. Y. et al. A deep learning model to predict recurrence of atrial fibrillation after pulmonary vein isolation. *Int. J. Arrhythm.* **21** <https://doi.org/10.1186/s42444-020-00027-3> (2020).
- Razeghi, O. et al. Atrial fibrillation ablation outcome prediction with a machine learning fusion framework incorporating cardiac computed tomography. *J. Cardiovasc. Electrophysiol.* **34**, 1164–1174 (2023).
- Diprose, W. K. et al. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. *J. Am. Med. Inf. Assoc.* **27**, 592–600 (2020).
- Parmar, B. R. et al. Poor scar formation after ablation is associated with atrial fibrillation recurrence. *J. Inter. Card. Electrophysiol.* **44**, 247–256 (2015).
- Tutuianu, C., Szilagyi, J., Pap, R. & Saghy, L. Very long-term results of atrial fibrillation ablation confirm that this therapy is really effective. *J. Atr. Fibrillation* **8**, 1226 (2015).
- Calkins, H. et al. 2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation. *Heart Rhythm* **14**, e275–e444 (2017).

29. Siebermair, J., Kholmovski, E. G. & Marrouche, N. Assessment of left atrial fibrosis by late gadolinium enhancement magnetic resonance imaging. *Methodol. Clin. Implic. JACC Clin. Electrophysiol.* **3**, 791–802 (2017).
30. January, C. T. et al. 2019 AHA/ACC/HRS focused update of the 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. *J. Am. Coll. Cardiol.* **74**, 104–132 (2019).
31. Kirchhof, P. et al. Outcome parameters for trials in atrial fibrillation: recommendations from a consensus conference organized by the German Atrial Fibrillation Competence NETwork and the European Heart Rhythm Association. *Europace* **9**, 1006–1023 (2007).
32. Jadidi, A. S. et al. Inverse relationship between fractionated electrograms and atrial fibrosis in persistent atrial fibrillation: combined magnetic resonance imaging and high-density mapping. *J. Am. Coll. Cardiol.* **62**, 802–812 (2013).
33. Akoum, N. et al. MRI Assessment of Ablation-Induced Scarring in Atrial Fibrillation: Analysis from the DECAAF Study. *J. Cardiovasc. Electrophysiol.* **26**, 473–480 (2015).
34. Bifulco, S. F. et al. Computational modeling identifies embolic stroke of undetermined source patients with potential arrhythmic substrate. *Elife* **10**, <https://doi.org/10.7554/eLife.64213> (2021).
35. Roney, C. H. et al. Universal atrial coordinates applied to visualisation, registration and construction of patient specific meshes. *Med Image Anal.* **55**, 65–75 (2019).
36. Cochet, H. et al. Relationship between fibrosis detected on late gadolinium-enhanced cardiac magnetic resonance and re-entrant activity assessed with electrocardiographic imaging in human persistent atrial fibrillation. *JACC Clin. Electrophysiol.* **4**, 17–29 (2018).
37. Sakata, K. et al. Assessing the arrhythmogenic propensity of fibrotic substrate using digital twins to inform a mechanisms-based atrial fibrillation ablation strategy. *Nat. Cardiovasc. Res.* **3**, 857–868 (2024).
38. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996).
39. Ghosh, P. et al. Efficient prediction of cardiovascular disease using machine learning algorithms with relief and LASSO feature selection techniques. *IEEE Access* **9**, 19304–19326 (2021).
40. Shi, H. et al. Predicting drug-target interactions using Lasso with random forest based on evolutionary information and chemical structure. *Genomics* **111**, 1839–1852 (2019).
41. Lundberg, S. S.-I. L. A Unified approach to interpreting model predictions. *arXiv* <https://doi.org/10.48550/arXiv.1705.07874> (2017).
42. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
43. Lundberg, S. M. et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2**, 749–760 (2018).
44. Labarthe, S. et al. A bilayer model of human atria: mathematical background, construction, and assessment. *Europace* **16**, iv21–iv29 (2014).
45. Courtemanche, M., Ramirez, R. J. & Nattel, S. Ionic mechanisms underlying human atrial action potential properties: insights from a mathematical model. *Am. J. Physiol.* **275**, H301–H321 (1998).
46. Krummen, D. E. et al. Mechanisms of human atrial fibrillation initiation: clinical and computational studies of repolarization restitution and activation latency. *Circ. Arrhythm. Electrophysiol.* **5**, 1149–1159 (2012).
47. Boyle, P. M. et al. The fibrotic substrate in persistent atrial fibrillation patients: comparison between predictions from computational modeling and measurements from focal impulse and rotor mapping. *Front. Physiol.* **9**, 1151 (2018).
48. Boyle, P. M. et al. Comparing reentrant drivers predicted by image-based computational modeling and mapped by electrocardiographic imaging in persistent atrial fibrillation. *Front. Physiol.* **9**, 414 (2018).
49. Plank, G. et al. The openCARP simulation environment for cardiac electrophysiology. *Comput. Methods Prog. Biomed.* **208**, 106223 (2021).
50. Santangeli, P. & Marchlinski, F. E. Techniques for the provocation, localization, and ablation of non-pulmonary vein triggers for atrial fibrillation. *Heart Rhythm* **14**, 1087–1096 (2017).
51. Clayton, R. H., Zhuchkova, E. A. & Panfilov, A. V. Phase singularities and filaments: simplifying complexity in computational models of ventricular fibrillation. *Prog. Biophys. Mol. Biol.* **90**, 378–398 (2006).
52. Boyle, P. M., Masse, S., Nanthakumar, K. & Vigmond, E. J. Transmural IK(ATP) heterogeneity as a determinant of activation rate gradient during early ventricular fibrillation: mechanistic insights from rabbit ventricular models. *Heart Rhythm* **10**, 1710–1717 (2013).
53. Stridh, M. & Sornmo, L. Spatiotemporal QRST cancellation techniques for analysis of atrial fibrillation. *IEEE Trans. Biomed. Eng.* **48**, 105–111 (2001).
54. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830. <http://jmlr.org/papers/v12/pedregosa11a.html> (2011).
55. Bifulco, S. F. et al. Predicting arrhythmia recurrence post-ablation in atrial fibrillation using explainable machine learning: atrial meshes. *Dryad*. <https://doi.org/10.5061/dryad.kkwh70sg0> (2025).
56. Bifulco, S. F. et al. Predicting arrhythmia recurrence post-ablation in atrial fibrillation using explainable machine learning: code repository. *Dryad*. <https://doi.org/10.5061/dryad.4tmpg4fp9> (2025).
57. Ma, Y. et al. Explainable machine learning model reveals its decision-making process in identifying patients with paroxysmal atrial fibrillation at high risk for recurrence after catheter ablation. *BMC Cardiovasc. Disord.* **23**, 91 (2023).
58. Kuppahally, S. S. et al. Echocardiographic left atrial reverse remodeling after catheter ablation of atrial fibrillation is predicted by preablation delayed enhancement of left atrium by magnetic resonance imaging. *Am. Heart J.* **160**, 877–884 (2010).
59. Zhuang, J. et al. Association between left atrial size and atrial fibrillation recurrence after single circumferential pulmonary vein isolation: a systematic review and meta-analysis of observational studies. *Europace* **14**, 638–645 (2012).
60. Benali, K. et al. Recurrences of atrial fibrillation despite durable pulmonary vein isolation: the PARTY-PVI study. *Circ. Arrhythm. Electrophysiol.* **16**, e011354 (2023).
61. Atta-Fosu, T. et al. A new machine learning approach for predicting likelihood of recurrence following ablation for atrial fibrillation from CT. *BMC Med Imaging* **21**, 45 (2021).
62. Wen, S. et al. Association of postprocedural left atrial volume and reservoir function with outcomes in patients with atrial fibrillation undergoing catheter ablation. *J. Am. Soc. Echocardiogr.* **35**, 818–828 e813 (2022).
63. Zou, R., Kneller, J., Leon, L. J. & Nattel, S. Substrate size as a determinant of fibrillatory activity maintenance in a mathematical model of canine atrium. *Am. J. Physiol. Heart Circ. Physiol.* **289**, H1002–H1012 (2005).
64. Kalifa, J. et al. Intra-atrial pressure increases rate and organization of waves emanating from the superior pulmonary veins during atrial fibrillation. *Circulation* **108**, 668–671 (2003).
65. Marrouche, N. F. et al. Association of atrial tissue fibrosis identified by delayed enhancement MRI and atrial fibrillation catheter ablation: the DECAAF study. *JAMA* **311**, 498–506 (2014).
66. Pak, H. N. et al. Sex differences in mapping and rhythm outcomes of a repeat atrial fibrillation ablation. *Heart* **107**, 1862–1867 (2021).
67. Packer, D. L. et al. Ablation versus drug therapy for atrial fibrillation in heart failure: results from the CABANA trial. *Circulation* **143**, 1377–1390 (2021).
68. Packer, D. L. et al. 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* **321**, 1261–1274 (2019).

69. Turagam, M. K. et al. Clinical outcomes by sex after pulsed field ablation of atrial fibrillation. *JAMA Cardiol.* **8**, 1142–1151 (2023).
70. Bahnson, T. D. et al. Association between age and outcomes of catheter ablation versus medical therapy for atrial fibrillation: results from the CABANA trial. *Circulation* **145**, 796–804 (2022).
71. Marrouche, N. F. et al. Effect of MRI-guided fibrosis ablation vs conventional catheter ablation on atrial arrhythmia recurrence in patients with persistent atrial fibrillation: the DECAAF II randomized clinical trial. *JAMA* **327**, 2296–2305 (2022).
72. Kinoshita, M. et al. Role of smoking in the recurrence of atrial arrhythmias after cardioversion. *Am. J. Cardiol.* **104**, 678–682 (2009).
73. Brusa, E., Cibrario, L., Delprete, C. & Di Maggio, L. G. Explainable AI for machine fault diagnosis: understanding features' contribution in machine learning models for industrial condition monitoring. *Appl. Sci.* **13**, 2038 (2023).
74. Chao, T. F. et al. Clinical outcome of catheter ablation in patients with nonparoxysmal atrial fibrillation: results of 3-year follow-up. *Circ. Arrhythm. Electrophysiol.* **5**, 514–520 (2012).
75. Letsas, K. P. et al. CHADS₂ and CHA₂DS₂-VASc scores as predictors of left atrial ablation outcomes for paroxysmal atrial fibrillation. *Europace* **16**, 202–207 (2014).
76. Prakosa, A. et al. Personalized virtual-heart technology for guiding the ablation of infarct-related ventricular tachycardia. *Nat. Biomed. Eng.* **2**, 732–740 (2018).
77. Boyle, P. M. et al. Computationally guided personalized targeted ablation of persistent atrial fibrillation. *Nat. Biomed. Eng.* **3**, 870–879 (2019).

Acknowledgements

We thank the UW Department of Bioengineering and Division of Cardiology for supporting our research teams. We also thank our sources of funding: ARCS Foundation (SFB, MJM), Catherine Holmes Wilkins Charitable Foundation (PMB), and John Locke Charitable Trust (NA). Research reported in this publication was also supported by the National Heart, Lung, and Blood Institute and the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award numbers R01HL158668 (NA, PMB) and T32EB001650 (SFB). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and interpretation, or the decision to publish the work.

Author contributions

Conceptualization—S.F.B., N.A., P.M.B. Formal Analysis—S.F.B., M.J.M., Y.C., I.K., F.M. Curation, management, and interpretation of clinical data—

Y.C., F.M., N.A. Writing of original draft—S.F.B. Preparation of revised manuscript, including new analysis—M.J.M. Review, editing, and approval of manuscript—all co-authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43856-025-01058-4>.

Correspondence and requests for materials should be addressed to Nazem Akoum or Patrick M. Boyle.

Peer review information *Communications Medicine* thanks Omer Berenfeld, Junaid A. B. Zaman, Julien Oster, and the other anonymous reviewer(s) for their contribution to the peer review of this work. [A peer review file is available].

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025