

Multi-modal artificial intelligence algorithm for the prediction of left atrial low-voltage areas in atrial fibrillation patient based on sinus rhythm electrocardiogram and clinical characteristics: a retrospective, multicentre study

Yirao Tao^{1,2,†}, Deyun Zhang^{3,4,†}, Naidong Pang^{1,2}, Shijia Geng^{3,4}, Chen Tan⁵, Ying Tian⁶, Shenda Hong^{7,8,*}, and XingPeng Liu^{1,2,6,*}

¹Department of Cardiology, Beijing Chaoyang Hospital, Capital Medical University, No. 8 Workers' Stadium South Road, Chaoyang District, Beijing 100020, China; ²Heart Center and Beijing Key Laboratory of Hypertension, Beijing Chaoyang Hospital, Capital Medical University, No. 8 Worker's Stadium South Road, Chaoyang District, Beijing 100020, China; ³Artificial Intelligence lab, HeartVoice Medical Technology, Hefei, China; ⁴Artificial Intelligence lab, HeartRhythm-HeartVoice Joint Laboratory, Beijing, China; ⁵Department of Cardiology, Hebei Yanda Hospital, Langfang, Hebei Province, China; ⁶Department of Cardiology, Dezhou HTRM Cardiovascular Hospital, Dezhou, Shandong Province, China; ⁷National Institute of Health Data Science, Peking University, Xueyuan Road No. 5, Haidian District, Beijing 100191, China; and ⁸Institute of Medical Technology, Health Science Center of Peking University, Xueyuan Road No. 38, Haidian District, Beijing 100191, China

Received 8 August 2024; revised 6 October 2024; accepted 15 November 2024; online publish-ahead-of-print 13 December 2024

Aims

We aimed to develop an artificial intelligence (AI) algorithm capable of accurately predicting the presence of left atrial low-voltage areas (LVAs) based on sinus rhythm electrocardiograms (ECGs) in patients with atrial fibrillation (AF).

Methods and results

The study included 1133 patients with AF who underwent catheter ablation procedures, with a total of 1787 12-lead ECG images analysed. Artificial intelligence-based algorithms were used to construct models for predicting the presence of LVAs. The DR-FLASH and APPLE clinical scores for LVAs prediction were calculated. A receiver operating characteristic (ROC) curve and a calibration curve were used to evaluate model performance. Multicentre validation included 92 AF patients from five centres, with a total of 174 ECGs. The data obtained from the participants were split into training ($n = 906$), validation ($n = 113$), and test sets ($n = 114$). Low-voltage areas were detected in 47.4% of all participants. Using ECG alone, the convolutional neural network (CNN) model achieved an area under the ROC curve (AUROC) of 0.704, outperforming both the DR-FLASH score (AUROC = 0.601) and the APPLE score (AUROC = 0.589). Two multimodal AI models, which integrated ECG images and clinical features, demonstrated higher diagnostic accuracy (AUROC 0.816 and 0.796 for the CNN-Multimodal and CNN-Random Forest-Multimodal models, respectively). Our models also performed well in the multicentre validation dataset (AUROC 0.711, 0.785, and 0.879 for the ECG alone, CNN-Multimodal, and CNN-Random Forest-Multimodal models, respectively).

Conclusion

The multimodal AI algorithm, which integrated ECG images and clinical features, predicted the presence of LVAs with a higher degree of accuracy than ECG alone and the clinical LVA scores.

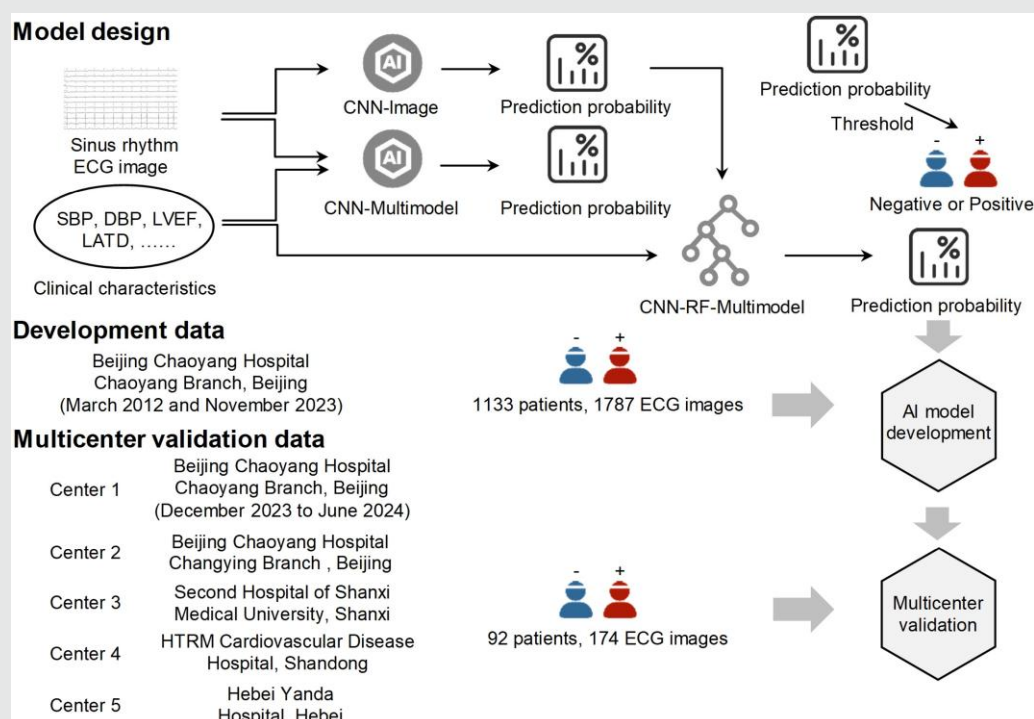
* Corresponding author. Tel: +86 0523 2789120, Email: xpliu71@hotmail.com (X.L.), Tel: +86 010 82806532, Email: hongshenda@pku.edu.cn (S.H.)

[†]These authors have contributed equally to this work.

© The Author(s) 2024. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Graphical Abstract



This figure provides an overview of the study. We developed three distinct artificial intelligence (AI) models specifically designed to determine the status of LVAs from sinus rhythm ECGs and clinical characteristics. These three AI models were tested in a multicentre cohort. CNN, convolutional neural network; RF, random forest; other abbreviations as in [Table 1](#).

Keywords

Atrial fibrillation • Low-voltage area • Electrocardiogram • Artificial intelligence algorithm

Introduction

Atrial fibrillation (AF) is the most prevalent sustained arrhythmia encountered in clinical practice. The 2024 European Society of Cardiology (ESC) guidelines for the management of AF recommend catheter ablation as a first-line treatment for patients with symptomatic AF.¹ However, the success rate of catheter ablation remains suboptimal, with ~22–33% of patients experiencing recurrence post-procedure.² Several studies demonstrated that left atrial (LA) low-voltage areas (LVAs) are significantly associated with AF recurrence post-ablation.^{3–5}

Identifying LVAs in patients with AF is crucial, as it not only aids in assessing the risk of recurrence post-catheter ablation but also enhances procedural outcomes. Studies have indicated that in AF patients with LVAs, incorporating LVA ablation with circumferential pulmonary vein isolation (CPVI) enhances the success rate of the procedure compared to CPVI alone.^{6,7} Therefore, pre-operative identification of LVAs permits the customization of individualized ablation strategies, thereby improving the overall success rate of ablation.

In recent years, growing studies highlight the potential for artificial intelligence enabled electrocardiogram (AI-ECG) in the diagnosis and management of cardiovascular disease, including AF, heart failure, and structural cardiomyopathy.^{8–11} In the current study, we aimed to evaluate the efficacy of AI algorithms based on sinus rhythm electrocardiograms (ECGs) alone or in combination with clinical characteristics for predicting LVAs in patients with AF.

Methods

Study population

We retrospectively identified 1133 patients with AF who underwent radio-frequency ablation between March 2012 and November 2023 at the Department of Cardiology, Beijing Chaoyang Hospital. These participants had a total of 1787 ECG records available for analysis. The data were divided into training ($n = 906$), validation ($n = 113$), and test sets ($n = 114$). [Figure 1](#) provides an overview of the patient selection process and dataset division.

We used the following inclusion criteria: paroxysmal or persistent AF, defined according to the 2024 ESC Guidelines for the management of AF¹²; availability of LA voltage maps; and availability of a minimum 10-s 12-lead ECG in sinus rhythm.

The DR-FLASH¹³ and APPLE¹⁴ scores were assessed based on specific parameters. For APPLE score, the factors considered were age > 65 years, estimated glomerular filtration rate (eGFR) < 60 mL/min/1.73 m², persistent AF, LA diameter ≥ 43 mm, and a left ventricular ejection fraction < 50%. The DR-FLASH included criteria such as diabetes mellitus, an eGFR < 90 mL/min/1.73 m², persistent AF, a LA diameter > 45 mm, age > 65 years, female gender, and hypertension. Baseline data, including demographic information, comorbidities, and echocardiographic parameters, were collected, and both the CHA2DS2-VASc¹⁵ and HAS-BLED¹⁶ scores were calculated. The eGFR was determined using the Cockcroft-Gault formula.¹⁷ This study was approved by the Institutional Ethics Committee and adhered to the Declaration of Helsinki principles.

The overall cohort consisted of 463 patients with persistent AF and 670 with paroxysmal AF. Of these, 47.4% ($n = 537$) had low-voltage areas

Table 1 Baseline characteristics of patients

| | LVA positive (n = 537) | LVA negative (n = 596) | P-value |
|--------------------------------------|---------------------------|---------------------------|------------------|
| Age (years) | 67.2 ± 10.1 | 61.6 ± 10.9 | <0.001 |
| Female, n | 310 | 164 | <0.001 |
| BMI (kg/m ²) | 26.3 ± 10.8 | 26.4 ± 12.5 | 0.792 |
| SBP (mmHg) | 130.7 ± 17.7 | 130.5 ± 17.2 | 0.438 |
| DBP (mmHg) | 76.9 ± 12.0 | 77.7 ± 10.9 | 0.124 |
| Hypertension, n | 364 | 356 | 0.005 |
| Diabetes, n | 135 | 136 | 0.366 |
| Prior stroke, n | 43 | 36 | 0.210 |
| CAD, n | 150 | 134 | 0.039 |
| HF, n | 42 | 39 | 0.421 |
| PsAF, n | 258 | 205 | <0.001 |
| Prior ablation | 104 | 76 | 0.003 |
| Post-ablation AAD use | 423 | 499 | 0.044 |
| LAD (mm) | 41.2 ± 6.3 | 39.5 ± 5.4 | <0.001 |
| LALD (mm) | 55.7 ± 6.7 | 53.2 ± 6.1 | <0.001 |
| LATD (mm) | 42.1 ± 5.9 | 40.1 ± 5.2 | <0.001 |
| RALD (mm) | 49.8 ± 6.3 | 48.4 ± 5.9 | <0.001 |
| RATD (mm) | 37.3 ± 5.3 | 36.4 ± 5.3 | 0.002 |
| LVEDD (mm) | 47.5 ± 4.9 | 48.2 ± 4.5 | 0.009 |
| LVESD (mm) | 30.4 ± 5.4 | 30.8 ± 5.4 | 0.114 |
| IVS (mm) | 10.4 ± 3.3 | 10.4 ± 3.1 | 0.461 |
| LVEF (%) | 64.3 ± 8.5 | 64.7 ± 8.4 | 0.272 |
| LA volume (mL) | 126.9 ± 56.2 | 120.8 ± 40.5 | 0.034 |
| LA surface area | 141.6 ± 38.6 | 143.4 ± 39.7 | 0.462 |
| LVA area (mm ²) | 22.4 ± 18.2 | 2.0 ± 2.4 | <0.001 |
| LVA (%) | 16.0 ± 12.3 | 1.3 ± 1.5 | <0.001 |
| CHA2DS2-VASc score | 2.8 ± 1.5 | 1.9 ± 1.4 | <0.001 |
| HAS-BLED score | 1.1 ± 0.8 | 1.1 ± 0.8 | 0.061 |
| DR-FLASH score | 3.3 ± 1.6 | 2.3 ± 1.5 | <0.001 |
| APPLE score | 1.7 ± 1.0 | 1.2 ± 1.1 | <0.001 |
| Creatinine (μmol/L) | 75.9 ± 41.9 | 76.4 ± 30.2 | 0.395 |
| eGFR < 60 mL/min/1.73 m ² | 94 | 70 | 0.007 |
| Mapping points | 661.8 ± 924.5 | 688.5 ± 693.7 | 0.582 |

AAD, anti-arrhythmic drugs; BMI, body mass index; CAD, coronary heart disease; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; HF, heart failure; IVS, interventricular septum; LA, left atrium; LAD, left atrial diameter; LALD, left atrial long diameter; LATD, left atrial transverse diameter; LVA, low-voltage area; LVEDD, left ventricular end-diastolic diameter; LVEF, left ventricular ejection fraction; LVESD, left ventricular end-systolic diameter; RALD, right atrial long diameter; RATD, right atrial transverse diameter; SBP, Systolic blood pressure; PsAF, persistent atrial fibrillation. Note: Bold values indicate statistically significant differences ($P < 0.05$).

(LVAs). The mean age of participants was 64.2 ± 10.9 years, with 41.8% being women. Baseline demographic data are summarized in [Table 1](#). Compared to participants without LVAs, those with LVAs were more likely to be older (67.2 ± 10.1 vs. 61.6 ± 10.9 years, $P < 0.001$), female (57.7% vs. 27.5%, $P < 0.001$), and have higher incidences of hypertension (67.8% vs. 59.7%, $P = 0.005$), coronary heart disease (27.9% vs. 22.5%, $P = 0.039$), and persistent AF (48.0% vs. 34.4%, $P < 0.001$). Additionally, patients with LVAs had larger LA diameter, larger LA volume (126.9 ± 56.2 vs. 120.8 ± 40.5 mL, $P = 0.034$), larger right atrial diameter, and smaller left ventricular

end-diastolic diameter (47.5 ± 4.9 vs. 48.2 ± 4.5 mm, $P = 0.009$). The LVA-positive group also had higher CHA2DS2-VASc scores (2.8 ± 1.5 vs. 1.9 ± 1.4 , $P < 0.001$), DR-FLASH scores (3.3 ± 1.6 vs. 2.3 ± 1.5 , $P < 0.001$), and APPLE scores (1.7 ± 1.0 vs. 1.2 ± 1.1 , $P < 0.001$). A higher proportion of LVA-positive patients had prior ablation (19.4% vs. 12.8%, $P = 0.003$), and 78.8% of them received post-procedural anti-arrhythmic drugs compared to 83.7% in the LVA-negative group ($P = 0.044$).

The mean LVA area in the LVA-positive group was 22.4 ± 18.2 mm², covering 16.0 ± 12.3 % of the LA surface. Interestingly, despite the more advanced disease characteristics in the LVA-positive group, creatinine levels between the two groups were similar (75.9 ± 41.9 vs. 76.4 ± 30.2 μmol/L, $P = 0.395$). However, the proportion of patients with eGFR < 60 mL/min/1.73 m² was significantly higher in the LVA-positive group (17.5% vs. 11.7%, $P = 0.007$). The number of mapping points was also similar between the two groups (661.8 ± 924.5 vs. 688.5 ± 693.7 , $P = 0.582$).

Additionally, a multicentre validation study was conducted, including 92 patients who underwent radiofrequency ablation for AF between December 2023 and June 2024 across five centres, contributing a total of 174 ECGs.

Electrocardiogram evaluation

A standard 12-lead ECG, with a duration of 10 s, was recorded at an amplitude of 10 mm/mV and a speed of 25 mm/s. For patients with paroxysmal AF, the ECG was recorded within 3 days before radiofrequency ablation in sinus rhythm, while for patients with persistent AF, the ECG was recorded within 3 days after radiofrequency ablation in sinus rhythm. All included individuals had at least one ECG recorded during sinus rhythm. We excluded ECGs with missing leads, significant interference, or those recorded during atrial pacing. A total of 430 patients had more than one ECG available.

Assessment of left atrial voltage maps

If the patient presents with sinus rhythm, LA anatomical reconstruction and LA voltage mapping are performed prior to ablation using a three-dimensional electroanatomical system (Carto 3, Biosense Webster, Diamond Bar, CA, USA) with a multi-electrode catheter (PentaRay, Biosense Webster, Diamond Bar, CA, USA). If the patient presents with AF, LA voltage mapping is performed again in sinus rhythm following radiofrequency ablation and conversion to sinus rhythm. The LA appendage and pulmonary veins were excluded from the LA geometry, while the mitral valve was delineated. LVAs were defined as regions with a voltage < 0.5 mV during sinus rhythm, covering more than 5% of the LA surface area (see [Supplementary material online, Figure S1](#)). The LVA percentage was defined as the LVA area divided by the LA surface area.

Ablation strategy

The ablation strategy for paroxysmal AF was CPVI. For patients with persistent AF, the ablation strategy has been previously detailed.¹⁸ To summarize, the initial approach also starts with CPVI. Following CPVI, all remaining patients received a low dose of ibutilide (0.004 mg/kg). Patients whose AF persisted were monitored for 30 min. No further ablation was done if AF converted to sinus rhythm within this period. In cases where persistent AF transitioned to atrial flutter or focal atrial tachycardia, the critical isthmus or arrhythmia focus was ablated until sinus rhythm was restored. For patients whose persistent AF was resistant to CPVI combined with low-dose ibutilide, atrial electrogram-based ablation was carried out until AF termination. The target areas identified for ablation exhibited the following characteristics (i) clusters of bipolar electrograms that displayed spatial dispersion and (ii) faster local atrial deflections compared to adjacent sites. If persistent AF remained unresponsive to CPVI and 30-min electrogram-guided ablation of the left atrium, electrical cardioversion was performed (see [Supplementary material online, Figure S2](#)).

Data pre-processing

In the present investigation, the primary dataset comprised of digitized facsimiles of conventional 12-lead ECGs obtained from a paper-based recorder. Each ECG recording was carefully curated to encapsulate a minimum of 10 s of waveform data, ensuring a comprehensive representation of cardiac electrical activity for analytical rigour.

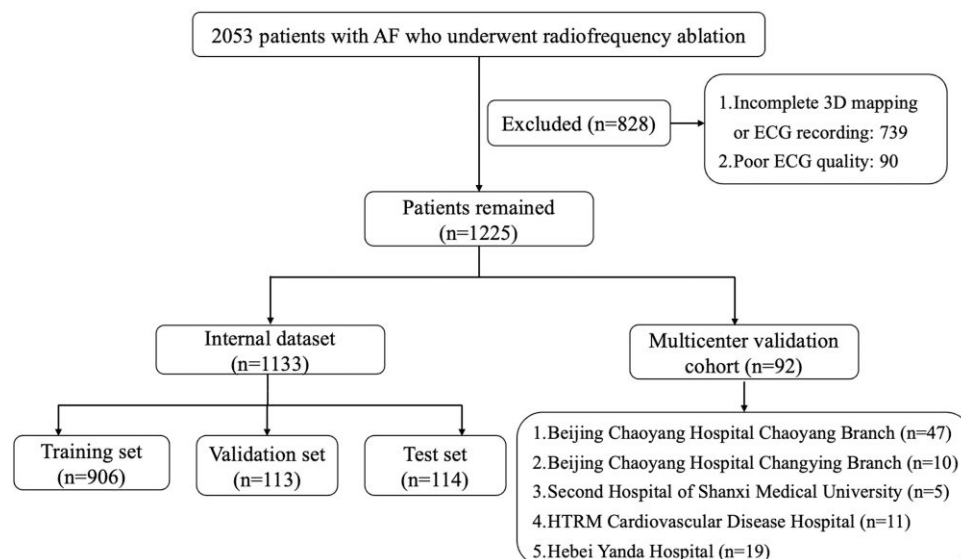


Figure 1 Flow diagram of the study. AF, atrial fibrillation; 3D, three dimensional; ECG, electrocardiogram; HTRM, heart rhythm.

The image processing in this study followed a structured protocol. Initially, each image was converted to greyscale to retain luminance information. The greyscale images were then binarized using a thresholding approach across the full intensity spectrum (0–255), resulting in 256 distinct binary images per original image. Pixels above the threshold were marked as features of interest (255), while those below were marked as background (0). Morphological transformations were subsequently applied to reduce noise, improving the signal-to-noise ratio. Finally, manual screening ensured that only representative and artefact-free images were included in the study (see [Supplementary material online, Figure S3](#)).

Model development

We have developed three distinct AI models specifically designed to aid cardiologists in determining LVAs status from sinus rhythm ECGs and clinical characteristics. This advanced system is engineered to integrate complex physiological data with clinical insights, thereby enhancing the accuracy of LVA detection and improving the efficacy of subsequent therapeutic interventions. In the realm of cardiovascular medicine, the precise identification of LVAs status is critical for informing clinical decision-making and optimizing treatment strategies.

Firstly, we employed a convolutional neural network (CNN) model, designated as CNN-Image, which is specifically trained on ECG images. This model exploits the spatial hierarchies and morphological patterns inherent to these visual representations. Secondly, we developed a hybrid AI model, termed CNN-Multimodal, which integrates ECG images with clinical features into the CNN architecture. This approach leverages the synergistic relationship between the visual ECG data and comprehensive clinical parameters, thereby augmenting the model's diagnostic accuracy. Thirdly, we introduced an advanced AI model, CNN-random forest (RF)-Multimodal, which combines CNN with RF algorithms to amalgamate the outputs of CNN-Image with clinical features for outcome prediction. This sophisticated methodology aims to harness the predictive insights from ECG image analysis and refine them through the integration of clinical data, resulting in a more nuanced and robust diagnostic framework.

In addition, CNN utilized ResNet101, and machine learning models employed RFs. The patients were randomly assigned to the training, validation, and internal test sets in an 8:1:1 ratio. As the ECG images and clinical features were linked to individual patients, all data from each patient were allocated to a single set, ensuring no overlap between the training, validation, and test sets.

Statistical analysis

Continuous data were presented as mean (\pm standard deviation) and compared using the Student's *t*-test or the Mann–Whitney *U* test. Categorical data were presented as numbers and compared using the χ^2 test. A *P*-value <0.05 was considered statistically significant. Data analysis was performed using SPSS version 29.0 software. The model performance was assessed using a receiver operating characteristic (ROC) curve.

Results

Prediction of low-voltage areas using the clinical risk scores

The APPLE LVA risk score had an area under the ROC curve (AUROC) of 0.589 for LVA discrimination, with an accuracy of 0.661, a sensitivity of 0.858, a specificity of 0.319, and an F1 score of 0.763. The DR-FLASH LVA risk score performed similarly, with an AUROC of 0.601, an accuracy of 0.556, a sensitivity of 0.433, a specificity of 0.768, and an F1 score of 0.553 ([Figure 2A](#), [Table 2](#)). These results indicate that both clinical scores offer moderate predictive value for LVAs, though their specificity remains limited.

Prediction of low-voltage areas using the electrocardiogram alone

Relying exclusively on paper ECG records, the CNN-Image model showed improved predictive power compared to the clinical scores, achieving an AUROC of 0.703, an accuracy of 0.635, a sensitivity of 0.857, a specificity of 0.396, and an F1 score of 0.709 ([Figure 2A](#), [Table 2](#)). The calibration curve demonstrated good linearity ([Figure 3A](#)), suggesting that the model's predictions were well-aligned with actual outcomes.

In addition, we visualized the ECG features learned by the CNN-Image using the Uniform Manifold Approximation and Projection dimensionality reduction technique. As shown in [Figure 3D](#), ECG features from patients with LVAs are more clustered compared to those without LVAs. To enhance the interpretability of

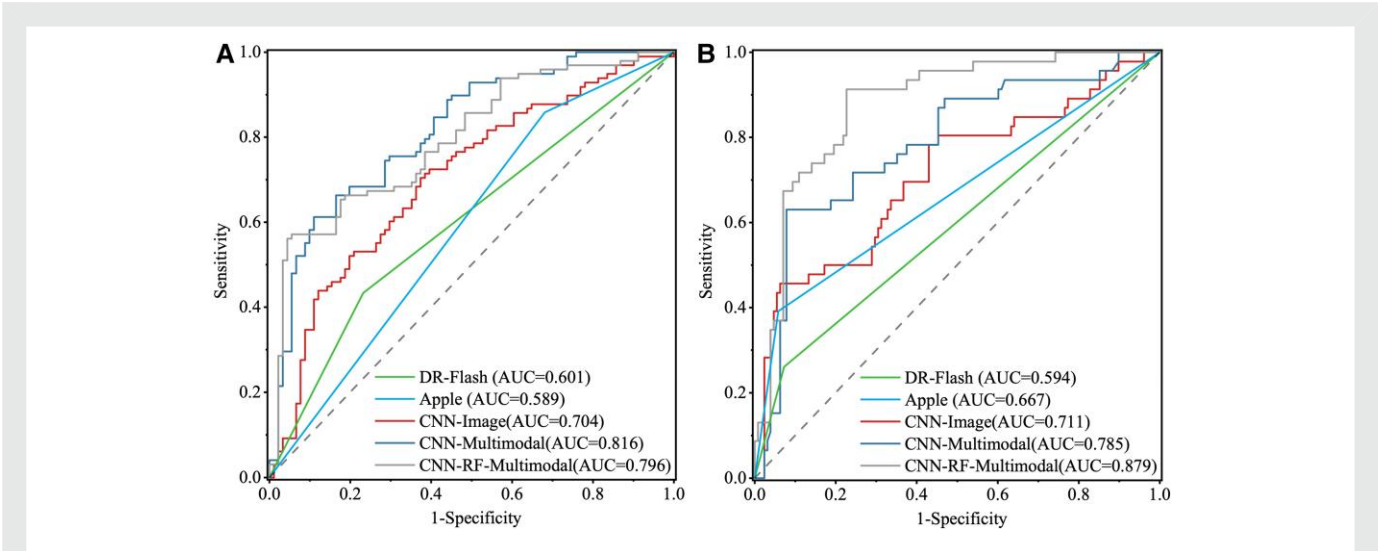


Table 2 Results of low-voltage areas prediction

| | AUROC | Sensitivity | Specificity | Accuracy | F1 score |
|-------------------|-------|-------------|-------------|----------|----------|
| Hold test | | | | | |
| DR-FLASH | 0.601 | 0.433 | 0.768 | 0.556 | 0.553 |
| APPLE | 0.589 | 0.858 | 0.319 | 0.661 | 0.763 |
| CNN-Image | 0.703 | 0.857 | 0.396 | 0.635 | 0.709 |
| CNN-Multimodal | 0.816 | 0.929 | 0.505 | 0.725 | 0.778 |
| CNN-RF-Multimodal | 0.796 | 0.939 | 0.429 | 0.693 | 0.760 |
| Extend test | | | | | |
| DR-FLASH | 0.594 | 0.261 | 0.928 | 0.761 | 0.353 |
| APPLE | 0.667 | 0.391 | 0.942 | 0.804 | 0.500 |
| CNN-Image | 0.711 | 0.804 | 0.391 | 0.500 | 0.460 |
| CNN-Multimodal | 0.785 | 0.934 | 0.234 | 0.420 | 0.460 |
| CNN-RF-Multimodal | 0.879 | 0.978 | 0.305 | 0.483 | 0.500 |

CNN, convolutional neural network; RF, random forest.

our model's predictions, we utilized Gradient-weighted Class Activation Mapping (Grad-CAM) to highlight the regions in the ECG images that were critical for predicting the presence of LVAs¹⁹ (Figure 4).

Prediction of low-voltage areas integrating electrocardiogram and clinical features

Supplementary material online, Figure S4 illustrates the importance scores of all 35 features, including the prediction LVAs probability by the CNN-Image model based on ECG alone (PLVA) and 34 clinical features. The importance score quantifies how much each feature improves the decision-making process of the model. The figure illustrates a rapid decline in the importance of features, with the least significant features approaching zero. To optimize predictive performance, the top 23 most important features were selected to construct the CNN-RF-Multimodal model, as these features yielded the highest

accuracy. The CNN-RF-Multimodal model based on these 23 features achieved an AUROC of 0.796, with an accuracy of 0.693, sensitivity of 0.939, specificity of 0.429, and an F1 score of 0.76 (Figure 2A, Table 2). This model outperformed both clinical scores in terms of AUROC, demonstrating that the combination of ECG and clinical data improves LVA prediction. The calibration curve demonstrated good linearity (Figure 3C). Additionally, we generated a SHAP (SHapley Additive exPlanations) plot depicting the top 23 features to enhance the interpretability of the CNN-RF-Multimodal model (Figure 5). This graphical representation clarifies the impact of individual variables on the model's diagnostic output, with higher SHAP values indicating a greater likelihood of LVAs occurrence.

In addition to the CNN-RF-Multimodal model, we also developed a CNN-Multimodal model that integrates ECG images and clinical features to evaluate its efficacy in predicting LVAs. Relying on paper ECG records and clinical features, the CNN-RF-Multimodal model achieved the highest AUROC of 0.816, with an accuracy of 0.725, a

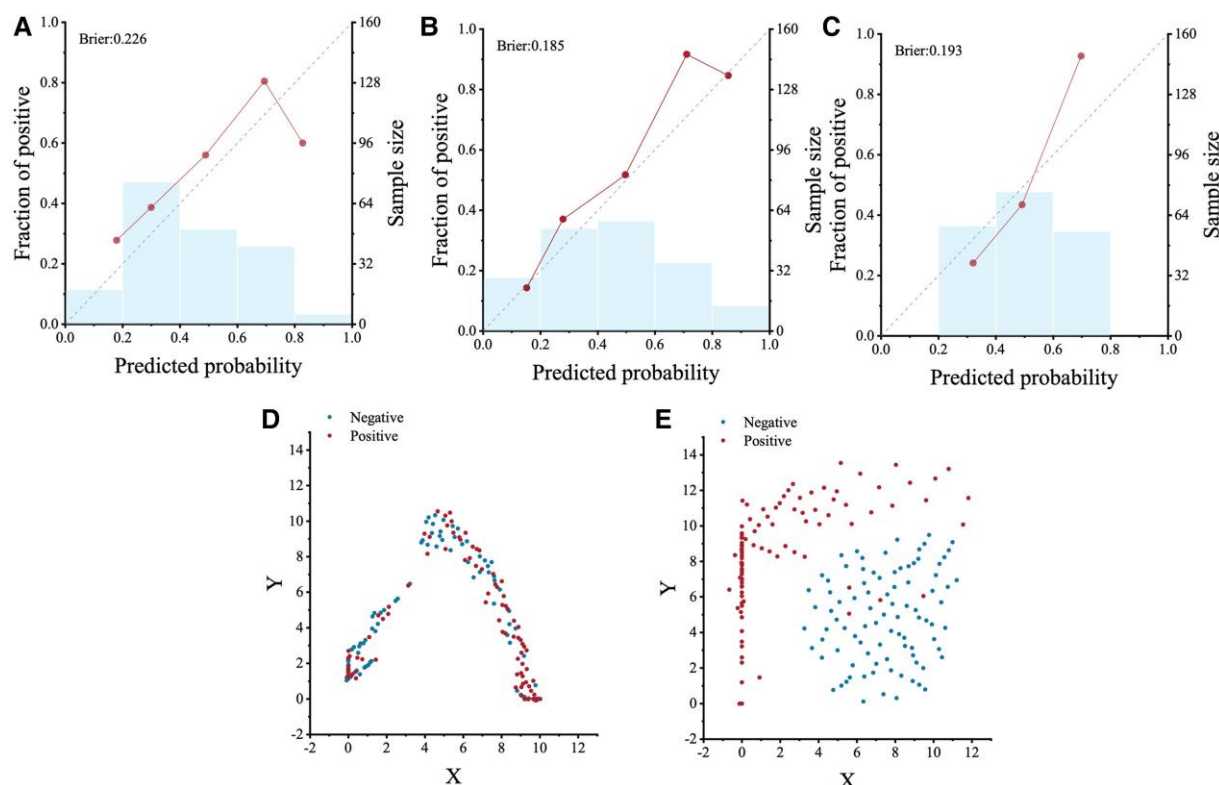


Figure 3 Model calibration curves for CNN-Image (A), CNN-Multimodal (B), and CNN-RF-Multimodal (C). Uniform Manifold Approximation and Projection visualization of electrocardiogram features learned by the CNN-Image model (D) and CNN-Multimodal model (E). CNN, convolutional neural network; RF, random forest.

sensitivity of 0.929, a specificity of 0.505, and an F1 score of 0.778 (Figure 2A, Table 2). The calibration curve demonstrated good linearity (Figure 3B). Uniform Manifold Approximation and Projection was used to visualize the ECG features learned by the CNN-Multimodal model. As shown in Figure 3E, the distinct separation of features highlights the model's ability to distinguish between patients with and without LVAs.

Multicentre validation

The multicentre validation included 92 patients from five centres, encompassing a total of 174 paper ECGs. Baseline characteristics of patients in the multicentre validation were shown in Supplementary material online, Table S1. Additionally, Supplementary material online, Table S2 provides a comparison of the clinical characteristics between the internal test cohort and the multicentre validation cohort. In the multicentre validation, the APPLE and DR-FLASH scores achieved AUROCs of 0.667 and 0.594, respectively, indicating moderate predictive ability in external cohorts. The CNN-Image model, relying solely on ECG data, performed better than the clinical scores, with an AUROC of 0.711, an accuracy of 0.5, a sensitivity of 0.804, a specificity of 0.391, and an F1 score of 0.460. The CNN-RF-Multimodal model and the CNN-Multimodal model that integrated ECG and clinical features demonstrated better predictive performance compared to using ECG alone. The CNN-RF-Multimodal model showed the highest predictive ability, with an AUROC of 0.879, a sensitivity of 0.978, and an F1 score of 0.5, despite a lower specificity of 0.305. Similarly, the CNN-Multimodal model also improved upon ECG alone, achieving an AUROC of 0.785, with a sensitivity of 0.934 (Figure 2B, Table 2).

Discussion

Main findings

In summary, our study demonstrated that the CNN-Image model, relying solely on ECG data, outperformed the clinical LVA scores in predicting LVAs in AF patients. Moreover, the CNN-Multimodal and CNN-RF-Multimodal models, which integrated both ECG and clinical data, exhibited significantly enhanced predictive performance for LVAs compared to the CNN-Image model.

The limitations in the current methods for the detecting low-voltage areas

Low-voltage areas, representing advanced AF substrate, are associated with LA fibrosis as indicated by late gadolinium enhancement cardiac magnetic resonance imaging.²⁰ Several studies suggest that various demographic and clinical characteristics, such as persistent AF, advanced age, impaired eGFR, female gender, heart failure, and enlarged left atrium, may facilitate the detection of LVAs.^{21,22} Based on these clinical characteristics, several scoring systems have been developed to predict LVAs in AF patients, including the DR-FLASH score,¹³ APPLE score,¹⁴ ZAQ score,²³ mAPPLE score,²¹ and ANP score.²⁴ However, these existing clinical scoring systems have been developed in relatively small cohorts, with the DR-FLASH score being the largest, including 238 AF patients. In our previous study, an AI-ECG model was developed to predict LVAs using ECGs recorded during AF. However, the study was limited by a relatively small sample size of 587 patients and lacked external validation.²⁵

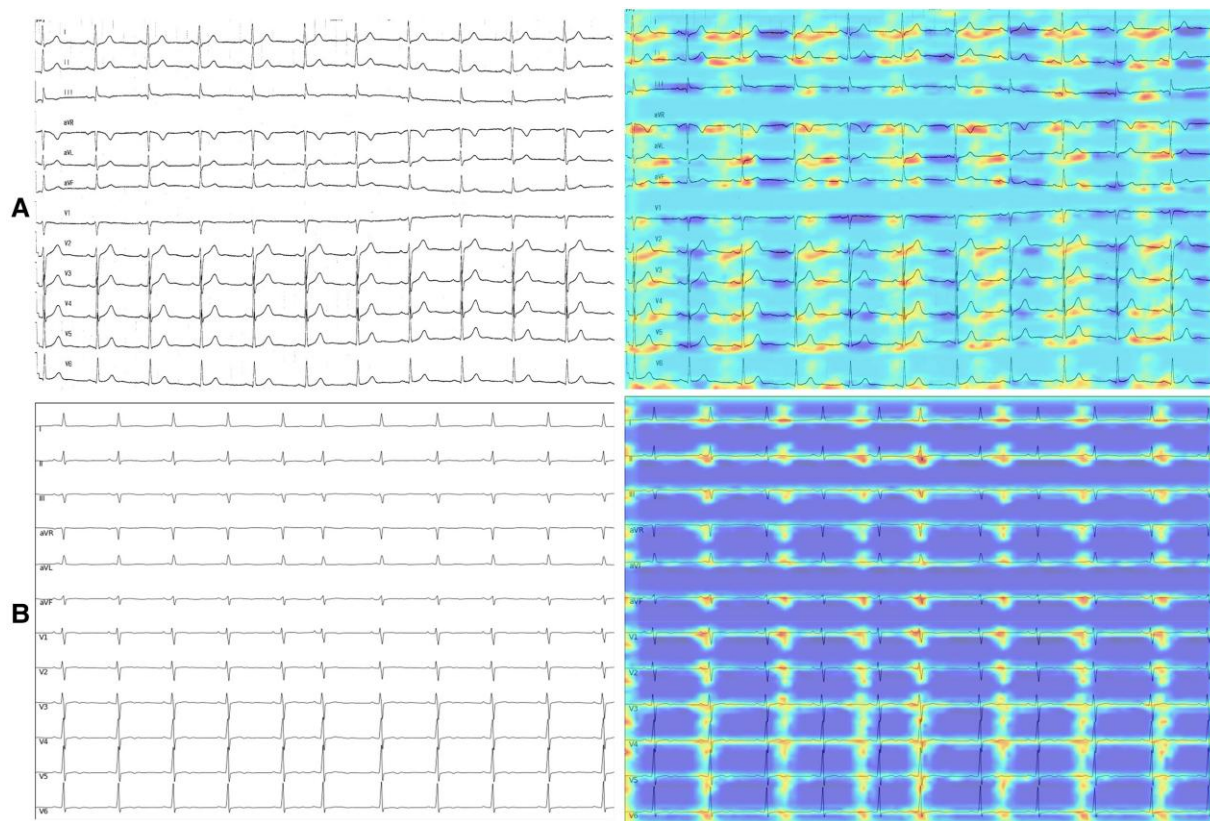


Figure 4 Comparison of electrocardiogram and heat map patterns between patients without (A) and with low-voltage areas (B). The heat map represent the importance of different regions to the model's prediction. Regions with higher importance, having a stronger influence on the model's decision, are represented with higher intensity, while regions with lower importance are shown with lower intensity, indicating less influence on the prediction.

The advantages of multimodal model for detecting low-voltage areas

Our study demonstrated that CNN models, particularly the CNN-RF-Multimodal and CNN-Multimodal models, exhibited higher sensitivity and lower false-negative rates in predicting LVAs in AF patients compared to clinical scoring systems. High sensitivity in detecting LVAs is clinically significant as it allows for earlier and more accurate identification of patients at risk for adverse outcomes. Early detection facilitates timely intervention, which can potentially improve patient outcome. Additionally, higher sensitivity minimizes the chance of false negatives, ensuring that fewer patients with LVAs are missed, thereby enhancing overall patient care and management.

Although there are currently several clinical scoring systems available to predict LVAs, our study stands out as the largest to date. We enrolled a total of 1225 AF patients, distributed across the training set, validation set, test set, and a multicentre validation set. This extensive inclusion of patients not only enhances the robustness of our findings but also ensures that our results are widely applicable across different clinical settings. Additionally, to the best of our knowledge, this study is unique in being the only multi-centre research conducted on an LVAs prediction model. This multicentre approach further strengthens the validity of our model and underscores its potential utility in diverse clinical environments.

In our study, a heat map was utilized to visualize the distribution and intensity of ECG features across different patient groups (Figure 5). The

Grad-CAM highlighted the significance of the P-wave in distinguishing between positive and negative predictions for LVAs. This finding is relatively easy to understand, as several studies demonstrated the association between LVAs and P-wave characteristics. Low-voltage areas have the potential to slow LA activation time, consequently impacting the duration, amplitude, and morphology of the P-wave.

When LVAs are present at a mild level, they can lead to an increase in P-wave duration. More extensive LVAs, on the other hand, can induce significant changes in P-wave morphology, such as the appearance of bi-phasic positive-negative P-waves in the inferior leads or the development of a 'late-terminal P' pattern.^{26,27} Additionally, LVAs can also affect P-wave amplitude. Park et al.²⁸ found that a P-wave amplitude of <0.1 mV in lead I is associated with LA remodelling and interatrial conduction block. Recent research has shown that the P-wave duration-amplitude ratio is positively correlated with the proportion of LVAs.²⁹ Interestingly, the heat map also revealed a relationship between the T-wave, atrial premature beat and LVAs. However, this correlation and its underlying mechanisms require further investigation to be clearly understood.

Clinical implications

In a recent large-scale study involving 1488 patients with AF, it was found that those with LVAs had a significantly higher risk of recurrence following radiofrequency ablation compared to patients without LVAs.³⁰ Additionally, the presence of LVAs was associated with poor

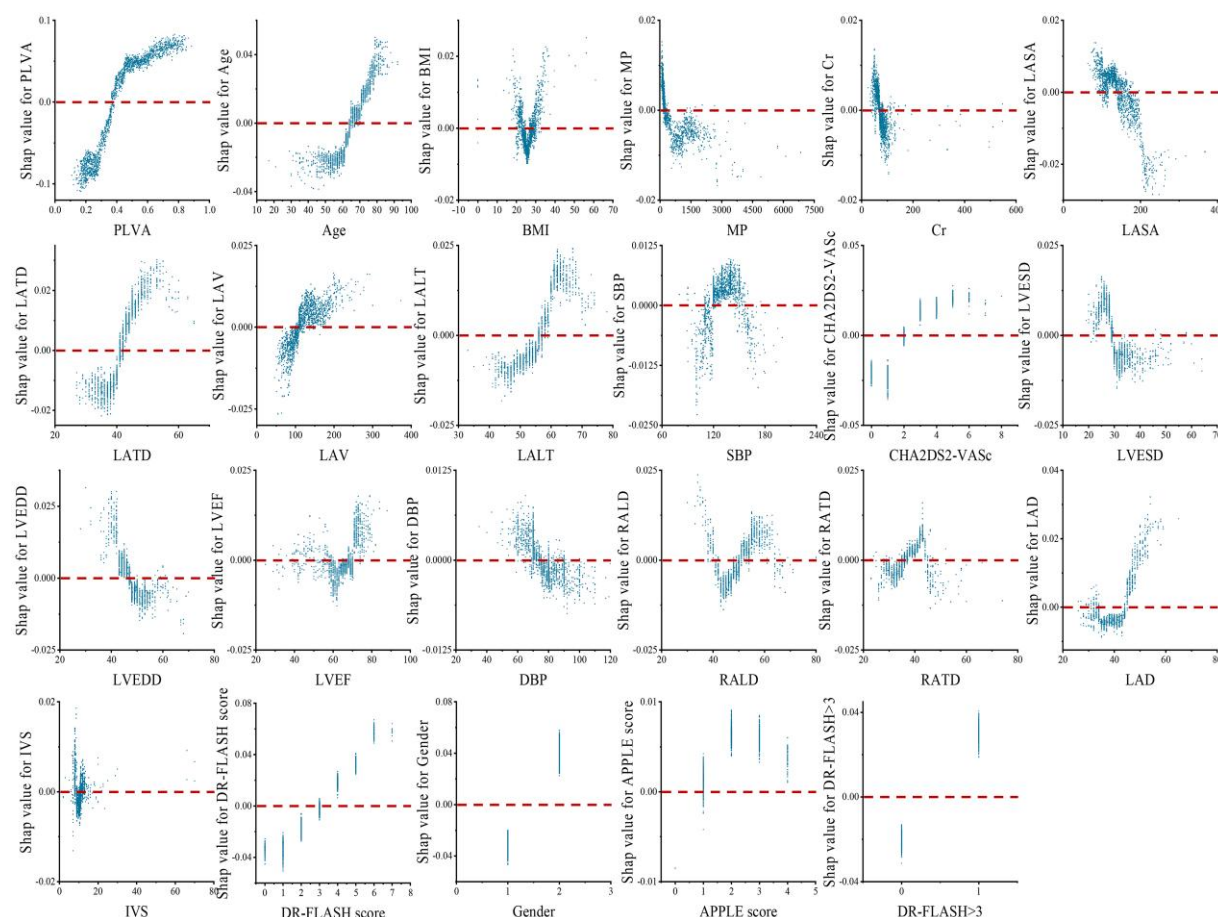


Figure 5 The SHapley Additive exPlanations plot of the 23 representative features for predicting low-voltage areas. Abbreviations as in [Table 1](#). CNN, convolutional neural network; Cr, creatinine; LASA, left atrial surface area; LAV, left atrial volume; MP, mapping points; PLVA, LVAs probability by the CNN-Image model based on electrocardiogram; other abbreviations as in [Table 1](#).

long-term composite endpoints, including death, heart failure, and stroke.³⁰ Although LVAs can increase the risk of AF recurrence, studies have shown that patients with LVAs who undergo additional LVAs ablation during procedure have similar success rates compared to those without LVAs.^{31,32}

The development of an AI model to predict LVAs in patients with AF holds significant clinical value. The model enables precise preoperative identification of patients with LVAs, facilitating more tailored and effective ablation strategy. For patients identified with LVAs, incorporating additional LVAs-targeted ablation into the CPVI can significantly enhance the success rates of the ablation procedure. This personalized strategy helps in reducing the likelihood of AF recurrence post-ablation. Moreover, LVAs is associated with an increased risk of AF recurrence and composite adverse events, including mortality, heart failure, and stroke. Therefore, patients with identified LVAs should undergo more intensive follow-up. Regular and systematic monitoring allows for the early detection of AF recurrence or complications, enabling timely and appropriate therapeutic interventions to mitigate these risks.

Limitations

This study has several limitations that should be acknowledged. First, the sample size is relatively small, with 1225 cases, which may limit the generalizability of the findings. Second, the study population

consists exclusively of Asian individuals, which restricts the applicability of the model to other ethnic groups. Third, the ECG data used in this study were all recorded using a 1 × 12 lead configuration, meaning the model cannot be applied to patients with different lead configurations, such as 2 × 6 or 3 × 4 leads. Future research will aim to address these limitations by validating the model in larger, more diverse populations and in both real-world and randomized controlled trial settings.

Conclusion

The multimodal AI algorithm, which integrates ECG images and clinical features, markedly enhances the accuracy of predicting LVAs in patients with AF compared to using ECG alone and traditional clinical scores. The robust performance of our models, validated across multicentre, underscores their broad applicability in diverse clinical settings. This innovative approach has the potential to significantly improve the clinical management and treatment outcomes for AF patients.

Supplementary material

[Supplementary material](#) is available at *European Heart Journal – Digital Health*.

Author contributions

Y.T. and D.Z. drafted the manuscript. N.P. and S.G. performed the literature search. N.P., T.Y., and C.T. collected the data. X.L. and S.H. checked and revised the manuscript. All authors read and approved the final version.

Funding

This study was supported by funds from the National Natural Science Foundation of China (Nos. 62102008 and 8227020788); Clinical Medicine Plus X—Young Scholars Project of Peking University, the Fundamental Research Funds for the Central Universities (PKU2024LCXQ030), the PKU-OPPO Fund (B0202301) and Hebei Science and Technology Project (22377785D).

Conflict of interest: none declared.

Data availability

Datasets included in this study are available from the corresponding author upon reasonable request.

References

- Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J* 2021;**42**:373–498.
- Brahier MS, Friedman DJ, Bahnson TD, Piccini JP. Repeat catheter ablation for atrial fibrillation. *Heart Rhythm* 2024;**21**:471–483.
- Starek Z, Di Cori A, Betts TR, Clerici G, Gras D, Lyan E, et al. Baseline left atrial low-voltage area predicts recurrence after pulmonary vein isolation: WAVE-MAP AF results. *Europace* 2023;**25**:euad194.
- Masuda M, Asai M, Iida O, Okamoto S, Ishihara T, Nanto K, et al. Additional low-voltage-area ablation in patients with paroxysmal atrial fibrillation: results of the randomized controlled VOLCANO trial. *J Am Heart Assoc* 2020;**9**:e015927.
- Chiang D, Sugumar H, Hunt A, Ling LH, Segal L, Al-Kaisey A, et al. Impact of posterior left atrial voltage on ablation outcomes in persistent atrial fibrillation: CAPLA substudy. *JACC Clin Electrophysiol* 2023;**9**:2291–2299.
- Chen H, Li C, Han B, Xiao F, Yi F, Wei Y, et al. Circumferential pulmonary vein isolation with vs without additional low-voltage-area ablation in older patients with paroxysmal atrial fibrillation: a randomized clinical trial. *JAMA Cardiol* 2023;**8**:765–772.
- Junarta J, Siddiqui MU, Riley JM, Dikdan SJ, Patel A, Frisch DR. Low-voltage area substrate modification for atrial fibrillation ablation: a systematic review and meta-analysis of clinical trials. *Europace* 2022;**24**:1585–1598.
- Khurshid S, Friedman S, Reeder C, Di Achille P, Diamant N, Singh P, et al. ECG-based deep learning and clinical risk factors to predict atrial fibrillation. *Circulation* 2022;**145**:122–133.
- Christopoulos G, Graff-Radford J, Lopez CL, Yao X, Attia ZI, Rabinstein AA, et al. Artificial intelligence-electrocardiography to predict incident atrial fibrillation: a population-based study. *Circ Arrhythm Electrophysiol* 2020;**13**:e009355.
- Akbilgic O, Butler L, Karabayir I, Chang PP, Kitzman DW, Alonso A, et al. ECG-AI: electrocardiographic artificial intelligence model for prediction of heart failure. *Eur Heart J Digit Health* 2021;**2**:626–634.
- Siontis KC, Liu K, Bos JM, Attia ZI, Cohen-Shelly M, Arruda-Olson AM, et al. Detection of hypertrophic cardiomyopathy by an artificial intelligence electrocardiogram in children and adolescents. *Int J Cardiol* 2021;**340**:42–47.
- Van Gelder IC, Rienstra M, Bunting KV, Casado-Arroyo R, Caso V, Crijns H, et al. 2024 ESC Guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). *Eur Heart J* 2024;**45**:3314–3414.
- Kosiuk J, Dinov B, Kornej J, Acou WJ, Schönbauer R, Fiedler L, et al. Prospective, multi-center validation of a clinical risk score for left atrial arrhythmogenic substrate based on voltage analysis: DR-FLASH score. *Heart Rhythm* 2015;**12**:2207–2212.
- Kornej J, Büttner P, Sommer P, Dagres N, Dinov B, Schumacher K, et al. Prediction of electro-anatomical substrate using APPLE score and biomarkers. *Europace* 2019;**21**:54–59.
- Friberg L, Rosenqvist M, Lip GY. Evaluation of risk stratification schemes for ischaemic stroke and bleeding in 182 678 patients with atrial fibrillation: the Swedish Atrial Fibrillation cohort study. *Eur Heart J* 2012;**33**:1500–1510.
- Pisters R, Lane DA, Nieuwlaar R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. *Chest* 2010;**138**:1093–1100.
- Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. *Nephron* 1976;**16**:31–41.
- He SN, Tian Y, Shi L, Wang YJ, Xie BQ, Li XX, et al. Identification of circumferential pulmonary vein isolation responders among patients with persistent atrial fibrillation: clinical value of the sequential low-dose ibutilide test. *Europace* 2020;**22**:1197–1205.
- Sangha V, Mortazavi BJ, Haimovich AD, Ribeiro AH, Brandt CA, Jacoby DL, et al. Automated multilabel diagnosis on electrocardiographic images and signals. *Nat Commun* 2022;**13**:1583.
- Zghaib T, Keramati A, Chrispin J, Huang D, Balouch MA, Ciuffo L, et al. Multimodal examination of atrial fibrillation substrate: correlation of left atrial bipolar voltage using multi-electrode fast automated mapping, point-by-point mapping, and magnetic resonance image intensity ratio. *JACC Clin Electrophysiol* 2018;**4**:59–68.
- Seewöster T, Kosich F, Sommer P, Bertagnoli L, Hindricks G, Kornej J. Prediction of low-voltage areas using modified APPLE score. *Europace* 2021;**23**:575–580.
- Masuda M, Fujita M, Iida O, Okamoto S, Ishihara T, Nanto K, et al. An E/e' ratio on echocardiography predicts the existence of left atrial low-voltage areas and poor outcomes after catheter ablation for atrial fibrillation. *Europace* 2018;**20**:e60–e68.
- D'Ambrosio G, Romano S, Althman O, Frommhold M, Borisov G, El Garhy M, et al. Computed tomography-derived left atrial volume index, sex, and age to predict the presence and the extent of left atrial low-voltage zones in patients with atrial fibrillation: the ZAQ score. *J Cardiovasc Electrophysiol* 2020;**31**:895–902.
- Seewöster T, Büttner P, Zeynalova S, Hindricks G, Kornej J. Are the atrial natriuretic peptides a missing link predicting low-voltage areas in atrial fibrillation? Introducing the novel biomarker-based atrial fibrillation substrate prediction (ANP) score. *Clin Cardiol* 2020;**43**:762–768.
- Tao Y, Zhang D, Tan C, Wang Y, Shi L, Chi H, et al. An artificial intelligence-enabled electrocardiogram algorithm for the prediction of left atrial low-voltage areas in persistent atrial fibrillation. *J Cardiovasc Electrophysiol* 2024;**35**:1849–1858.
- Müller-Edenborn B, Chen J, Allgeier J, Didenko M, Moreno-Weidmann Z, Neumann FJ, et al. Amplified sinus-P-wave reveals localization and extent of left atrial low-voltage substrate: implications for arrhythmia freedom following pulmonary vein isolation. *Europace* 2020;**22**:240–249.
- Jadidi A, Müller-Edenborn B, Chen J, Keyl C, Weber R, Allgeier J, et al. The duration of the amplified sinus-P-wave identifies presence of left atrial low voltage substrate and predicts outcome after pulmonary vein isolation in patients with persistent atrial fibrillation. *JACC Clin Electrophysiol* 2018;**4**:531–543.
- Park JK, Park J, Uhm JS, Joung B, Lee MH, Pak HN. Low P-wave amplitude (<0.1 mV) in lead I is associated with displaced inter-atrial conduction and clinical recurrence of paroxysmal atrial fibrillation after radiofrequency catheter ablation. *Europace* 2016;**18**:384–391.
- Zhang ZR, Ragot D, Massin SZ, Suszko A, Ha ACT, Singh SM, et al. P-wave duration/amplitude ratio quantifies atrial low-voltage area and predicts atrial arrhythmia recurrence after pulmonary vein isolation. *Can J Cardiol* 2023;**39**:1421–1431.
- Masuda M, Matsuda Y, Uematsu H, Sugino A, Ooka H, Kudo S, et al. Prognostic impact of atrial cardiomyopathy: long-term follow-up of patients with and without low-voltage areas following atrial fibrillation ablation. *Heart Rhythm* 2024;**21**:378–386.
- Yamaguchi T, Tsuchiya T, Nakahara S, Fukui A, Nagamoto Y, Murotani K, et al. Efficacy of left atrial voltage-based catheter ablation of persistent atrial fibrillation. *J Cardiovasc Electrophysiol* 2016;**27**:1055–1063.
- Zhou W, Wang L, Zhou B, Wu L. Catheter ablation of paroxysmal atrial fibrillation using high-density mapping-guided substrate modification. *Pacing Clin Electrophysiol* 2018;**41**:1630–1634.