Qiu et al. BMC Medical Informatics and Decision Making (2024) 24:225
https://doi.org/10.1186/s12911-024-02616-x

BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Deep learning-based multimodal fusion of the surface ECG and clinical features in prediction of atrial fibrillation recurrence following catheter ablation

Yue Qiu[1†], Hongcheng Guo[2†], Shixin Wang[1], Shu Yang[1], Xiafeng Peng[1], Dongqin Xiayao[1], Renjie Chen[2], Jian Yang[2], Jiaheng Liu[2], Mingfang Li[1], Zhoujun Li[2], Hongwu Chen[1] and Minglong Chen[1*]

Abstract

Background Despite improvement in treatment strategies for atrial fibrillation (AF), a significant proportion of patients still experience recurrence after ablation. This study aims to propose a novel algorithm based on Transformer using surface electrocardiogram (ECG) signals and clinical features can predict AF recurrence.

Methods Between October 2018 to December 2021, patients who underwent index radiofrequency ablation for AF with at least one standard 10-second surface ECG during sinus rhythm were enrolled. An end-to-end deep learning framework based on Transformer and a fusion module was used to predict AF recurrence using ECG and clinical features. Model performance was evaluated using areas under the receiver operating characteristic curve (AUROC), sensitivity, specificity, accuracy and F1-score.

Results A total of 920 patients (median age 61 [IQR 14] years, 66.3% male) were included. After a median follow-up of 24 months, 253 patients (27.5%) experienced AF recurrence. A single deep learning enabled ECG signals identified AF recurrence with an AUROC of 0.769, sensitivity of 75.5%, specificity of 61.1%, F1 score of 55.6% and overall accuracy of 65.2%. Combining ECG signals and clinical features increased the AUROC to 0.899, sensitivity to 81.1%, specificity to 81.7%, F1 score to 71.7%, and overall accuracy to 81.5%.

Conclusions The Transformer algorithm demonstrated excellent performance in predicting AF recurrence. Integrating ECG and clinical features enhanced the models' performance and may help identify patients at low risk for AF recurrence after index ablation.

Keywords Deep learning, Transformer, Atrial fibrillation recurrence, Electrocardiogram, Clinical features, Pulmonary vein isolation

[†]Yue Qiu and Hongcheng Guo contributed equally to this work.

*Correspondence:
Minglong Chen
chenminglong@njmu.edu.cn

[1]Department of Cardiology, The First Affiliated Hospital of Nanjing Medical University, Guangzhou Road, Nanjing 210029, Jiangsu Province, China
[2]State Key Lab of Software Development Environment, Beihang University, Beijing 100191, China

Background

Atrial fibrillation (AF) is the most common cardiac arrhythmia associated with increased morbidity and mortality [1, 2]. The recent updated guideline emphasizes a multifaceted approach to AF management, focusing on risk assessment, therapeutic strategies, and the integration of new evidence to optimize patient outcomes [3]. For symptomatic AF patients, catheter ablation is highlighted as a beneficial treatment, particularly for those with symptomatic paroxysmal AF and few comorbidities [3, 4]. Despite improvements in technology, the single-procedure success rate was as high as 72.5–75.9% for paroxysmal AF (PAF) and 50–60% for non-paroxysmal AF (non-PAF) [5, 6]. Our recent study demonstrated that female tend to exhibit more advanced structural remodeling at the time of ablation, which is associated with lower atrial voltage compared to men. This advanced remodeling in female often results in less favorable ablation outcomes [7]. Consequently, it is necessary to choose a more optimal strategy to identify patients who are more likely to benefit from ablation. Several clinical scores have been developed to predict success of ablation with modest performance [8, 9].

Deep learning (DL) is a subfield of machine learning (ML) that is capable of extracting complex patterns from data without requiring manual, expert-dependent feature engineering [10]. The Transformer, a state-of-the-art DL model, has gained prevalence in natural language processing (NLP) and Computer Vision using the attention mechanism [11]. Recently, the Transformer has demonstrated promising results in medical imaging applications, as it is adept at understanding contextual information [12, 13]. Previous studies using ML to predict AF recurrence by directly predicting shape descriptors from magnetic resonance imaging (MRI) [14]. Additionally, ML methods have been integrated with personalized computational modeling to predict recurrence following PVI [15]. Furthermore, handcrafted features extracted from computerized tomography (CT) scans have been linked to the probability of AF recurrence after ablation [16]. The electrocardiogram (ECG) contains a large amount of information that directly reflects underlying cardiac physiology associated with cardiac electrical and structural variations. A recent study showed that Convolutional Neural Networks (CNNs) and multimodal fusion framework can be used to predict AF recurrence using ECG, electrogram (EGM) and clinical features [17].

In this study, we propose a novel AF recurrence prediction algorithm based on Transformer using surface ECG signals and clinical features to identify patients who are more likely to benefit from index ablation.

Methods

Study population

Between October 2018 and December 2021, a total of 1,264 patients with either PAF (790, 65.2%) or non-PAF (474, 34.8%) who underwent first ablation were recruited. Non-PAF includes persistent and long-standing AF, persistent AF is defined as a sustained episode lasting > 7 days and < 1 year and long-standing AF is > 1 year and < 3 years. Patients with the following conditions were excluded: previous history of AF ablation, absence of data, poor ECG quality, and loss of follow-up (Fig. 1). This study was approved by the Ethics Committee of The First Affiliated Hospital of Nanjing Medical University.

ECG pre-processing

For each patient, a standard 12-lead ECG of 10 s during sinus rhythm (SR) were obtained (filter range 0.5–100 Hz; AC filter 50 Hz, 25 mm/s, 10 mm/mV) by ECG recording system (RAGE-12, Nalong Technology Co., Ltd., Xiamen, China). Preoperative ECG during SR was obtained for all PAF patients. For non-PAF patients, the ECG during SR was obtained within three days following the ablation procedure to minimize the effect of potential cardiac structural changes. When multiple ECG recordings were available, the one closest to the index procedure was selected. Details of the pre-processing of the surface ECG can be found in the Supplemental Methods.

Preprocessing of clinical variables

The following demographic variables were collected: age at the time of ablation, sex, height, weight, body mass index (BMI). Comorbidities including hypertension, diabetes, heart failure, stroke/transient ischemic attack (TIA), coronary artery disease (CAD) were included. Left atrial anteroposterior diameter (LAD), left ventricular ejection fraction (LVEF) and left ventricular end diastolic diameter (LVDd) were extracted from transthoracic echocardiograms that were obtained within 3 months before index ablation procedure. Pro-B type natriuretic peptide (Pro-BNP) and e-GFR, ml/(min 1.73 m^2) were obtained within 1 week before the first ablation procedure. The simple deletion method was used to handle missing values, which is the most primitive method for the treatment of missing values. Details regarding the pre-processing of the clinical features are provided in the Supplemental Methods. The number of missing values for each clinical variable are presented in Table S3.

Transformer network

The Transformer model was developed to predict the likelihood of AF recurrence by analyzing ECG signals and clinical data. The Transformer encoder, as outlined in Figure S1 (see Methods in the Data Supplement for further details), was employed to extract pertinent

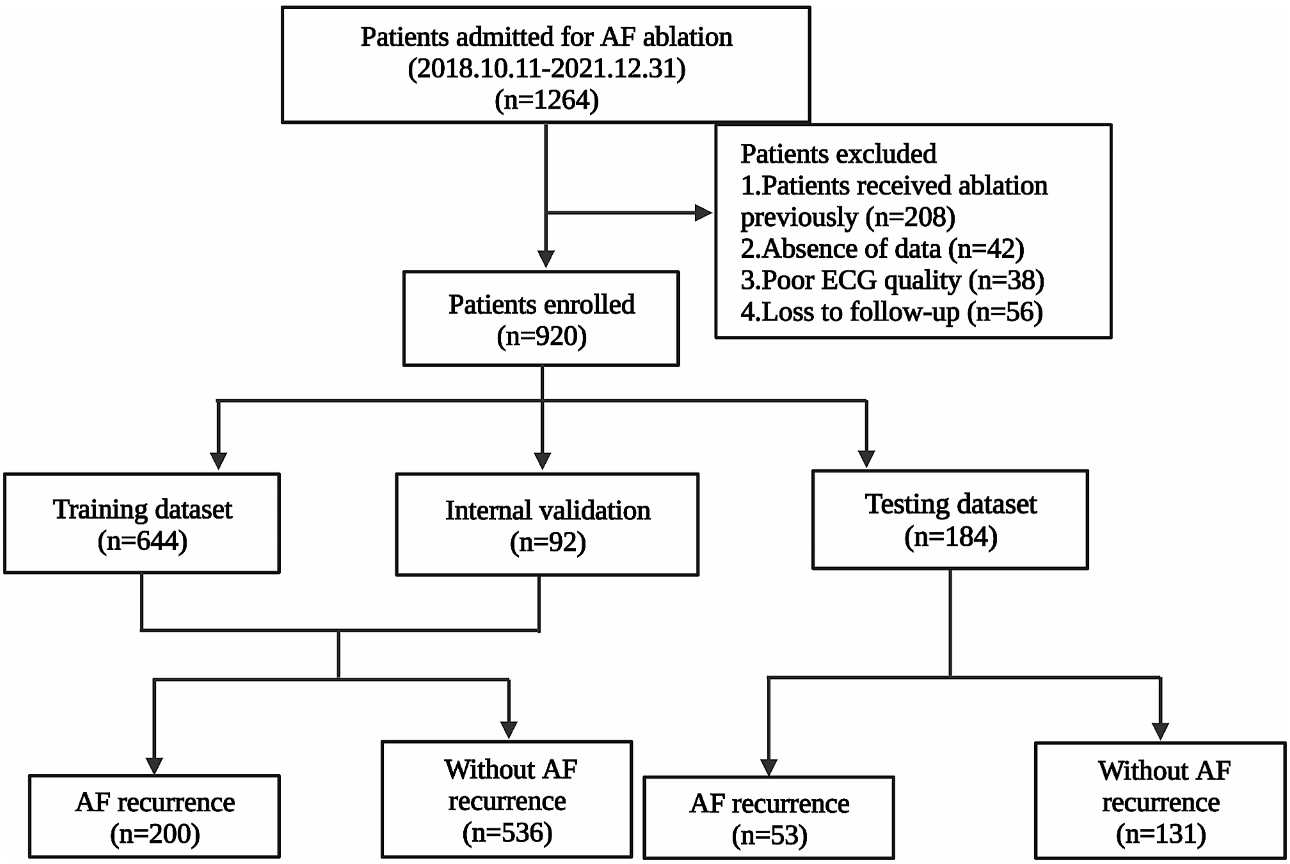

Fig. 1 The flowchart for inclusion and exclusion criteria

information from both data sources. Briefly, the Transformer encoder consists of various modules, including the multi-head attention module, residual module, fully-connected layer and the normalization layer. Training details of the Transformer model, such as the loss function and accuracy, are provided in Figure S3.

Fusion module

To enhance the model's performance, we developed a multimodal fusion framework (Figure S2). The pre-processed ECG images were first sent to a Multilayer Perceptron (MLP) layer for linear coding, which downscaled the relevant data before being sent to the Transformer encoder for further encoding. After six layers of encoding, the features were sent to the MLP for scaling. Similarly, the clinical data was first encoded using one-hot encoding and sent to a MLP layer, followed by the Transformer encoder for encoding. After six layers of encoding, the features were also sent to the MLP for scaling. In this way, the image and clinical data features were in the same dimension before fusion module. Therefore, we can directly concatenate these two vectors in the fusion module. Merged features were fed to the MLP classifier for further prediction.

Model training and validation

All patients and ECGs in SR were randomly assigned in a 7:1:2 ratio to training, internal validation, and testing datasets. The training and evaluation were conducted on 4 NVIDIA RTX V100 (16G). The maximum training epoch was set to 40, with a batch size of 32. We used 6 encoder layers for each modality and set the number of attention heads in all Transformer layers to 12. Adam optimizer was employed with a learning rate of 2e-6, while the dropout and label-smoothing rates were set to 0.3 and 0.1, respectively. To determine the most critical clinical characteristics that the model primarily relies on to derive the final prediction, the importance of each variable is determined by a self-attentive mechanism calculation. A probability threshold was selected based on the ROC curve of the internal validation set.

Ablation procedure

All antiarrhythmic drugs were discontinued for 5 half-lives and amiodarone for 2 months before the index procedure. The ablation procedure was performed as previously described in detail [18, 19]. In brief, a 3-D mapping system (Biosense Webster.; Diamond Bar, CA, USA) was used to guide the mapping and ablation procedures. Circumferential pulmonary vein isolation (CPVI) was initially performed by point-to-point ablation with an ablation index (AI) of 500 to 550 (power 30 to 40 w, contact force 5 to 30 g). If AF persisted after CPVI, SR was restored by electrical cardioversion. After the CPVI procedure, substrate mapping was applied with LA body and left atrial appendage during SR. Low-voltage areas (LVAs) were defined as areas with amplitude less than 0.5 mV in more than 3 adjacent low voltage points with space difference of 0.5 cm. The detailed techniques for substrate modification have been described previously [18, 20]. In brief, the intervention steps included homogenization ablation in the LVAs, defragmentation in the transitional zones (TZs), and dechanneling of the substrate if necessary.

Follow-up

All patients were treated with anticoagulation for the first 3 months and anti-arrhythmic drugs if not contraindicated. Follow-up visits were scheduled for the patients at 1, 3, 6, and 12 months after index ablation, and at least once a year after that. During these visits, patients underwent a surface ECG and 24-h Holter monitoring. AF recurrence was defined as any episode lasting longer than 30 s on Holter recordings or ECG during clinical visits after index ablation procedure.

Statistical analysis

Continuous variables were presented as mean ± standard deviation (SD) or median (interquartile range [IQR]) and compared between groups using the Student t test or Mann-Whitney U test. Categorical variables were expressed as percentages and compared using the X^2-test. Multivariate Cox proportional hazards regression was performed including variables with $P<0.1$ on univariate analysis. Kaplan-Meier (KM) incidence free survival analysis was performed in the testing set. Statistical analyses were performed using R software (version 3.3.0) and python (version 3.6.8).

Results

Population characteristics

Between October 2018 and December 2021, a total of 1,264 patients who underwent catheter ablation for AF were recruited. Of these, 208 patients with a history of AF ablation were excluded, 38 patients were excluded due to poor ECG quality and 56 patients were lost to follow-up. Finally, a total of 920 patients were enrolled in the study (Fig. 2). Demographics and clinical data of the study population are summarized in Table 1. The median age was 61 years and PAF accounted for 580 (63.0%) patients. CPVI alone was performed in 430 patients (46.7%) while 490 patients (53.3%) underwent additional LVA ablation. After a median follow-up of 24 months, 253 patients (27.5%) experienced AF recurrence. Patients with recurrent AF were predominantly female ($P=0.009$), had a history of stroke/TIA ($P=0.005$) and had a larger LAD ($P=0.002$). On multivariate analysis, both female sex (HR 1.36, 95% CI 1.02–1.83, $P=0.037$) and LAD (HR 1.03, 95% CI 1.00–1.06, $P=0.046$) proved to be independent predictors of AF recurrence (Table 2).

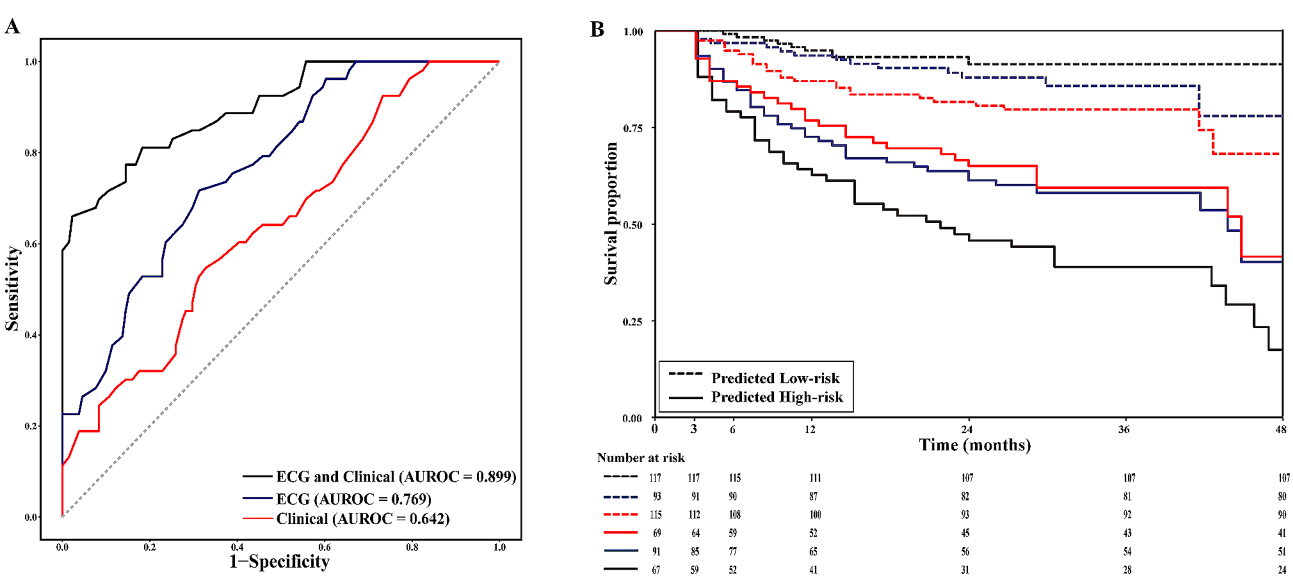

Fig. 2 Receiver operating characteristic curves and incidence-free KM survival curves for the 3 models. The 3 models are transformer model with clinical features only (Trans-Clinical red); transformer model with ECG only (Trans-ECG; blue); and transformer model with ECG and clinical (Trans-ECG-Clinical; black) for the testing cohort. (**A**) ROC curves of the testing cohort for the 3 models. (**B**) Incidence-free KM curves for the high- and low-risk groups of the 3 models after a median follow-up of 24 months. Trans indicates transformer; ROC: receiver operating characteristic; AUROC: areas under the receiver operating characteristic curve; KM: Kaplan-Meier

Table 1 Baseline characteristics of the study population

Characteristics	All ($n=920$)	No AF Recurrence ($n=667$)	AF Recurrence ($n=253$)	*P*Value
Age, years	61.0 (54.0,68.0)	61.0 (54.0,67.0)	61.0 (54.0,68.0)	0.359
Male, n (%)	610 (66.3)	459 (68.8)	151 (59.7)	0.009
BMI, kg/m^2	25.0 (23.1,27.1)	24.9 (23.2,27.0)	25.2 (22.9,27.4)	0.496
Medical history				
Hypertension, n (%)	468 (50.9)	327 (49.0)	141 (55.7)	0.069
Diabetes, n (%)	125 (13.6)	94 (14.1)	31 (12.3)	0.467
HF, n (%)	172 (18.7)	116 (17.4)	56 (22.1)	0.099
Stroke/TIA, n (%)	75 (8.5)	44 (6.6)	31 (12.3)	0.005
CAD, n (%)	66 (7.2)	43 (6.4)	23 (9.1)	0.165
CHA_2DS_2-VASc score	1.0 (1.0,3.0)	1.0 (1.0,3.0)	2.0 (1.0,3,0)	0.001
0, n (%)	186 (20.2)	148 (22.2)	38 (15.0)	
1, n (%)	279 (30.3)	210 (31.5)	69 (27.3)	
2, n (%)	196 (21.3)	134 (20.1)	62 (24.5)	
3, n (%)	144 (15.7)	100 (15.0)	44 (17.4)	
> 3, n (%)	115 (12.5)	75 (11.2)	40 (15.8)	
Procedure				0.318
CPVI alone, n (%)	430 (46.7)	305 (45.7)	125 (49.4)	
CPVI + LVA, n (%)	490 (53.3)	362 (54.3)	128 (50.6)	
Periprocedural complications				
Cardiac tamponade, n (%)	6 (0.7)	2 (0.3)	4 (1.6)	0.090
Pseudoaneurysm, n (%)	4 (0.4)	3 (0.4)	1 (0.4)	1.000
Echocardiogram parameters				
LAD, mm	40.0 (36.0,43.0)	39.0 (36.0,42.0)	40.0 (37.0,44.0)	0.002
LVDd, mm	48.0 (45.0,50.0)	48.0 (45.0,50.0)	48.0 (45.0,51.0)	0.476
LVEF, %	63.0 (61.4,64.4)	63.0 (61.0,64.7)	62.7 (61.7,64.4)	0.483
Biomarkers				
Pro-BNP	409.0 (152.0,855.0)	347.0 (138.0,823.0)	496.2 (206.0,935.0)	0.002
Dyslipidemia, n (%)	381 (41.4)	287 (43.0)	94 (37.2)	0.106
e-GFR, ml/(min 1.73 m^2)	90.0 (71.4,109.3)	90.5 (71.4,109.3)	89.7 (73.5,111.8)	0.995
AF type				0.018
Paroxysmal	580 (63.0)	436 (65.4)	144 (56.9)	
Non-paroxysmal	340 (37.0)	231 (34.6)	109 (43.1)	

AF, atrial fibrillation; BMI, body mass index; HF, heart failure; TIA, transient ischemic attack; CAD, coronary artery disease; CHA_2DS_2-VASc score, congestive heart failure, hypertension, age≥75 years (doubled), diabetes, stroke/transient ischemic attack/thromboembolism (doubled), vascular disease (prior myocardial infarction, peripheral artery disease, or aortic plaque), age 65–75 years, sex category (female); CPVI, circumferential pulmonary vein isolation; LVA, low-voltage area; LAD, left atrial diameter; LVDd, left ventricular diastolic diameter; LVEF, left ventricular ejection fraction; Pro-BNP, Pro-Brain Natriuretic Peptide; e-GFR, estimated of glomerular filtration rate

Clinical scores

To access the predictive ability of clinical risk scoring methods for AF recurrence, we compared the predictive performance of the CHA_2DS_2-VASc Score [9], APPLE score [21] and the DR-FLASH score [22]. The AUROCs for these models were 0.538, 0.563 and 0.548, respectively, demonstrating inferiority to the Transformer model (Table 3).

Prediction of AF recurrence using ECG or clinical features alone

The receiver operating characteristic (ROC) and KM curves for the three different models are illustrated in Fig. 2. The Transformer models demonstrated HRs of 3.9 (95% CI, 2.14–7.11) and 1.92 (95% CI, 1.11–3.31) in Trans-ECG and Trans-Clinical, respectively. The predictive performance of the Transformer models for all patients is presented in Table 3. In the model developed and tested based only on surface ECG, the AUROC was 0.769, which out-performs the performance of the CHA_2DS_2-VASc Score, APPLE score and the DR-FLASH score. The AUROC of the model using clinical features alone in the testing dataset was 0.642 with a sensitivity and specificity of 52.8% and 68.7%, respectively. Feature selection algorithm rankings are provided in Fig. 3. The top five most important features are LAD, AF type, BMI, stroke/TIA and age.

Table 2 Univariate and multivariate analysis

Characteristics	Univariable			Multivariable		
	HR	95%CI	*P* value	HR	95%CI	*P* value
Age	1.00	0.99 to 1.02	0.503			
Female sex	**1.37**	**1.06 to 1.76**	**0.015**	**1.36**	**1.02 to 1.83**	**0.037**
BMI	1.01	0.97 to 1.05	0.708			
Hypertension	1.22	0.95 to 1.56	0.120			
Diabetes	0.85	0.58 to 1.23	0.387			
HF	1.26	0.94 to 1.70	0.124			
Stroke/TIA	**1.78**	**1.22 to 2.59**	**0.003**	1.43	0.90 to 2.27	0.131
CAD	1.40	0.91 to 2.15	0.126			
CHA_2DS_2-VASc score	**1.13**	**1.05 to 1.22**	**0.002**	1.02	0.91 to 1.13	0.762
CPVI + LVA	0.92	0.72 to 1.17	0.487			
LAD	**1.04**	**1.02 to 1.07**	**0.002**	**1.03**	**1.00 to 1.06**	**0.046**
LVDd	1.01	0.98 to 1.05	0.384			
Pro-BNP	1.00	1.00 to 1.00	0.092			
Dyslipidemia	0.79	0.61 to 1.02	0.071			
e-GFR	1.00	1.00 to 1.00	0.826			
Non-PAF	**1.35**	**1.05 to 1.73**	**0.019**	1.17	0.88 to 1.55	0.281

BMI: body mass index; HF: heart failure; TIA: transient ischemic attack; CAD: coronary artery disease; CPVI: circumferential pulmonary vein isolation; LVA: low-voltage area; LAD: left atrial diameter; LVDd: left ventricular diastolic diameter; Pro-BNP: Pro-Brain Natriuretic Peptide; e-GFR: estimated of glomerular filtration rate

Table 3 The performance of the clinical scores and transformer models to predict AF recurrence

	AUROC	Sensitivity	Specificity	Accuracy	F1-score
CHA_2DS_2-VASc Score	0.538	0.585	0.496	0.522	0.413
DR-FLASH score	0.548	0.623	0.443	0.495	0.415
APPLE score	0.563	0.151	0.954	0.723	0.239
ECG	0.769	0.755	0.611	0.652	0.556
Clinical features	0.642	0.528	0.687	0.641	0.459
ECG and clinical features	0.899	0.811	0.817	0.815	0.717

Prediction of AF recurrence combining ECG and clinical features

Combining ECG and clinical features in the Transformer model resulted in an improved AUROC of 0.899, outperforming the performance of the model trained on ECG or clinical features alone (Table 3). Cox regression analysis indicated that high-risk patients had a significantly higher incidence of AF recurrence during the follow-up period than low-risk patients (HR: 32.2; 95% CI: 7.83–132.5; $P<0.001$). The transformer model demonstrated superior performance compared to both the CNN approach trained solely on surface ECG and the CNN approach trained using a combination of surface ECG and clinical features (Table S4). Figure S3 demonstrates representative examples of the fusion model's performance in patients with and without AF recurrence.

Subgroup analysis

The performance of the fusion model in subgroup analysis is presented in Fig. 4. The fusion model exhibited similar performance among male and female patients, but it demonstrated exceptional performance in older (≥65 years old) patients (sensitivity 90.5%; specificity 84.5%; HR: 24.3; 95% CI: 5.6 to 104.8). Regarding PAF patients, the Transformer model yielded AUROCs of 0.782 using ECG alone and 0.889 with the incorporation of clinical features. For non-PAF patients, the model resulted in AUROCs of 0.924 and 0.751 with and without the addition of clinical features, respectively (Table S1) The model performance was similar among patients with prior-ablation ECG and those with post-ablation ECG (Table S3).

Discussion

The main findings of this study are as follows: 1), a novel Transformer-based end-to-end approach demonstrated high predictive performance for the incidence of AF recurrence. 2), combining surface ECG and clinical variables significantly improved the performance of the model. 3), the Transformer model showed promise in identifying patients at low risk of AF recurrence after index ablation.

DL is a rapidly evolving and dynamic field of computer science that has garnered significant attention in recent years [23]. Recently, the transformer architecture has emerged as a state-of-the-art technique in NLP tasks that involve sequential input data. This has led to breakthroughs in tasks such as language translation, sentiment

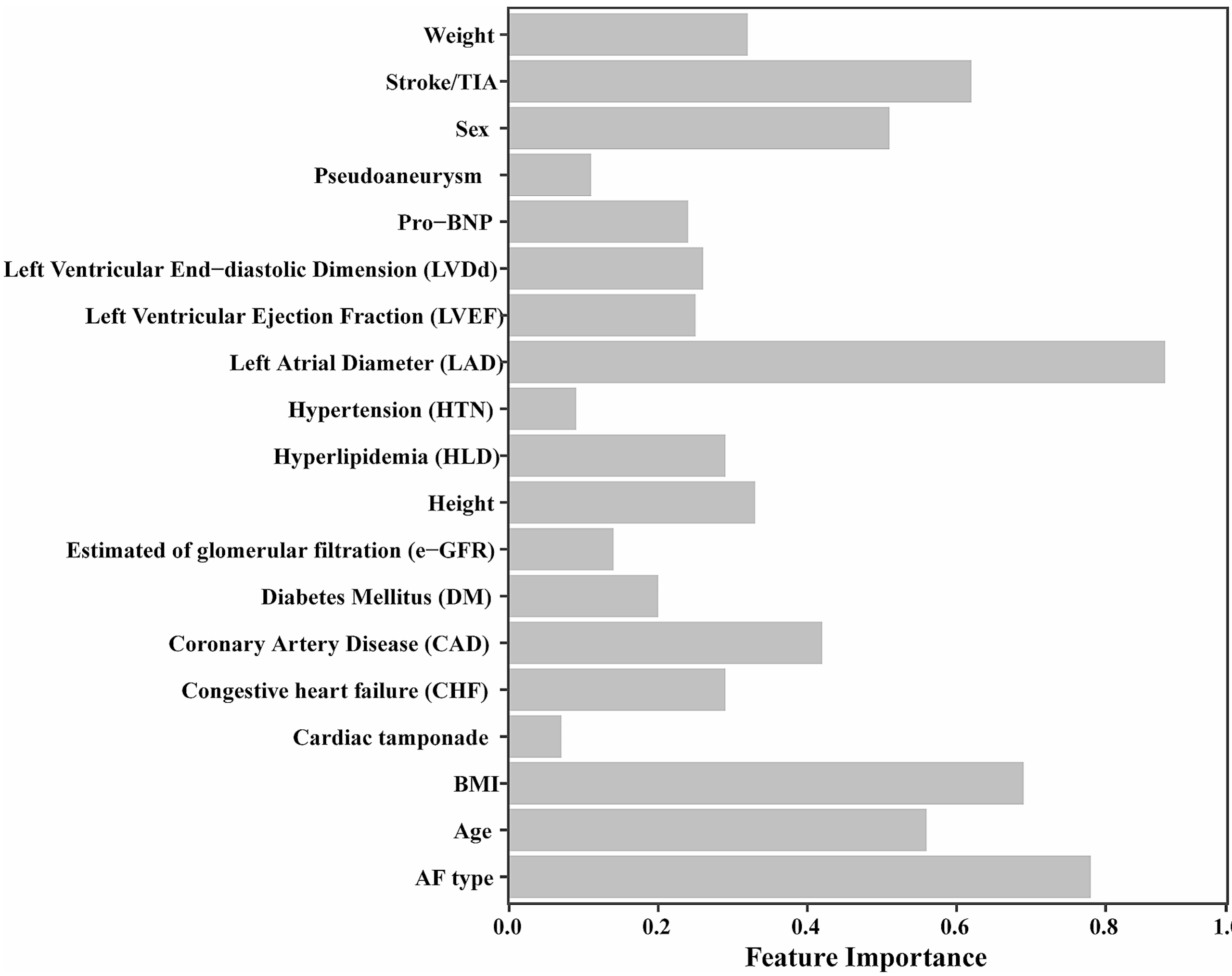

Fig. 3 Feature importance of the clinical variables before the ablation procedure

analysis and text generation. Moreover, DL techniques have demonstrated remarkable success in the medical field. The inherent capability of DL to autonomously extract pertinent features and comprehend intricate patterns from ECG or Holter data has significantly improved the accuracy and efficiency of arrhythmias diagnosis and risk stratification [24–26]. A transformer-based automatic system, combining denoising and segmentation modules, was utilized to identify ST-segment and J point abnormalities in patients using long-term Holter ECG data [24]. Additionally, a novel CNN-based input structure was developed to enhance the feature extraction capability from dynamic ECG signals, enabling the detection of premature ventricular contractions (PVC) and supraventricular premature beats (SPB) [26].

Clinical factors and imaging-based features have been reported to be associated with AF recurrence after ablation [27–29]. Many risk factors have been identified to predict AF recurrence after ablation, including left atrial enlargement, female, age and AF type [28, 30]. The CAAP-AF score predicted long-term freedom from AF following the final ablation and provided a practical expectation for individual patient outcomes following AF ablation [27]. Previous studies have also assessed the utility of ML in predicting AF recurrence using ECG or EGM, imaging data and clinical data. Shade et al [15] demonstrated that ML and personalized computational modeling could be combined to predict AF recurrence in 32 patients with an AUROC of 0.82. In their study, random forests and a Quadratic Discriminant Analysis (QDA) classifier were trained using features derived from magnetic resonance imaging (MRI) images and simulations. Roney et al [31] constructed 100 patient-specific models to predict long-term response to AF ablation. They found that models based on clinical history, imaging and simulation stress tests outperformed those trained on clinical history and imaging or clinical history alone. In recent research, the CNN model (a fusion of EGM, ECG signals, and clinical features) showed excellent predictive performance with an AUROC, sensitivity, specificity, and

Subgroup	Sensitivity	Specificity	Hazard ratio(95% CI)
Over all	81.1	81.7	10.7 (5.4-21.3)
Age, y			
<65	75	83.9	7.1 (3.2-15.9)
≥65	90.5	84.5	24.3 (5.6-104.8)
Sex			
Male	80.6	78.4	22.2 (5.3-92.6)
Female	82.4	88.4	15.2 (4.3-52.9)
AF type			
PAF	79.4	83.9	11.4 (5.0-26.3)
Non-PAF	84.2	76.3	8.9 (2.6-30.7)
LAD, mm			
<40	87.5	85.9	20.2 (6-68)
≥40	75.9	76.7	6.5 (2.7-15.2)
CHA_2DS_2 - VASc Score			
<3	77.8	81.1	8.6 (3.9-18.9)
≥3	88.2	82.9	18.8 (4.3-82.6)

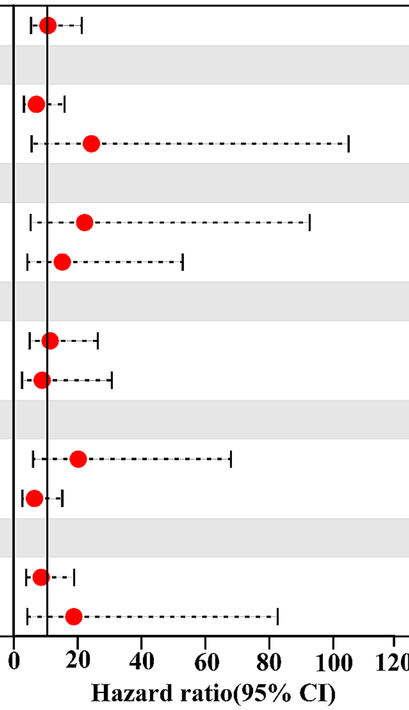

Fig. 4 The performance of the fusion model in subpopulations. CI: confidence interval

accuracy of 0.859, 87.0%, 86.7%, and 86.6%, respectively. However, 28% of patients in this study had prior AF ablation, which may have affected characteristics of the EGM or ECG signals.

In our study, we utilized the Transformer algorithm to predict AF recurrence by extracting temporal information from ECG images. Image ECGs offer a more intuitive and accessible format for clinicians, potentially enhancing diagnostic accuracy and efficiency. Furthermore, image ECGs are more compatible with ML algorithms, facilitating automated analysis and advancing cardiac health monitoring [32]. Adding clinical features further improved the model's performance. Compared to CNNs, which only process fixed-length time-series vectors and lose some feature information, the Transformer network focuses more on the temporal continuity of the data and captures the data's hidden deep features well [33, 34]. The structural changes that occur with AF recurrence, such as myocyte hypertrophy and fibrosis, which may cause subtle ECG changes. Thus, DL could afford the ability to consider complex datasets of all contained data and detect subtle ECG changes associated with structural changes in AF.

Recently, our STABLE-SR-III trial demonstrated that CPVI combined with LVA modification increased the success rate compared to CPVI alone in older patients with PAF [35]. Furthermore, our recent study found that older women exhibited more advanced atrial substrate and might benefit more from additional LVA modification than men [7]. In this study, female tended to have a significantly higher incidence of AF recurrence and proved to be an independent predictor of AF recurrence. Subgroup analyses indicated that the fusion model had better predictive performance in older patients compared to younger patients. This difference could be attributed to the higher prevalence of comorbidities like stroke/TIA, HF, and CAD in the older patients. Additionally, the Transformer exhibited satisfactory performance in both PAF and non-PAF patients, with relatively better results in non-PAF patients, indicating that our approach is applicable to subgroups with diverse AF types. Furthermore, the model performance did not differ significantly among patients with prior-ablation ECG and those with post-ablation ECG.

Limitations

Firstly, this is a single center and retrospective study. Secondly, due to the unavailability of a comparable dataset, external validation could not be conducted, and we have relied on internal validation to assess the performance of our data. Thirdly, surface ECG in SR was not obtained before ablation in all patients, and the ablation procedure might have impacted the ECG characteristics for non-PAF patients. However, we found that the model performance did not differ significantly among patients with prior-ablation ECG and those with post-ablation ECG. Finally, DL models are still perceived as black boxes and improving interpretability is an ongoing research.

Conclusions

Deep learning based on Transformer algorithm demonstrated outstanding performance in predicting AF recurrence. Integrating ECG and clinical features enhanced the models' performance compared to using clinical data or ECG alone.

Abbreviations

AF	Atrial Fibrillation
AUROC	Areas Under the Receiver Operating Characteristic Curve
BMI	Body Mass Index
CAD	Coronary Artery Disease
CNN	Convolutional Neural Network
CPVI	Circumferential Pulmonary Vein Isolation
DL	Deep Learning
ECG	Electrocardiogram
LAD	Left Atrial Anteroposterior Diameter
LVA	Low Voltage Area
LVDd	Left Ventricular End Diastolic Diameter
LVEF	Left Ventricular Ejection Fraction
ML	Machine Learning
MLP	Multilayer Perceptron
NLP	Nature Language Processing
SR	Sinus Rhythm
TIA	Transient Ischemic Attack

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-024-02616-x.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

QY and GHC contributed to the conceptualization and design of the study. QY, WSX, YS, PXF and XYDQ collected the data. QY, GHC, CRJ, YJ and LJH conducted the analysis. QY and GHC led the writing of the original draft. LMF, LZJ, CHW and CML edited the manuscript, discussed results, and provided feedback regarding the manuscript. All authors had full access to the data and approved the manuscript for publication.

Funding

This work was supported by the National Natural Science Foundation of China (grant number 82170322).

Data availability

Any additional data are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of The First Affiliated Hospital of Nanjing Medical University. All patients provided written informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 23 January 2024 / Accepted: 22 July 2024
Published online: 08 August 2024

References

1. Writing Committee M, Joglar JA, Chung MK, Armbruster AL, Benjamin EJ, Chyou JY, et al. 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: a report of the American College of Cardiology/American Heart Association Joint Committee on clinical practice guidelines. J Am Coll Cardiol. 2024;83(1):109–279.
2. Bizhanov KA, CKB capital A AK, Baimbetov AB, Sarsenbayeva, Lyan E. Atrial fibrillation: epidemiology, pathophysiology, and clinical complications (literature review). J Cardiovasc Electrophysiol. 2023;34(1):153–65.
3. Joglar JA, Chung MK, Armbruster AL, Benjamin EJ, Chyou JY, Cronin EM et al. 2023 ACC/AHA/ACCP/HRS guideline for the diagnosis and management of atrial fibrillation: a report of the American College of Cardiology/American Heart Association Joint Committee on clinical practice guidelines. Circulation. 2024;149(1):e1–e156.
4. Calkins H, Hindricks G, Cappato R, Kim YH, Saad EB, Aguinaga L, et al. 2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation. Heart Rhythm. 2017;14(10):e275–444.
5. Reddy VY, Dukkipati SR, Neuzil P, Natale A, Albenque JP, Kautzner J, et al. Randomized, controlled trial of the safety and effectiveness of a contact force-sensing irrigated catheter for ablation of paroxysmal atrial fibrillation: results of the TactiCath contact force ablation catheter study for atrial fibrillation (TOCCASTAR) study. Circulation. 2015;132(10):907–15.
6. Kirchhof P, Calkins H. Catheter ablation in patients with persistent atrial fibrillation. Eur Heart J. 2017;38(1):20–6.
7. Wu N, Liu W, Yang G, Chen H, Ju W, Li M, et al. Sex difference in atrial scar prevalence. What can we learn from the STABLE-SR-III trial? Heart Rhythm. 2024.
8. Kosich F, Schumacher K, Potpara T, Lip GY, Hindricks G, Kornej J. Clinical scores used for the prediction of negative events in patients undergoing catheter ablation for atrial fibrillation. Clin Cardiol. 2019;42(2):320–9.
9. Jacobs V, May HT, Bair TL, Crandall BG, Cutler M, Day JD, et al. The impact of risk score (CHADS2 versus CHA2DS2-VASc) on long-term outcomes after atrial fibrillation ablation. Heart Rhythm. 2015;12(4):681–6.
10. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of go with deep neural networks and tree search. Nature. 2016;529(7587):484–9.
11. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30.
12. Campello VM, Gkontra P, Izquierdo C, Martín-Isla C, Sojoudi A, Full PM, et al. Multi-centre, multi-vendor and multi-disease cardiac segmentation: the M&Ms challenge. IEEE Trans Med Imaging. 2021;40:3543–54.
13. Li J, Chen J, Tang Y, Wang C, Landman BA, Zhou SK. Transforming medical imaging with transformers? A comparative review of key properties, current progresses, and future perspectives. Med Image Anal. 2023;85:102762.
14. Bhalodia R, Goparaju A, Sodergren T, Morris AK, Kholmovski E, Marrouche NF et al. Deep learning for end-to-end atrial fibrillation recurrence estimation. 2018 Computing in Cardiology Conference (CinC). 2018;45:1–4.
15. Shade JK, Ali RL, Basile D, Popescu D, Akhtar T, Marine JE, et al. Preprocedure application of machine learning and mechanistic simulations predicts likelihood of paroxysmal atrial fibrillation recurrence following pulmonary vein isolation. Circ Arrhythm Electrophysiol. 2020;13(7):e008213.
16. Firouznia M, Feeny AK, LaBarbera MA, McHale M, Cantlay C, Kalfas N, et al. Machine learning-derived fractal features of shape and texture of the left atrium and pulmonary veins from cardiac computed tomography scans are associated with risk of recurrence of atrial fibrillation postablation. Circ Arrhythm Electrophysiol. 2021;14(3):e009265.
17. Tang S, Razeghi O, Kapoor R, Alhusseini MI, Fazal M, Rogers AJ, et al. Machine learning-enabled multimodal fusion of intra-atrial and body surface signals in prediction of atrial fibrillation ablation outcomes. Circ Arrhythm Electrophysiol. 2022;15(8):e010850.
18. Yang B, Jiang C, Lin Y, Yang G, Chu H, Cai H et al. STABLE-SR (electrophysiological substrate ablation in the Left Atrium during Sinus Rhythm) for the treatment of nonparoxysmal atrial fibrillation: a prospective, multicenter randomized clinical trial. Circ Arrhythm Electrophysiol. 2017;10(11).
19. Yang G, Zheng L, Jiang C, Fan J, Liu X, Zhan X et al. Circumferential pulmonary vein isolation plus low-voltage area modification in persistent atrial fibrillation: the STABLE-SR-II trial. JACC Clin Electrophysiol. 2022;8(7):882–91.
20. Yang G, Yang B, Wei Y, Zhang F, Ju W, Chen H, et al. Catheter ablation of nonparoxysmal atrial fibrillation using electrophysiologically guided substrate modification during sinus rhythm after pulmonary vein isolation. Circ Arrhythm Electrophysiol. 2016;9(2):e003382.

21. Kornej J, Hindricks G, Arya A, Sommer P, Husser D, Bollmann A. The APPLE score - a novel score for the prediction of rhythm outcomes after repeat catheter ablation of atrial fibrillation. PLoS ONE. 2017;12(1):e0169933.
22. Kosiuk J, Dinov B, Kornej J, Acou WJ, Schonbauer R, Fiedler L, et al. Prospective, multicenter validation of a clinical risk score for left atrial arrhythmogenic substrate based on voltage analysis: DR-FLASH score. Heart Rhythm. 2015;12(11):2207–12.
23. Chen D, Liu S, Kingsbury P, Sohn S, Storlie CB, Habermann EB, et al. Deep learning and alternative learning strategies for retrospective real-world clinical data. NPJ Digit Med. 2019;2:43.
24. Wu S, Cao Q, Chen Q, Jin Q, Liu Z, Zhuang L, et al. Using multi-task learning-based framework to detect ST-segment and J-point deviation from holter. Front Physiol. 2022;13:912739.
25. Nadarajah R, Wu J, Frangi AF, Hogg D, Cowan C, Gale C. Predicting patient-level new-onset atrial fibrillation from population-based nationwide electronic health records: protocol of FIND-AF for developing a precision medicine prediction model using artificial intelligence. BMJ Open. 2021;11(11):e052887.
26. Meng L, Tan W, Ma J, Wang R, Yin X, Zhang Y. Enhancing dynamic ECG heartbeat classification with lightweight transformer model. Artif Intell Med. 2022;124:102236.
27. Winkle RA, Jarman JW, Mead RH, Engel G, Kong MH, Fleming W, et al. Predicting atrial fibrillation ablation outcome: the CAAP-AF score. Heart Rhythm. 2016;13(11):2119–25.
28. den Uijl DW, Bax JJ. Left atrial size as a predictor of successful radiofrequency catheter ablation for atrial fibrillation. Europace. 2009;11(10):1255–6.
29. Gal P, Marrouche NF. Magnetic resonance imaging of atrial fibrosis: redefining atrial fibrillation to a syndrome. Eur Heart J. 2017;38(1):14–9.
30. Abecasis J, Dourado R, Ferreira A, Saraiva C, Cavaco D, Santos KR, et al. Left atrial volume calculated by multi-detector computed tomography may predict successful pulmonary vein isolation in catheter ablation of atrial fibrillation. Europace. 2009;11(10):1289–94.
31. Roney CH, Sim I, Yu J, Beach M, Mehta A, Solis-Lemus JA, et al. Predicting atrial fibrillation recurrence by combining population data and virtual cohorts of patient-specific left atrial models. Circ Arrhythm Electrophysiol. 2022;15(2):e010253.
32. Ao R, He G. Image based deep learning in 12-lead ECG diagnosis. Front Artif Intell. 2022;5:1087370.
33. Khan S, Naseer M, Hayat M, Zamir SW, Khan FS, Shah M. Transformers in vision: a survey. ACM computing surveys (CSUR). 2022;54(10s):1–41.
34. Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS et al. Transformers in medical imaging: a survey. arXiv preprint. 2022;arXiv:2201.09873.
35. 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(8):765–72.

Publisher's Note