Hybrid machine learning to localize atrial flutter substrates using the surface 12-lead electrocardiogram

Giorgio Luongo [1]*[†], **Gaetano Vacanti**[2]*[†], **Vincent Nitzke**[1], **Deborah Nairn**[1], **Claudia Nagel**[1], **Diba Kabiri**[2], **Tiago P. Almeida**[3], **Diogo C. Soriano**[4], **Massimo W. Rivolta**[5], **Ghulam André Ng**[3], **Olaf Dössel** [1], **Armin Luik** [2], **Roberto Sassi**[5], **Claus Schmitt**[2], and **Axel Loewe** [1]

[1]Institute of Biomedical Engineering (IBT), Karlsruhe Institute of Technology (KIT), Fritz-Haber-Weg 1, 76131 Karlsruhe, Germany; [2]Medizinische Klinik IV, Städtisches Klinikum Karlsruhe, Moltkestrasse, 90, 76182, Karlsruhe, Germany; [3]Department of Cardiovascular Sciences, University of Leicester, NIHR Leicester Biomedical Research Centre, Leicester, UK; [4]Engineering, Modelling and Applied Social Sciences Centre, ABC Federal University, São Bernardo do Campo, Brazil; and [5]Dipartimento di Informatica, Università degli Studi di Milano, Milan, Italy

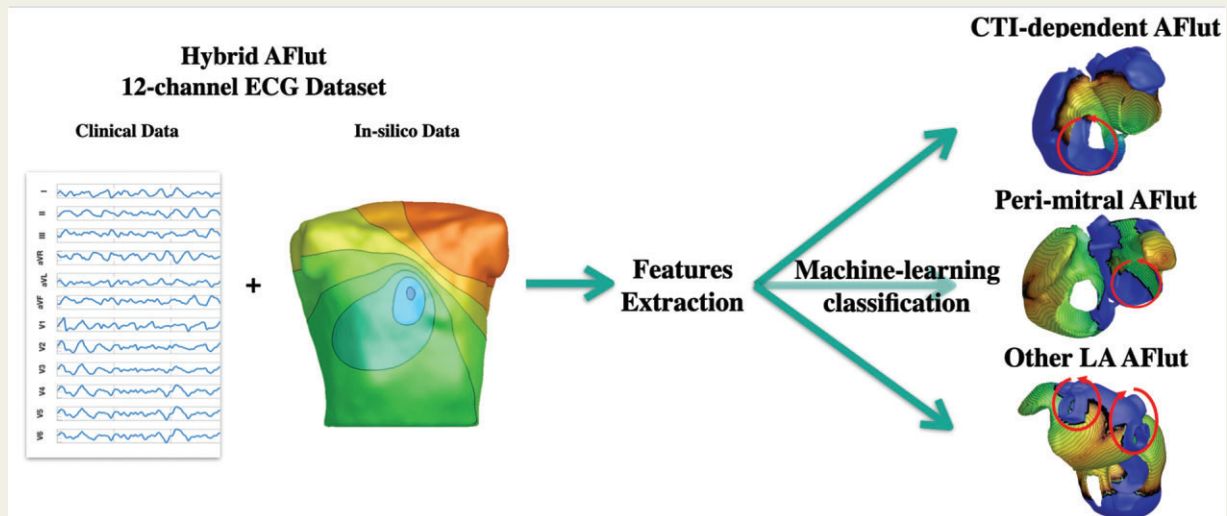|	
Aims	Atrial flutter (AFlut) is a common re-entrant atrial tachycardia driven by self-sustainable mechanisms that cause excitations to propagate along pathways different from sinus rhythm. Intra-cardiac electrophysiological mapping and catheter ablation are often performed without detailed prior knowledge of the mechanism perpetuating AFlut, likely prolonging the procedure time of these invasive interventions. We sought to discriminate the AFlut location [cavotricuspid isthmus-dependent (CTI), peri-mitral, and other left atrium (LA) AFlut classes] with a machine learning-based algorithm using only the non-invasive signals from the 12-lead electrocardiogram (ECG).
Methods and results	Hybrid 12-lead ECG dataset of 1769 signals was used (1424 *in silico* ECGs, and 345 clinical ECGs from 115 patients—three different ECG segments over time were extracted from each patient corresponding to single AFlut cycles). Seventy-seven features were extracted. A decision tree classifier with a hold-out classification approach was trained, validated, and tested on the dataset randomly split after selecting the most informative features. The clinical test set comprised 38 patients (114 clinical ECGs). The classifier yielded 76.3% accuracy on the clinical test set with a sensitivity of 89.7%, 75.0%, and 64.1% and a positive predictive value of 71.4%, 75.0%, and 86.2% for CTI, peri-mitral, and other LA class, respectively. Considering majority vote of the three segments taken from each patient, the CTI class was correctly classified at 92%.
Conclusion	Our results show that a machine learning classifier relying only on non-invasive signals can potentially identify the location of AFlut mechanisms. This method could aid in planning and tailoring patient-specific AFlut treatments.

* Corresponding authors. Tel: +49 721608 42650; fax: +49 721608 42789. *E-mail address*: publications@ibt.kit.edu (G.L.); Tel: +49 7219742960. *E-mail address*: gaetano.vacanti@klinikum-karlsruhe.de (G.V.)

† The first two authors contributed equally to the study.

Graphical Abstract

What's new?

- Non-invasive diagnostic algorithm for atrial flutter location discrimination.
- Identification of atrial flutter mechanisms using short F-wave segments [atrial activity in the 12-lead electrocardiogram (ECG)].
- Hybrid machine learning classifier (training dataset: *in silico* + clinical ECGs) for atrial flutter location discrimination.
- Machine learning algorithm based on selected features.
- Patient-specific ablation therapy.

Introduction

Atrial Flutter (AFlut) represents one of the most common supraventricular arrhythmias.[1,2] They are defined as organized, macro-re-entrant atrial tachycardias. The re-entry could either revolve around the tricuspid annulus [so-called typical forms or cavotricuspid-isthmus (CTI)-dependent forms], or originating in other atrial regions, such as the mitral annulus, the superior vena cava, or the pulmonary veins (PV), mostly facilitated by previous atrial ablations.[3,4] The widespread use of PV isolation and other left atrial ablation procedures for the treatment of atrial fibrillation may alter the normal activation patterns in the left atrium (LA) as well.[5,6] As a consequence, the prevalence of atypical left AFlut and of CTI-dependent flutter with atypical electrocardiogram (ECG) patterns post-ablation procedures increased in the recent years. These complex arrhythmias pose new diagnostic and classification challenges.[7–9]

Clinical diagnosis of AFlut currently relies on the interpretation of a non-invasive 12-lead surface ECG. Although distinctive features for identifying typical CTI-dependent flutter have long been identified and often enable an easy diagnosis, atypical non-CTI-dependent flutter forms are more difficult to recognize.[9] Catheter ablation represents a curative therapy for all forms of AFlut. The CTI-dependent flutter requires relatively straightforward, right-sided procedures with an anatomically well-characterized target: the CTI. Conversely, non-CTI-dependent flutter requires longer and technically more challenging procedures. In those procedures, transseptal puncture and electroanatomical mapping require additional preparation and additional diagnostic examinations such as transoesophageal echocardiography, adding further possible complications for the patient.[10] Non-invasive identification of the AFlut location enables to delineate the most likely target prior to the procedure. This would allow to avoid time-consuming and potentially riskier electroanatomic mapping and even pre-procedural transoesophageal echocardiography if a right-side flutter is identified and it is therefore appealing for optimal planning of procedures and utilization of hospital resources.

Several clinical schemes based on visual, non-computational, inspection of P-wave morphology of the surface 12-lead ECGs have been proposed, with equivocal results.[11,12] Machine learning-based algorithms are an emerging tool in classifying many diseases and have shown promising results in the field of cardiac arrhythmia detection.[13,14] Compared with clinical schemes, machine learning classifiers have a larger flexibility to fit the data, and are less operator-dependent resulting in better performance and more standardized approaches. Moreover, a feature selected-based machine learning algorithm can lead to a clear interpretation of the results as clinical algorithms do. As opposed to a deep learning approach that is

Table 1 Test set patient characteristics mean values with univariate ANOVA test between groups (*P*-values < 0.05)

	All patients (*n* = 36)	CTI-dependent AFlut (*n* = 13)	Peri-mitral AFlut (*n* = 12)	Other LA AFlut (*n* = 13)	Univariate (*P*-value)	Multivariate (*P*-value)
Age (years)	70 (8)	68 (9)	69 (11)	71 (5)	0.59	0.39
Female	10 (27.8)	1 (7.7)	5 (41.7)	4 (30.8)	0.16	0.25
Body mass index (kg/m^2)	289 (4.9)	29 (3.8)	27 (4.4)	29 (5.0)	0.09	0.09
Coronary artery disease	14 (38.9)	6 (46.1)	5 (41.7)	3 (23.1)	0.51	–
LVEF (%)	55 (7)	51 (3)	56 (5)	56 (7)	0.16	–
LVEDD (mm)	47 (5.5)	50 (5.2)	45 (5.2)	43 (8.3)	0.06	0.12
LAD (mm)	44 (4)	46 (6)	42 (5)	46 (7)	0.48	–
Previous catheter ablation	24 (66.7)	3 (23.1)	11 (91.7)	10 (76.9)	<0.01	0.07
Machine learning classifier	–	–	–	–	–	0.03

Values are given as mean (±standard deviation) or number (%). Multivariate regression analysis performed between the variables with a univariate *P*-value <0.1 and our classifier.
LAD, left atrial diameter; LVEDD, left ventricular end-diastolic diameter; LVEF, left ventricular ejection fraction.

referred to as a black box model (data goes in, decisions come out, but the processes between input and output are not evident).

In a previous work, we developed a machine learning classifier that provides an accurate and reliable classification for AFlut location for *in silico* signals.[15] After further training with *in silico* and clinical ECGs (hybrid approach) to discriminate CTI-dependent AFlut vs. peri-mitral AFlut vs. other non-CTI-dependent flutter classes, we tested the performance of the classifier on clinical ECGs. In the present work, we offer a proof of concept for a clinical tool able to non-invasively predict the target location for therapeutic catheter ablation, fostering the enablement of a more personalized therapy and better allocation of medical resources.

Methods

Study population

Clinical data were retrospectively collected, including standard 12-lead surface ECGs and electrophysiological data of 115 patients who presented with AFlut on baseline ECGs between 2015 and 2020 and underwent an electrophysiological study and catheter ablation in Städtisches Klinikum Karlsruhe. Inclusion criteria were atrial arrhythmia ECG characteristics and subsequent invasively confirmed diagnosis of AFlut. Exclusion criteria were the absence of complete and clinically evaluable 12-lead ECG documentation of atrial arrhythmia.

Exact arrhythmic mechanisms were confirmed invasively during the electrophysiological examination by termination of the arrhythmias after successful catheter ablation of the target site or non-inducibility of the arrhythmia by pacing or pharmaceutical challenge after catheter ablation. Seventy-three patients were diagnosed with CTI-dependent AFlut, 20 peri-mitral flutter, and 22 other non-CTI-dependent flutter with a critical isthmus in the LA ('other LA AFlut'). Patients with right AFlut with non-CTI-dependent mechanism (such as re-entry in the superior vena cava) were not present in the study population. The clinical characteristics of the test set patient cohort are shown in *Table 1* (see Supplementary material online for the clinical characteristics of all the patients).

From each patient, three different single AFlut segments were extracted from the standard 12-lead ECG (10 s length) to run the analysis, resulting in a total of 345 clinical signals (219 CTI-dependent AFlut, 60 peri-mitral AFlut, and 66 other LA AFlut). An AFlut segment is a single flutter cycle in-between two consecutive QRS-T complexes (red segments in *Figure 1*). Segments were manually derived from ECGs. To obtain such a segment, a random point in the F-wave was selected as starting point, and the subsequent part of the signal was manually extracted until the signal itself returned to the same state as the selected starting point (*Figure 1*). The three independent (not in the same phase of the F-wave) clearest and least compromised by QRS-T complexes AFlut segments were randomly chosen from each patient for the analysis. Segment quality and its distinction from QRS-T complexes were the only criteria for selecting the three segments from each ECG to avoid bias in the acquisition protocol. The 12-lead ECGs (1 kHz sampling frequency) were notch filtered at 50 Hz and band-pass Butterworth filtered between 0.05 Hz and 100 Hz (*Figure 1*).

In silico population

A database with computational AFlut scenarios was setup based on computational studies conducted in previous work.[16] Cardiac excitation was modelled using the fast marching approach to solve the Eikonal equation on 100 bi-atrial anatomies generated from a statistical shape model.[17] Scars were added circumferentially around ipsilateral PVs representing previous PV isolation. More details regarding the simulations can be found in Luongo *et al.*[16]

In total, 15 mechanisms of AFlut were implemented and merged into the three classes under analysis in this study: CTI-dependent, peri-mitral, and other LA AFlut mechanisms (a complete list of the *in silico* mechanisms and their classification is shown in *Table 2*; re-entry path examples and ECGs of *in silico* cases in *Figure 2*). The simulated cardiac excitation was used to compute the body surface potential map (BSPM) on the mean geometry derived from a statistical shape model of the torso.[17] Finally, conventional 12-lead ECGs were extracted from the BSPM (1 kHz sampling frequency). From the *in silico* ECGs, the AFlut single cycle segments were extracted. A total of 1424 sets of signals were obtained (198 CTI-dependent, 186 peri-mitral, and 1040 other LA AFlut mechanisms). Due to anatomical reasons, the implementation and/or sustainment of some scenarios was not possible on some atrial geometries. The final simulated dataset was held unbalanced to keep the number of geometries used in generating the simulations constant to prevent introducing anatomical bias. The unbalanced dataset was then accounted for during classifier training as described in machine learning classification.

The AFlut from the acquired clinical data matched the simulated data. Particularly, in the other LA class, the clinical cases obtained
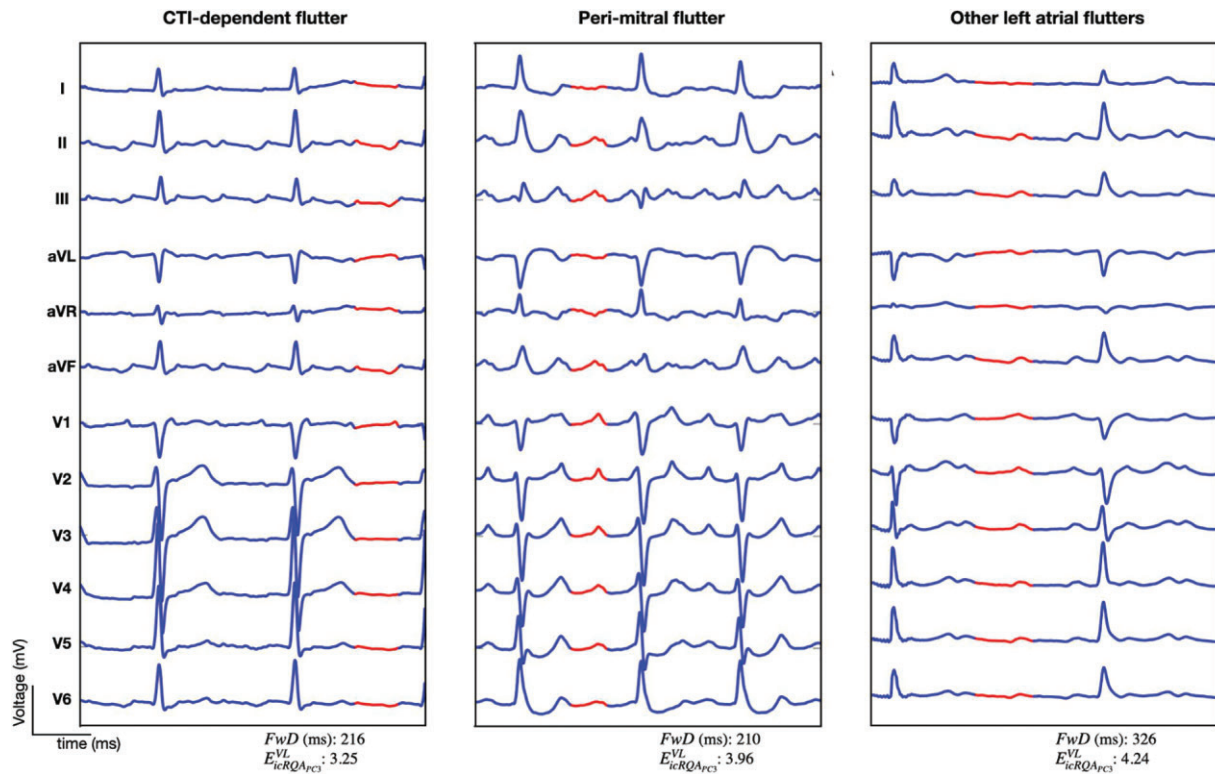
Figure 1 Example of clinical CTI-dependent, peri-mitral, and other LA AFlut 12-lead ECGs in the test set, respectively. The red segments represent one of the three AFlut single cycles extracted and used in this work for this specific patient. The values of the two most predictive metrics are reported for the three cases in example (FwD and $E_{icRQA_{PC3}}^{VL}$).

Table 2 *In silico* **AFlut mechanisms and attribution to the three classes under analysis**

Mechanism	Atrium	Position	Direction	Class
Macro-reentry	RA	Tricuspid valve	ccw	CTI-dependent
Macro-reentry	RA	Tricuspid valve	cw	CTI-dependent
Macro-reentry	LA	Mitral valve	ccw	Peri-mitral
Macro-reentry	LA	Mitral valve	cw	Peri-mitral
Scar-related re-entry	LA	LPV	post	Other LA
Scar-related re-entry	LA	LPV	ant	Other LA
Scar-related re-entry	LA	RPV	post	Other LA
Scar-related re-entry	LA	RPV	ant	Other LA
Figure-8 macro-reentry	LA	Both PVs	ant	Other LA
Figure-8 macro-reentry	LA	Both PVs	post	Other LA
Figure-8 macro-reentry	LA	RPVs	ant	Other LA
Focal source	LA	RSPV anterior		Other LA
Focal source	LA	RSPV posterior		Other LA
Focal source	LA	LSPV anterior		Other LA
Focal source	LA	LSPV posterior		Other LA

ant, anterior; cw, clockwise; ccw, counterclockwise; LA, left atrium; LPV, left pulmonary vein; LSPV, left superior pulmonary vein; post, posterior; PVs, pulmonary veins; RA, right atrium; RPV, right pulmonary vein; RSPV, right superior pulmonary vein.

correspond to the same simulated categories (i.e. the clinical other LA AFlut dataset is composed by 10 figure-of-eight re-entries, 3 focal sources, 5 micro-re-entries, and 4 scar-related re-entries). The pre-processing of the simulated and clinical signals was the same with the only difference that no filtering was applied to the simulated data, which contained no noise.
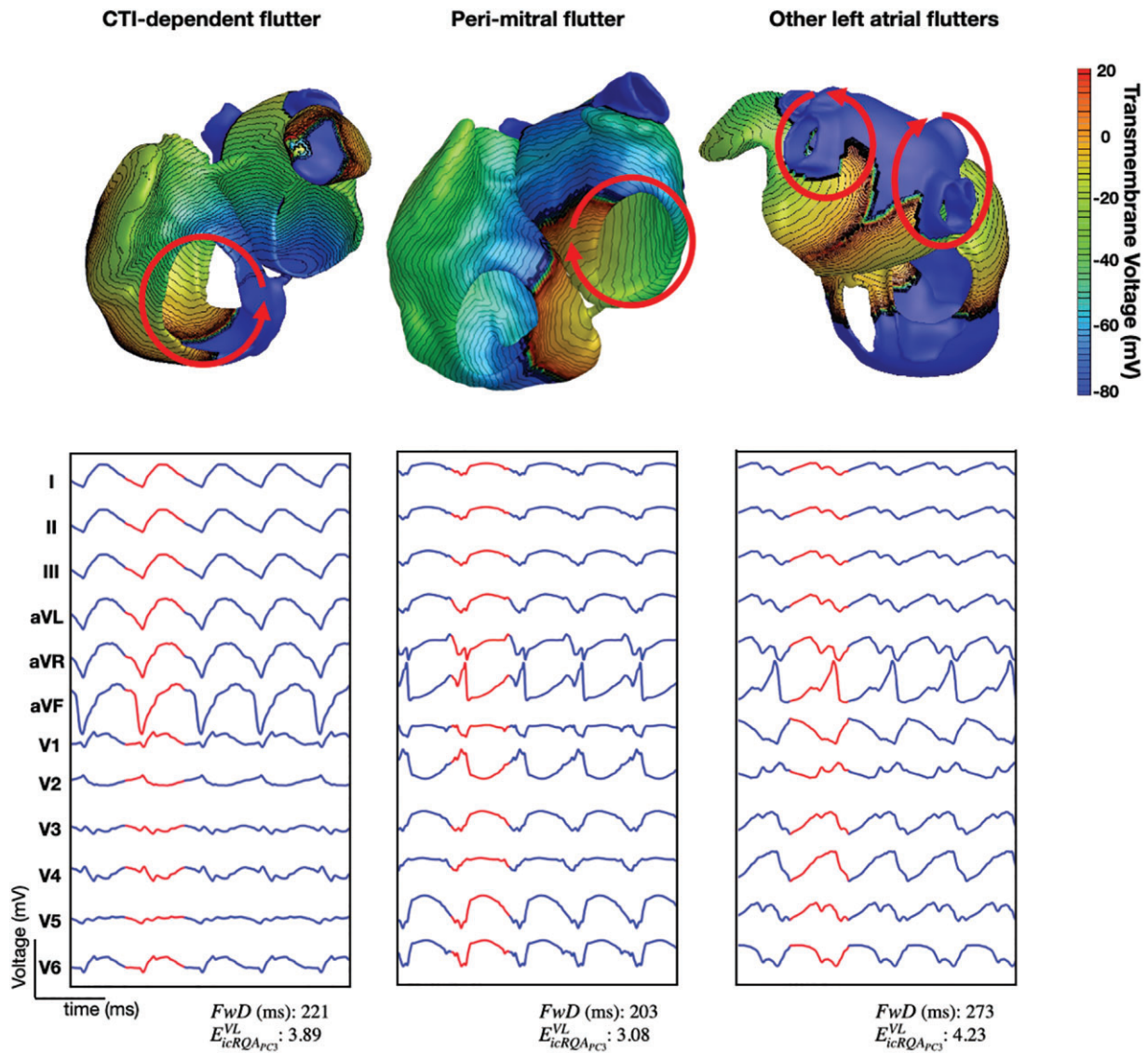
Figure 2 Example of different re-entry paths for each class in simulation with the respective 12-lead ECG. The top line shows a frame of one of the cardiac simulations computed in this study: CTI-dependent flutter with counterclockwise direction (left), peri-mitral flutter wit clockwise direction (middle), other left atrium flutter, i.e. figure-8 macro-re-entry with anterior direction (right). Following, in the bottom line there are three 12-lead ECGs as result of the top simulations. It can be seen that there is no QRS-T activity given the lack of ventricles from the simulations. The segments used for each signal have been highlighted in red. The values of the two most predictive metrics are reported for the three cases in example (*FwD* and $E_{icRQA_{PC3}}^{VL}$).

Feature extraction and selection

Seventy-seven features were extracted from the signals using several bio-signal processing methods from different domains, i.e. time, frequency, wavelet, entropy, and non-linear recurrence analysis. A table summarizing the features and more information regarding the feature extraction methods are provided in the Supplementary material online.

Among these features some proved to be particularly relevant in pilot analysis: the F-wave (flutter wave, P-wave during AFlut) duration, the duration of a complete cycle of atrial electrical activation of each specific AFlut mechanism, i.e. the length of the F-wave in the ECGs (the feature was manually derived and corresponded to the length of the extracted ECG segments), wavelet features which are able to extract spectral and temporal information simultaneously from the signals, recurrence quantification analysis features to analyse the regularity and stability of time domain signals,[16] morphological features such as the fragmented conduction index, the optimal model order (number of Gaussian functions needed to model the signals),[18] and symbolic dynamic features.

To reduce the degree of redundant information obtainable from the extracted features and therefore to avoid over-fitting by using too many features, the optimal set of features was selected with a greedy forward selection technique. This algorithm starts with an empty feature set and adds the feature which leads to the highest increase of classification accuracy in each iteration. The algorithm was stopped when the performance, based on the validation set, could not be further increased. In order to

Table 3　Clinical test set confusion matrixes for three-class classification

(A)		True class		
		CTI-dependent	Peri-mitral	Others
	CTI-dependent	35	7	7
Predicted class	Peri-mitral	2	27	7
	Others	2	2	25
(B)		True class		
		CTI-dependent	Peri-mitral	Others
	CTI-dependent	12	2	2
Predicted class	Peri-mitral	0	10	3
	Others	1	0	8

(A) Confusion matrix per ECG-segment (three different ECG segments were extracted from each patient independently). (B) Confusion matrix per patient.

handle possible correlations among features, the candidate feature was only added to the set if the correlation coefficient with any of the already included features was <0.6.

Shapley calculation was implemented to analyse *a posteriori* the importance of the features selected for classification once the model was trained. Shapley calculation was run 1000 times with random samples (Bootstrap approach) to calculate the standard deviation (SD).

Machine learning classification

A decision tree was implemented for three-class classification: CTI-dependent AFlut, peri-mitral AFlut, and other LA AFlut (similar results obtained with other machine learning algorithms are provided in the Supplementary material online). The classifier was trained and applied using the MATLAB (The MathWorks, Inc.) functions *fitctree* and *predict*, respectively. The dataset used to train, validate, and test the classifier was a hybrid of the *in silico* and clinical datasets with a total of 1769 signals. First, a multi-feature classification was performed with the feature set selected as described in the previous section. Hold-out classification was performed randomly, dividing the dataset into a training set, validation set, and test set with a ratio of 70, 15, and 15%, respectively. Signal segments from the same patient were always used solely in one of the sets to avoid over-fitting. The training set was used to tune the classifier parameters, while the validation set was used for the greedy feature selection optimization. During training, class imbalance was addressed by assigning a weight to each sample in a given class (by setting the *weights* parameter and the *Prior* model parameter to *uniform* in the MATLAB *fitctree* function). To solve the problem of the different number of data points and the different number of cases for each AFlut class between simulated and clinical data, the weight was inversely proportional to the number of samples in the class and extra weights were added for clinical cases to give them more importance during the training phase. The decision tree can also automatically provide a degree of confidence of the prediction. This degree of confidence is the posterior probability that a test sample is of a specific class, which is provided for all three classes. Lastly, the trained classifier was tested on the clinical test set (114 segments data from 38 patients—39, 36, and 39 ECG segments for each class, respectively) and on a simulated test set (90 simulated ECGs—30, 30, and 30 segments for each class, respectively).

A blind annotation of the clinical test ECG signals was performed by a trained physician in order to compare the physician's prediction
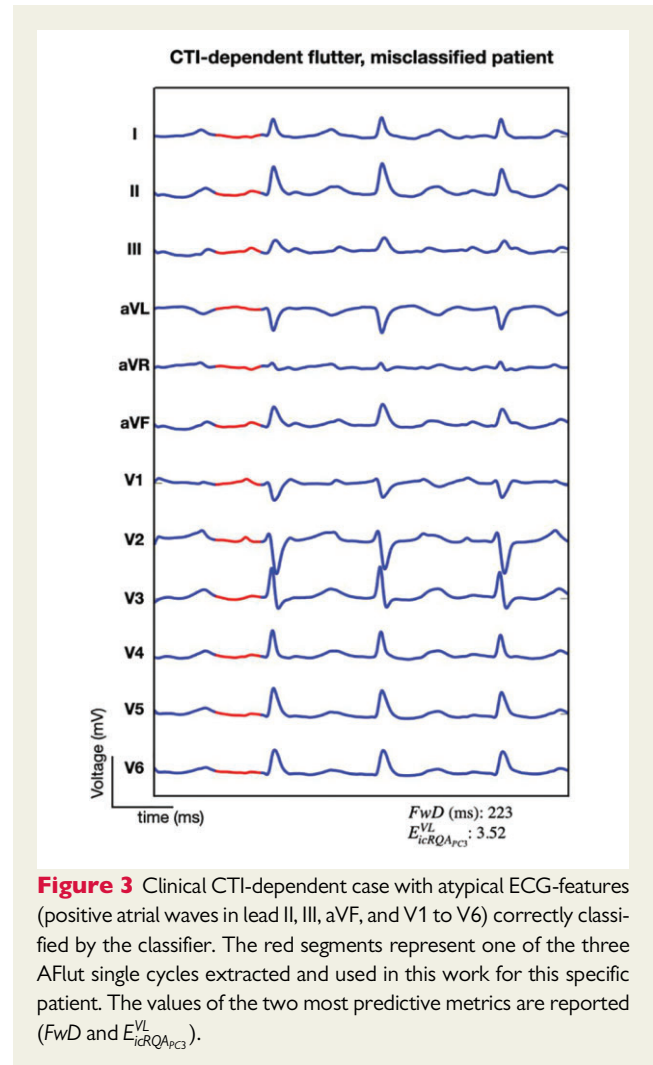

Figure 3 Clinical CTI-dependent case with atypical ECG-features (positive atrial waves in lead II, III, aVF, and V1 to V6) correctly classified by the classifier. The red segments represent one of the three AFlut single cycles extracted and used in this work for this specific patient. The values of the two most predictive metrics are reported (FwD and $E_{icRQA_{PC3}}^{VL}$).

with the classifier's prediction to evaluate the added value of the algorithm.

Ethics statement

The authors designed the study, gathered and analysed the data according to the Helsinki Declaration guidelines on human research. The research protocol used in this study was reviewed and approved by the institutional review board. All patients provided written informed consent.

Statistical analysis

Classifier performance was evaluated using the accuracy (ACC), sensitivity, and positive predictive value (PPV). Sensitivity and PPV were calculated for each class, considering the specific class as positive and the remaining classes as negative. The clinical characteristics of our test set patient cohort and the extracted features were evaluated with the ANOVA test between the classes (*P*-values <0.05 considered significant; *Table 1*, univariate column; similar analysis for the full dataset in the Supplementary material online). Next, multivariate regression analysis was performed on variables that differed between groups with a *P*-value <0.1 (from the previous ANOVA test) and the machine learning classifier (*Table 1*, Multivariate column). Age and sex were included in the multivariate model for their clinical relevance.

Results

Patient characteristics of the test set patients with associated univariate and multivariate analysis are provided in *Table 1* (for similar analysis for the full dataset in the Supplementary material online). The multivariate regression analysis between the variables age, sex, body mass, left ventricular ejection fraction (LVEF), left ventricular end-diastolic diameter (LVEDD), previous catheter ablation, and our classifier showed that the classifier was the only significant variable ($P = 0.03$). As such, it adds value beyond the routine clinical parameters. The blind classification performed by a trained physician achieved 44.7% correct classification on the clinical test ECGs: 84.6% CTI AFlut, 41.7% peri-mitral AFlut, and 7.7% other LA AFlut cases were correctly identified by the physician.

The hybrid decision tree for three-class classification (CTI-dependent AFlut, peri-mitral AFlut, and other LA AFlut—*Figures 1 and 2*) achieved an accuracy of 76.3% on the clinical test set with a sensitivity of 89.7%, 75.0%, and 64.1% and a PPV of 71.4%, 75.0%, and 86.2% for each class, respectively. Other classification methods were tested but yielded worse performance (see Supplementary material online). The mean degree of confidence that the classification obtained for each class when predicted was 81.63%, 81.18%, and 70.30% for the CTI, peri-mitral, and other LA AFlut classes, respectively. The decision tree was also tested on a simulated test set achieving 66.4% accuracy. *Table 3* shows the confusion matrices obtained from the decision tree classifier on the clinical test set both for ECG segments and patients. Considering the mean prediction out of the three segments for each patient, the classifier correctly classified 78.9% of the patients. The four AFlut segments misclassified for the CTI-dependent AFlut classes belong to three different patients, this means that two out of the three patients were correctly classified considering the mean prediction (92.3% sensitivity for CTI AFlut patients). Whereas the peri-mitral and other LA AFlut yielded 83.3% and 61.5% patient's classification sensitivity, respectively. The implemented classifier was also able to classify CTI-dependent AFlut with atypical characteristics in the surface ECGs (two patients in the clinical test set; see example in *Figure 3*).

The greedy forward selection technique reached the maximum accuracy with 18 features. The 18 selected features were: F-wave duration, 8 features from the dynamic symbolic analysis (the number of high valued segments, low valued segments, bottom valued segments, isolines, sequence of top–top, low–zero, low–low, and bottom–low segments in the signals), the minimum and mean optimal model orders to model each ECG lead with Gaussian functions, the 7th, 8th, and 9th coefficients from the discrete cosine transformation, the mean prominence of the 5th and 6th order wavelets, the minimum fragmented conduction index, and the vertical entropy calculated with the individual component recurrence quantification analysis on the 3rd principal component ($icRQA_{PC3}$).

The Shapley calculation to analyse *a posteriori* the importance of the 18 selected features showed that the most relevant features for the implemented classification were the F-wave duration (FwD) and the $icRQA_{PC3}$ vertical entropy ($E^{VL}_{icRQA_{PC3}}$) mainly supported by the mean optimal model order, and the mean prominence of the 5th order wavelet.

The definition of all extracted features, the feature set, the detailed feature importance analysis, and the performance achieved with a classifier trained with only simulated data or clinical data only can be found in the Supplementary material online.

Discussion

Our results suggest that a non-invasive machine learning approach based on surface ECG analysis can aid in the discrimination of the location of the substrate which is sustaining AFlut. This could improve protocols for clinical therapeutic decision-making and ablation procedure planning.

The 12-lead ECG is broadly used for cardiac diagnostics, including diagnosis and classification of AFlut from other cardiac rhythms.[11–14] Many ECG-based clinical schemes have been proposed, with mixed results. Due to novel and widespread left atrial ablation procedures in the last decades, rare clinical entities with diagnostic challenges (i.e. left AFlut or right AFlut with atypical ECG-features) are increasing in prevalence and entering the routine of electrophysiology labs.[7–9] In addition, the 12-lead ECG is the most common and available tool for arrhythmia diagnosis in clinical practice, thus its use would not lead to any extra expense. In contrast, methods with extra electrodes or body surface potential maps are not available in all hospitals and would, therefore, require additional investment by the health care system.

To the best of our knowledge, ours represent the first work to implement a decision tree classifier to discriminate the location of AFlut mechanisms by using *in silico* and clinical non-invasive ECG signals. The computational 12-lead ECG simulations provided an ideal and controlled environment to generate a consistent ground truth dataset of AFlut mechanisms. This *in silico* dataset served as a reinforcement to the clinical dataset to solve the problematic availability of clinical data where ground truth was ascertained by invasive electrophysiological analysis. The clinical data used during training helped to bridge the domain gap between simulated and measured ECGs. In addition, the machine learning algorithm implemented in this project uses carefully selected features. As a result, clinical knowledge is used to optimize the algorithm, rendering also the interpretation of the results easier than a deep learning approach.

This study builds on our previous work where we developed a classifier for 20 AFlut mechanisms discrimination for *in silico* signals.[15] Further analysis of the classifier showed that 20 classes of AFlut to be classified was a too complex goal at the time. In addition, the classifier trained with only simulated data proved to be unable to generalize on clinical data. Therefore, in this study we used a hybrid approach and a reduced number of classes. Of note, no other clinical characteristics were statistically relevant in correctly classifying the AFlut (*Table 1*).

Feature analysis

The Shapley calculation showed that, of the 18 features in the feature set needed for classification, 4 were most relevant: The F-wave duration, the $icRQA_{PC3}$ vertical entropy, the mean optimal model order, and the mean prominence of the 5th order wavelet. The F-wave duration already proved to be a fundamental feature in the AFlut characterization in our previous works[15,19] together with the vertical entropy $icRQA_{PC3}$. In fact, several previous results have reported how RQA approaches can be useful in the characterization and

discrimination of different cardiac arrhythmias.[16,20] These two features showed to be significantly higher in values in the 'other LA AFlut' class in comparison with the other classes. In contrast, there was no significant difference between the CTI-dependent and the peri-mitral AFlut classes. The F-wave duration is a feature strongly influenced by the size and shape of the atria and the conduction velocity of the tissue. Therefore, it can be assumed that similar sizes between the two valves or complementary conduction velocities of the tissue may result in similar F-wave durations (case of CTI and peri-mitral AFlut classes). In contrast, the vertical entropy $icRQA_{PC3}$ is an expression of signal complexity and it can be reasonably assumed that the other LA AFlut class has significantly higher values than the other two classes due to the higher complexity of the mechanisms composing the other LA AFlut class in comparison with the CTI and peri-mitral AFlut classes. Thus, the resulting ECGs (specifically, the third principal components calculated from 12-lead ECGs by PCA) are more complex and irregular. However, these two features were used as main features by the algorithm likely due to their expression of both complexity of the mechanisms and atrial geometry and tissue properties (i.e. conduction velocity). On the other hand, regarding the mean prominence on the 5th order wavelet and the mean optimal model order, there was no significant difference between the three classes. Therefore, the importance identified by the Shapley evaluation for these two last features, but not observed by the significance analysis among the classes, must be solely due to their use in the algorithm to better define the classification. The feature importance analysis, and the distribution of the values of the four most important features are shown in the Supplementary material online.

Classification

Previous studies have sought to find consistent and reliable non-invasive predictors of location for AFlut. The most investigated and promising features have been the morphology of the atrial waves in 12-lead ECGs. Several clinical decision-making schemes based on such features have been proposed with mixed results and there is currently no consensus on how to classify AFlut based on the surface ECG.[11,12] Nevertheless, as different AFlut types require different ablation procedures with more challenging preparations for left atrial types, a pre-procedural diagnosis and characterization of the location of ablation targets would increase the efficacy of the procedure by permitting direct targeting of the region of interest. Moreover, it would also allow tailoring the procedure in a time- and resource-saving fashion, avoiding potential unnecessary complications or delays.

Our decision tree classifier based on the surface 12-lead ECGs achieved an accuracy of 76.3% on the clinical test set, demonstrating the ability to correctly classify most of the AFlut segments extracted from the 12-lead ECGs. In particular, the high sensitivity for the CTI-dependent AFlut (89.7% for segments' classification and 92.3% for patients' classification) shows the potentiality of the algorithm in the identification of these patients. Moreover, the PPV for the other LA AFlut class (86.2% and 88.9% for segments' and patients' classification, respectively) demonstrated that most of the segments and patients classified into this group were correct, making our classifier reliable in the identification of these AFlut. The significant difference in accuracy between the clinical and simulated test sets demonstrates the effectiveness of our classifier training method in giving greater weight to clinical data as the final application of our algorithm. In general, the

results indicate a good classification ability, likely negatively influenced by the lower number of cases for the peri-mitral and other LA AFlut classes in our training set. In addition, the degree of confidence automatically provided by the system gives added support to clinicians by knowing the order of probability of the three types of AFlut.

As *Table 3A* shows, four segments got misclassified for the CTI-dependent AFlut class. These segments belong to three different patients, thus indicating that the classifier correctly identified 92.3% of the patients in this group when considering the majority prediction from the three segments of each patient. Not having relied on a single, random segment per patient, but rather three ECG segments per patient, the classification accuracy per patient (*Table 3B*) of 78.9% demonstrates robustness and reliability of the classifier. Moreover, the comparison between the accuracy achieved by the classifier and the blind classification performed by a trained physician is a demonstration of added value of the algorithm to clinical practice. Whereas the physician was able to predict almost solely CTI cases, our classifier was also able to discriminate between peri-mitral and other LA AFlut (confusion matrix of the blind classification performed by the physician in the Supplementary material online).

As the therapy spectrum for atrial ablation broadens, the prevalence of AFlut with atypical characteristics increases leading to new diagnostic challenges. For example, CTI-dependent flutters, which occurs after left atrial procedures tend to present atrial waves with unusual morphologies. Such cases can easily be misinterpreted as left AFlut and lead to unnecessary, lengthier, and potentially more dangerous workup and procedures. Our classifier was able to correctly identify the nature of the AFlut also for such challenging cases. The CTI cases misclassified by the physician in blind mode were exactly these unusual CTI ECGs. In *Figure 3*, we present one of the two cases included in the clinical test set of patients who developed, after PV isolation ablation and further ablation of peri-mitral AFlut, a CTI-dependent AFlut. Despite the unusual characteristics in the surface ECGs (positive atrial waves in leads II, III, aVF, and V1 through V6), our classifier correctly identified their location.

Limitations

This study is limited to a small clinical dataset. As a next step, the algorithm should be trained and tested on a more extensive patient cohort. In particular, more patients for the peri-mitral and other LA AFlut classes should be collected and added to the training set to improve the classifier performance. Even though the manual segmentation of the clinical ECG is not time-consuming, in view of future clinical applications, automated segmentation of clinical ECG should be implemented to extract the F-wave. However, due to the strong influence that the F-wave duration has on the final classifier, the automatic segmentation requires to be robust and reliable.

Fifteen AFlut mechanisms related to clinical situations have been implemented. Although the clinical AFlut and the simulated ones matched, they are just a general representation of the mechanisms that are most commonly found in the literature. More mechanisms could be included in the dataset, considering the heterogeneity seen in the clinical practice, e.g. slow conduction areas.

Imaging or scar information could help in the classification, going to give relevant information on the size of the atria and on the possible paths that the excitation may follow because of the scars present. However, this would reduce one of the main goals of this study: to

use non-invasive, easily available, and inexpensive systems (i.e. 12-lead ECGs) to identify the location of AFlut.

Conclusions

The results presented in this study show that a machine learning feature-based classifier can distinguish between CTI-dependent AFlut vs. peri-mitral AFlut vs. other LA AFlut using single AFlut loop segments. Additionally, a hybrid approach (*in silico* data + clinical data) to train the classifier can be successful when it is difficult to get enough clinical data for purely clinical machine learning. As such, a machine learning-based classifier leveraging the routinely available non-invasive 12-lead ECG can be valuable for clinical decision-making and increase the personalization of therapy. Our machine learning classifier can correctly non-invasively predict ablation targets avoiding the need for transseptal catheterization and LA mapping and consequently even pre-procedural transoesophageal echocardiography in some cases.

Supplementary material

Supplementary material is available at *Europace* online.

Authorship

All authors attest they meet the current Committee on Publication Ethics (COPE) and ICMJE criteria for authorship.

References

1. Granada J, Uribe W, Chyou P-H, Maassen K, Vierkant R, Smith PN *et al.* Incidence and predictors of atrial flutter in the general population. *J Am Coll Cardiol* 2000;**36**:2242–6.
2. Page RL, Joglar JA, Caldwell MA, Calkins H, Conti JB, Deal BJ *et al.*; Evidence Review Committee Chair‡. 2015 ACC/AHA/HRS guideline for the management of adult patients with supraventricular tachycardia: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. *Circulation* 2016;**133**:e506–74.
3. Cosío FG. Atrial flutter, typical and atypical: a review. *Arrhythm Electrophysiol Rev* 2017;**6**:55–62.
4. Saoudi N, Cosío F, Waldo A, Chen SA, Iesaka Y, Lesh M *et al.*; Working Group of Arrhythmias of the European of Cardiology and the North American Society of Pacing and Electrophysiology. A classification of atrial flutter and regular atrial tachycardia according to electrophysiological mechanisms and anatomical bases; a Statement from a Joint Expert Group from The Working Group of Arrhythmias of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Eur Heart J* 2001;**22**:1162–82.
5. Pappone C, Oreto G, Rosanio S, Vicedomini G, Tocchi M, Gugliotta F *et al.* Atrial electroanatomic remodeling after circumferential radiofrequency pulmonary vein ablation. *Circulation* 2001;**104**:2539–44.
6. Loewe A, Poremba E, Oesterlein T, Luik A, Schmitt C, Seemann G *et al.* Patient-specific identification of atrial flutter vulnerability—a computational approach to reveal latent reentry pathways. *Front Physiol* 2018;**9**:1910.
7. Mesas CE, Pappone C, Lang CCE, Gugliotta F, Tomita T, Vicedomini G *et al.* Left atrial tachycardia after circumferential pulmonary vein ablation for atrial fibrillation: electroanatomic characterization and treatment. *J Am Coll Cardiol* 2004;**44**:1071–9.
8. Chugh A, Oral H, Good E, Han J, Tamirisa K, Lemola K *et al.* Catheter ablation of atypical atrial flutter and atrial tachycardia within the coronary sinus after left atrial ablation for atrial fibrillation. *J Am Coll Cardiol* 2005;**46**:83–91.
9. Hoffmayer KS, Yang Y, Joseph S, Mccabe JM, Bhave P, Hsu J *et al.* Predictors of unusual ECG characteristics in cavotricuspid isthmus-dependent atrial flutter ablation. *Pacing Clin Electrophysiol* 2011;**34**:1251–7.
10. Tabbernor L. The radiofrequency ablation of atrial flutter. *Br J Card Nurs* 2006;**1**:191–7.
11. Gerstenfeld EP, Dixit S, Bala R, Callans DJ, Lin D, Sauer W *et al.* Surface electrocardiogram characteristics of atrial tachycardias occurring after pulmonary vein isolation. *Heart Rhythm* 2007;**4**:1136–43.
12. Pascale P, Roten L, Shah AJ, Scherr D, Komatsu Y, Ramoul K *et al.* Useful electrocardiographic features to help identify the mechanism of atrial tachycardia occurring after persistent atrial fibrillation ablation. *JACC Clin Electrophysiol* 2018;**4**:33–45.
13. Hemam ME, Dave AS, Rodríguez-Mañero M, Valderrábano M. Epiphenomenal re-entry and spurious focal activation detection by atrial fibrillation mapping algorithms. *JACC Clin Electrophysiol* 2021;**7**:923–32.
14. Luongo G, Azzolin L, Schuler S, Rivolta MW, Almeida TP, Martínez JP *et al.* Machine learning enables noninvasive prediction of atrial fibrillation driver location and acute pulmonary vein ablation success using the 12-lead ECG. *Cardiovasc Digit Health J* 2021;**2**:126–36.
15. Luongo G, Schuler S, Rivolta MW, Doessel O, Sassi R, Loewe A. 236 Automatic classification of 20 different types of atrial tachycardia using 12-lead ECG signals. *Europace* 2020;**22**:euaa162.048.
16. Luongo G, Schuler S, Luik A, Almeida TP, Soriano DC, Dossel O *et al.* Non-invasive characterization of atrial flutter mechanisms using recurrence quantification analysis on the ECG: a Computational Study. *IEEE Trans Biomed Eng* 2021;**68**:914–25.
17. Nagel C, Schuler S, Dössel O, Loewe A. A bi-atrial statistical shape model for large-scale in silico studies of human atria: model development and application to ECG simulations. *Med Image Anal* 2021;**74**:102210–8415.
18. Censi F, Calcagnini G, Ricci C, Ricci RP, Santini M, Grammatico A *et al.* P-wave morphology assessment by a Gaussian functions-based model in atrial fibrillation patients. *IEEE Trans Biomed Eng* 2007;**54**:663–72.
19. Luongo G, Schuler S, Rivolta MW, Dössel O, Sassi R, Loewe A. Automatic ECG-based discrimination of 20 atrial flutter mechanisms: influence of atrial and torso geometries. In *2020 Computing in Cardiology*. 2020. pp. 1–4.
20. Desai U, Martis RJ, Acharya UR, Nayak CG, Seshikala G, Shetty KR. Diagnosis of multiclass tachycardia beats using recurrence quantification analysis and ensemble classifiers. *J Mech Med Biol* 2016;**16**:1640005.