KHRS 2023

A New Artificial Intelligence Algorithms for Predicting Atrial Fibrillation Using Serial 12-lead Electrocardiograms

[1]Ji-Hoon Choi, [2]Kyoung-Min Park

[1]Division of Cardiology, Department of Internal Medicine, Konkuk University Medical Center
Konkuk University School of Medicine, Seoul, Republic of Korea

[2]Division of Cardiology, Department of Medicine, Samsung Medical Center
Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

Korean Heart Rhythm Society

COI Disclosure

Name of First Author: Ji-Hoon Choi, Sung-Hee Song

Corresponding author: Kyoung-Min Park

The authors have no financial conflicts of interest
to disclose concerning the presentation

This work was supported by a grant from the Korea Medical Device Development Fund funded by the Korean government

Introduction

- Most cases of strokes occurs without forewarning since early-stage atrial fibrillation (AF) is rarely detected.

- Artificial intelligence (AI) algorithms using 12-lead electrocardiogram (ECG) are being developed to detect AF in advance of stroke.

- A common feature in previously published studies is prediction of AF using deep-learning convolutional neural networks (CNNs) based on a single normal sinus rhythm (NSR) ECG.

- The agnostic approach of CNNs can provide high-quality results, but humans cannot understand how the AI mad the decision.

Introduction

- In contrast, explainable machine learning (ML) algorithms are expected to provided reliable, interpretable information in clinical medicine where high-stakes decision-making is required.

- Left atrial (LA) remodeling is the pathophysiologic basis of AF initiation and progression.

- Electrophysiological and structural remodeling of the LA promotes the development of AF.

Objective

- Unlike previous studies that derived deep-learning algorithms from a single NSR ECG,
- We hypothesized that analysis of an individual's serial NSR ECGs (paired ECGs performed within a certain period) could predict new-onset AF more accurately than analysis of a single NSR ECG by detecting the subtle cardiac remodeling that occurs immediately before AF occurrence.
- In addition, ECG features used for training ML algorithms were investigated to advance explainable AI.

Method – ECG collection

- All standard 12-lead ECGs obtained from patients at Samsung Medical Center who were older than 18 years between January 2010 and December 2021 were identified.

- All ECGs were conducted using a Philips ECG instruments at a 500 Hz sampling rate with 5 μV resolution, and raw data were stored in XML format.

- ECG data were divided into training, internal validation, and test tests at an 8:1:1 ratio without overlap.

- For external validation of the developed AI model, ECGs conducted using a Philips ECG machine at three other tertiary hospital were used.

Method – Study population

- Medical record and diagnostic codes were reviewed for all cases.
- Definite AF were designated only for those patients with a documented AF ECG (12-lead ECG or Holter monitoring) and AF diagnosis confirmed by medical records or diagnostic codes.
- All patients included in this study were classified into an AF group or NSR group according to the criteria for definite AF.
- Index AF ECG was defined as the first documented instance of AF according to any ECG modality.
- All NSR ECGs within 2 years of the index AF ECG were analyzed.

Method – Study population

- **Exclusion criteria**

✓ Patients diagnosed with AF in their medical records or a diagnostic code of AF before the index AF ECG

✓ Patients without an NSR ECG prior to the index AF ECG

✓ Patients with only one ECG

✓ Patients with a medical record or diagnostic code for AF but no AF ECG

✓ Insufficient medical records to evaluate a patient's medical status

✓ Missing or unsuitable digital ECG data

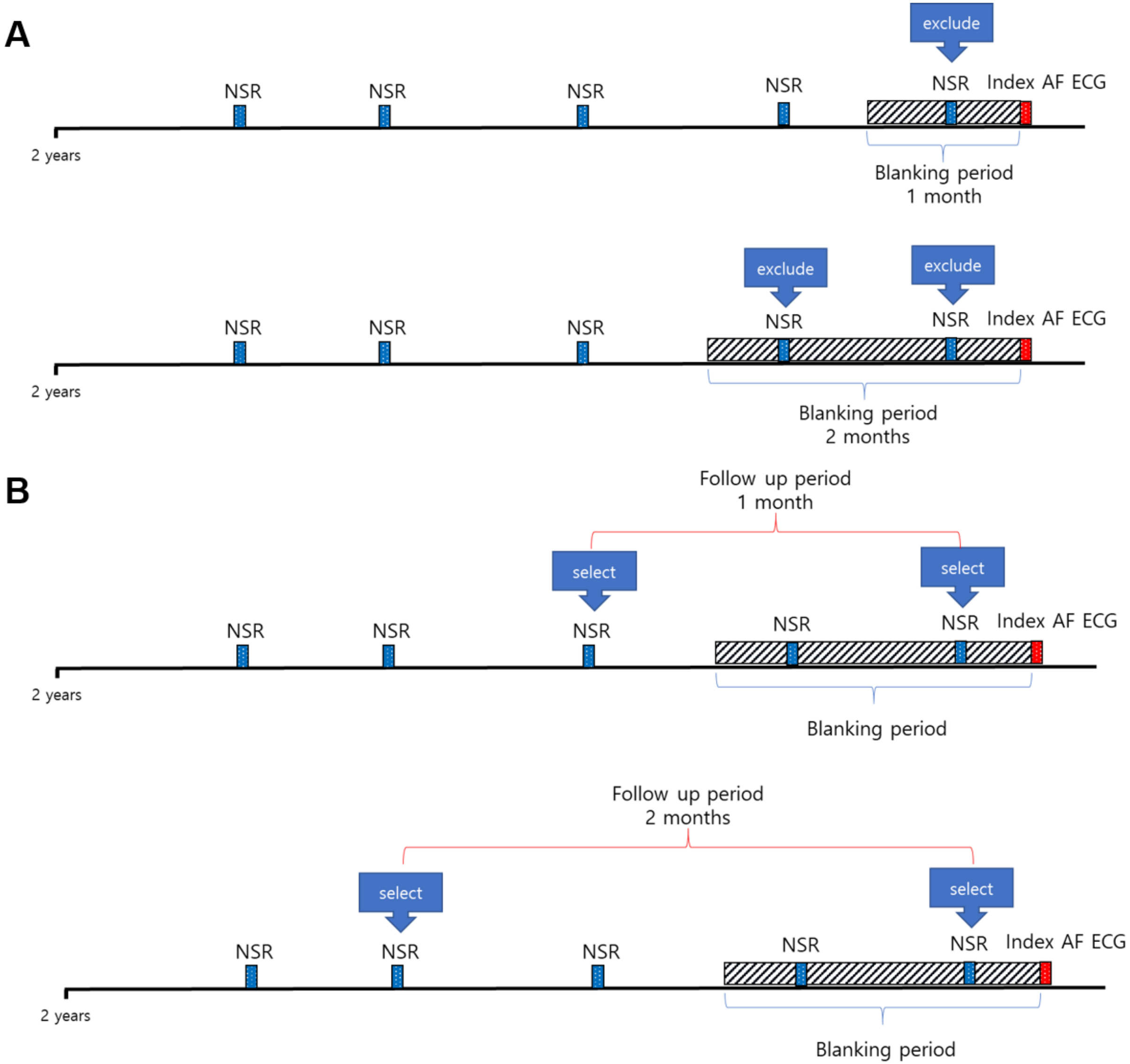
Method – Machine learning model development

- All ECG features (components of P-QRS-T waves including peaks, intervals, sed segments) were extracted from all 12-leads.

- For both single and serial ECG ML models, a light gradient boosting machine (LGBM) algorithm, which is a machine learning algorithm based on a gradient boosting decision tree (GBDT), was used.

- To select the best hyperparameters for the LGBM model, a Bayesian optimization method was used.

Method – Two machine learning models

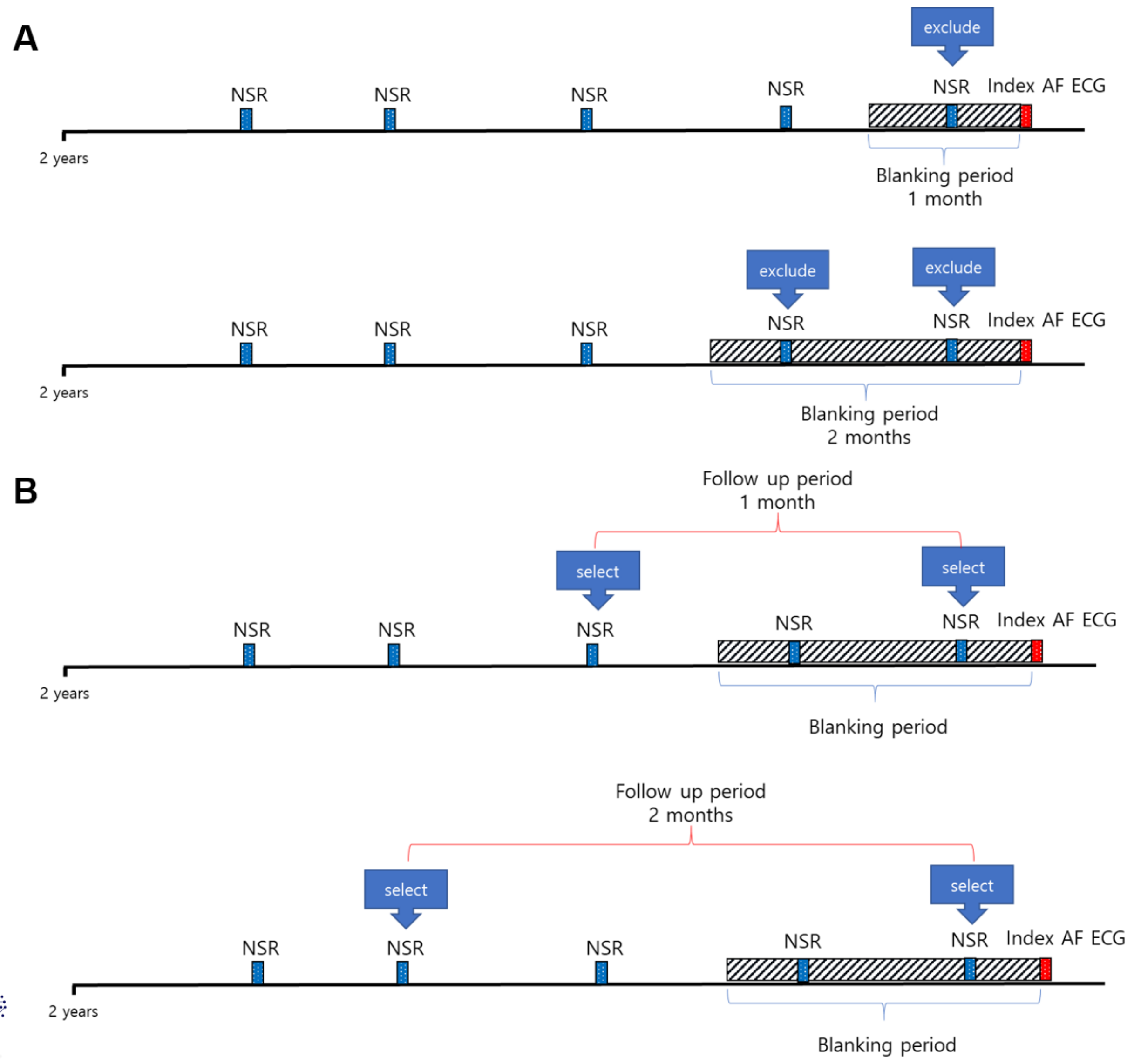


- Analysis was conducted using two ML models, single and serial ECG ML model, both of were designed to predict new-onset AF.

- One NSR ECG prior to the index AF ECG was used for the single ECG model (Figure A), while two NSR ECGs prior to the index AF ECG were used for the serial ECG model (Figure B).

- We defined the blanking period as the minimum period during which atrial remodeling occurs within which the ML algorithm was able to detect subtle changes in the atrium during NSR.

Method – Two machine learning models



- To determine the optimal blanking period, NSR ECGs over defined periods (1 month, 2 months, 3 months, etc.) were excluded from the index AF ECG using the single ECG ML model (Figure A).
- In the serial ECG ML model, we selected two ECGs, one ECG from the blanking period defined by the single ECG ML model, and the other prior to the blanking period (Figure B).
- To determine the optimal ECG follow-up duration for predicting AF, we compared NSR ECGs according to follow-up period.

Method – Outcomes

- The primary outcome was prediction of new-onset AF using standard 12-lead NSR ECGs using two ML models (single ECG model vs. serial ECG model) and comparison of two ML models.

- ML model performance was assessed using the area under the receiver operating characteristic curve (AUROC), precision-recall curve, sensitivity, specificity, accuracy, and F1 score.

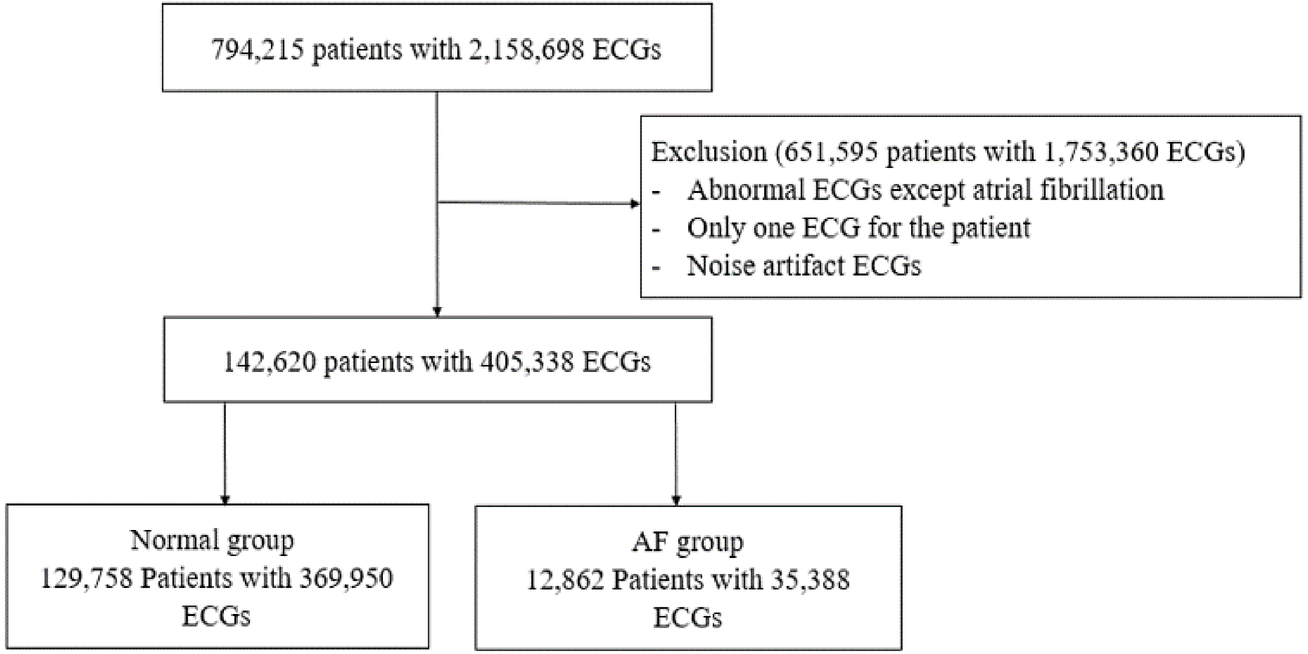
- We also determined the optimal blanking period and follow-up period for ECGs for each model.

- Identification of ECG features that were significant predictors of AF was a secondary outcome.

Method – Statistical analysis

- The ROC curves of the two models were compared using DeLong's test.
- Analysis of variance (ANOVA) and F-tests were performed to identify ECG parameters that were significant predictors of AF.
- Statistical significance was defined as a 2-tailed p value less than 0.05.
- All statistical analyses were conducted in R statistical software (version 4.2.1) and Python (version 3.8)

Results – Study population

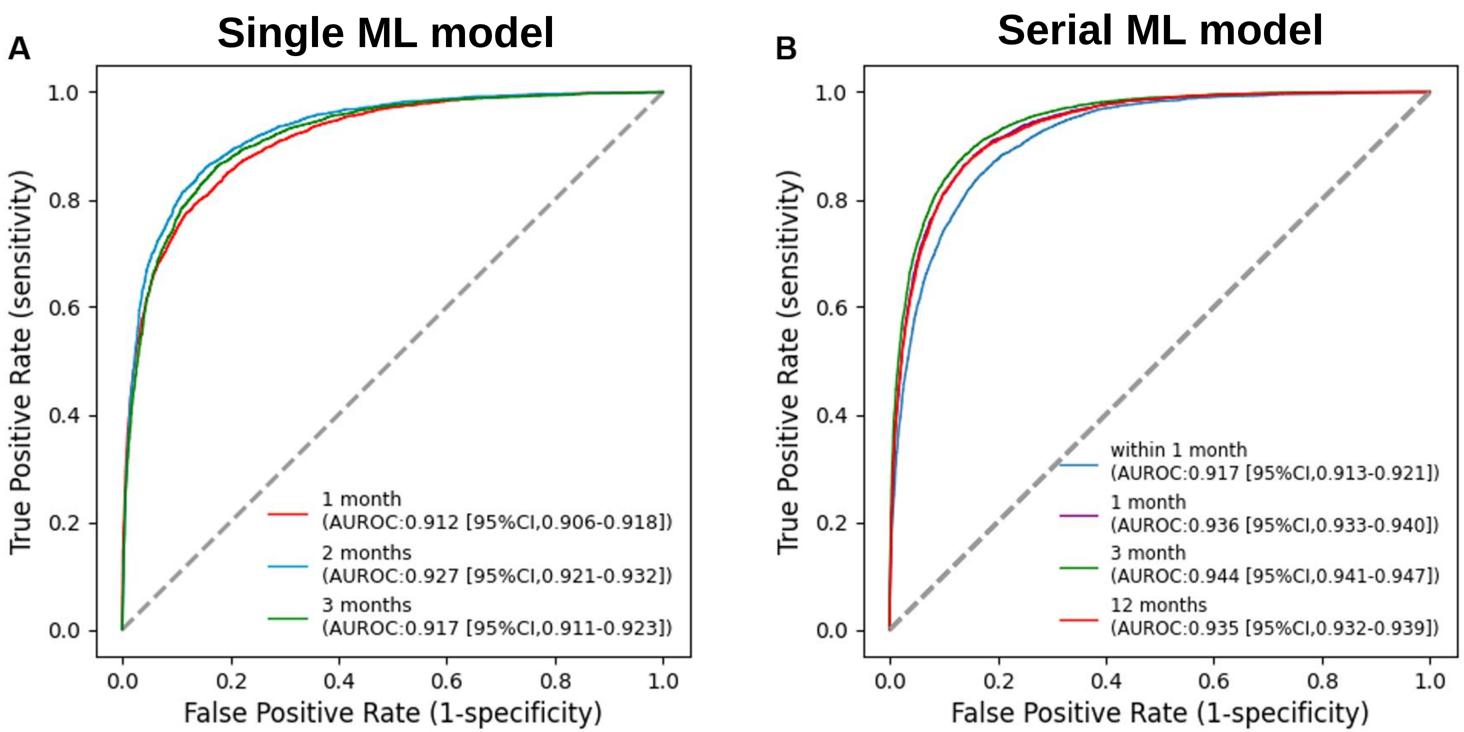


- A total of 2,158,698 ECGs was identified from 794,215 adult patients, and 1,753,360 ECGs from 651,595 patients were excluded based on the study criteria.

- We trained the single and serial models on 405,338 ECGs from 142,620 patients.

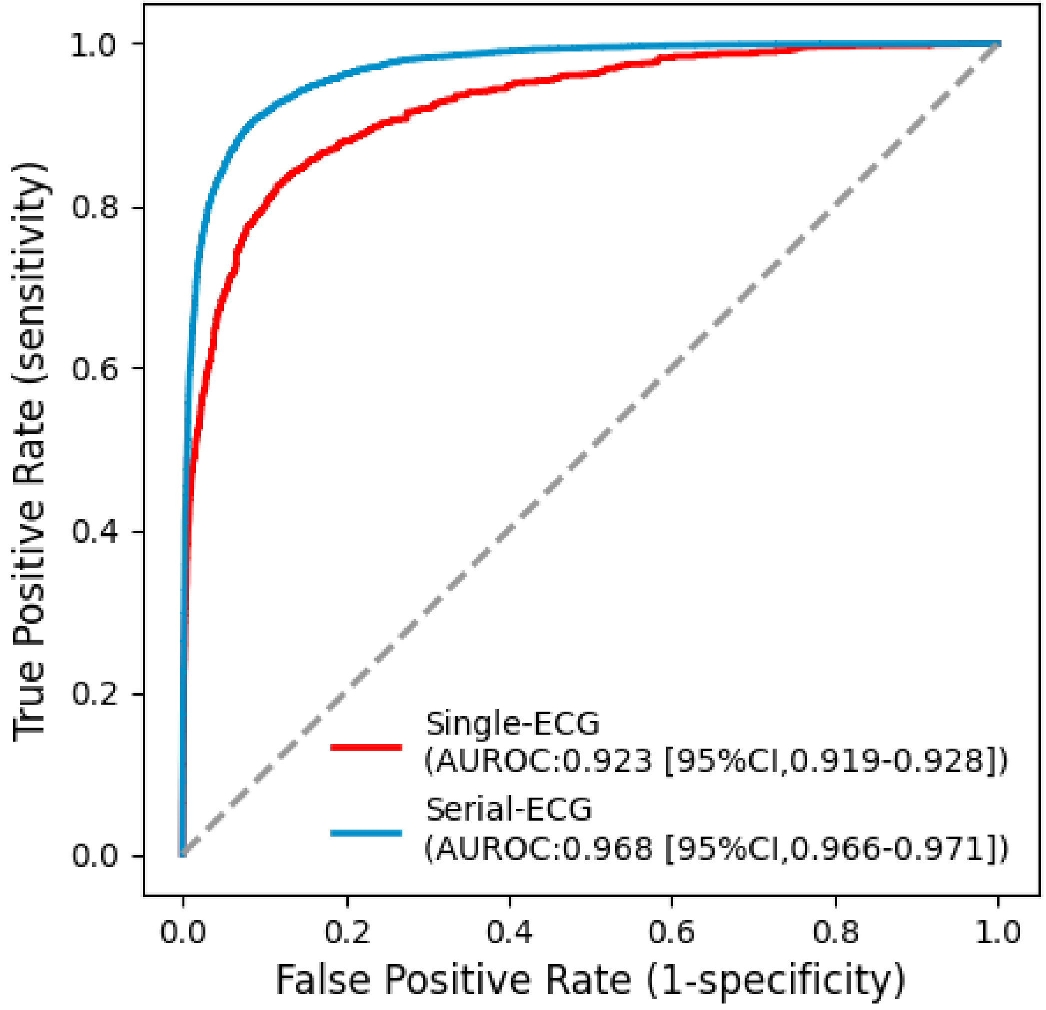
- Mean age was 66.5 ± 12.4 years, and 63,777 (44.7%) patients were male.

Results – Two ML models for predicting AF



	Blanking period	Sensitivity	Specificity	PPV	NPV	F1 score
Single ML model	1 month	0.655 (0.646 - 0.663)	0.9437 (0.939 - 0.947)	0.774 (0.766 - 0.781)	0.902 (0.897 - 0.907)	0.709 (0.701 - 0.717)
	2 months	0.683 (0.675 - 0.692)	0.949 (0.945 - 0.953)	0.809 (0.802 - 0.816)	0.906 (0.900 - 0.911)	0.741 (0.733 - 0.749)
	3 months	0.646 (0.637 - 0.651)	0.946 (0.942 - 0.951)	0.788 (0.780 - 0.796)	0.897(0.891 - 0.902)	0.710 (0.701 - 0.718)
Serial ML model	**Follow-up period**	**Sensitivity**	**Specificity**	**PPV**	**NPV**	**F1 score**
	1 month	0.313 (0.306 - 0.320)	0.992 (0.991 - 0.993)	0.939 (0.935 - 0.942)	0.795 (0.789 - 0.801)	0.469 (0.462 - 0.477)
	2 months	0.642 (0.635 - 0.649)	0.960 (0.957 - 0.963)	0.858 (0.853 - 0.863)	0.878 (0.873 - 0.883)	0.734 (0.728 - 0.741)
	3 months	0.683 (0.677 - 0.690)	0.958 (0.955 - 0.961)	0.859 (0.854 - 0.864)	0.890 (0.886 - 0.895)	0.761 (0.755 - 0.768)
	6 months	0.625 (0.618 - 0.632)	0.965 (0.962 - 0.968)	0.873 (0.869 - 0.878)	0.874 (0.869 - 0.879)	0.729(0.722 - 0.735)
	12 months	0.295 (0.289 - 0.302)	0.993 (0.992 - 0.994)	0.944 (0.941 - 0.947)	0.791 (0.785 - 0.797)	0.450 (0.443 - 0.457)

Results – Comparison of two ML models



- The AUROC of the single ECG ML model for the prediction of new onset AF was 0.923 (95% CI, 0.919–0.927) compared to 0.952 (95% CI, 0.966–0.971) for the serial ECG ML model.

- The serial ML model showed better performance for predicting new onset AF than the single ML model (single vs. serial ML model: sensitivity 0.684 vs. 0.804; specificity 0.950 vs. 0.967; accuracy 0.887 vs. 0.923; F1 score 0.741 vs. 0.849).

Result – External validation

	AUC	F1 score	Sen (%)	Spe (%)	PPV (%)	NPV (%)	ACC (%)
Single ECG AI model	0.925	0.802	0.730	0.910	0.890	0.771	0.820
Serial ECG AI model	0.964	0.877	0.817	0.953	0.946	0.839	0.885

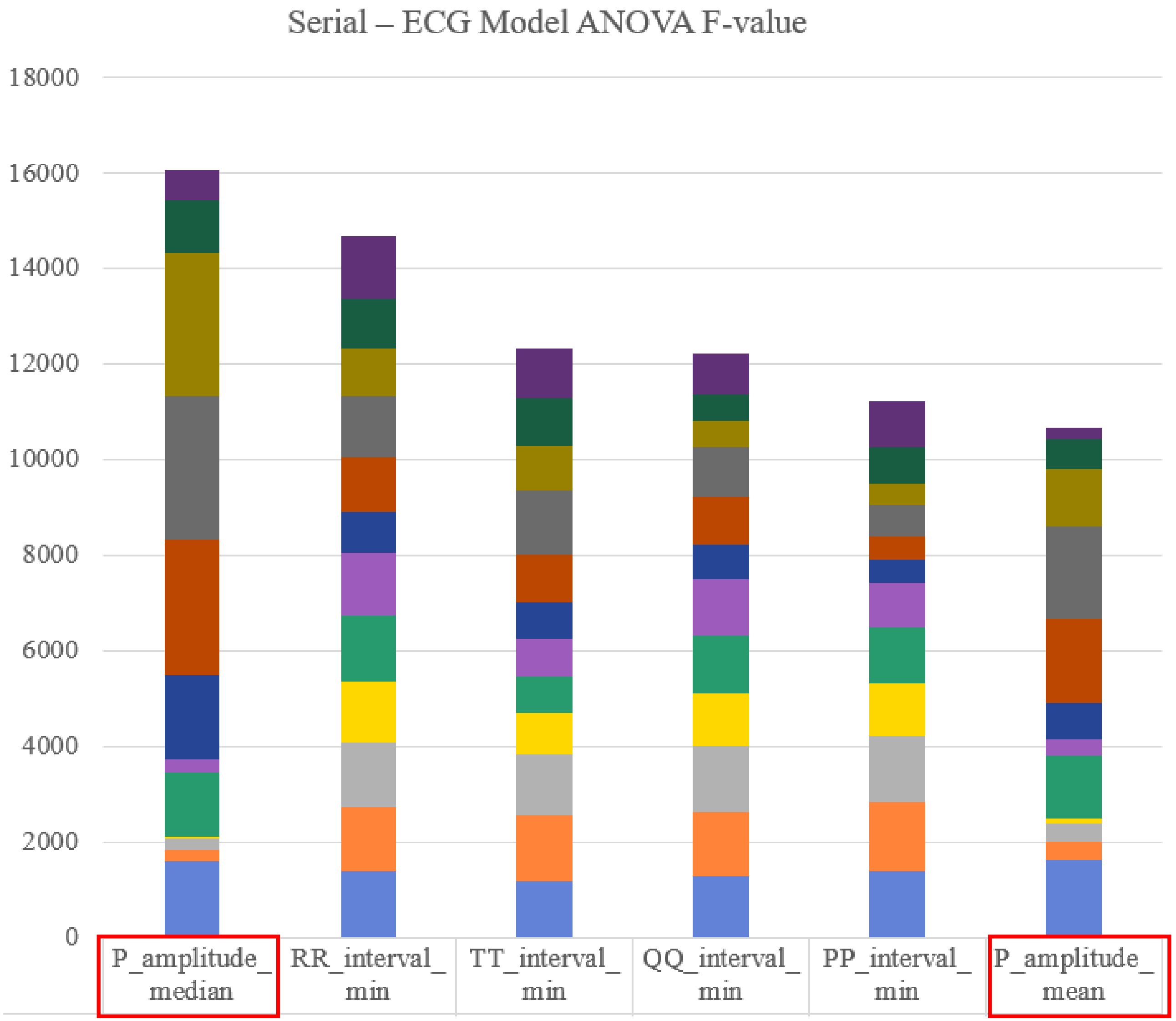
- The external validation results were similar to those of internal validation

Results – Comparison with state-of-the art DL model

Model	Sensitivity	Specificity	PPV	NPV	F1 score
DenseNet-121	0.728	0.934	0.775	0.917	0.751
Transformer	0.839	0.813	0.818	0.835	0.827
LGBM (Single)	0.683	0.949	0.809	0.906	0.741
LGBM (Serial)	0.804	0.967	0.903	0.930	0.849

- We also developed a state-of-the-art DL model and compared it to a ML model.
- The serial ECG model showed higher prediction accuracy than the deep learning model.

Results – Features of ECG parameters used in ML learning



- In this study, 1,554 features were selected and classified from our ML models.
- Top 6 features were sorted according to ANOVA F-value.
- In the serial ML model, the difference in P-wave voltage between two serial NSR ECGs was the most useful parameter for differentiating between AF and NSR groups.

Limitations

- The number of patients in the external validation group was small (n=600).
- Due to the nature of paroxysmal AF, undetected subclinical AF likely occurs prior to diagnosis using 12-lead ECG. Although only patients with at least two NSR ECGs were included, and extensive medical record review was performed to minimize this bias, the possibility of undetected AF in the NSR group cannot be excluded.
- The timing of AF diagnosis was unclear.
- NSR ECGs from the index AF ECG were not evenly distributed over time, which may have impacted our findings.
- Prospective studies are needed to address these issues.

Conclusions

- We demonstrated that an ML model based on serial ECGs can predict new-onset AF more accurately than an ML model based on a single ECG.
- The optimal follow-up interval between ECGs should be at least 3 months to reflect atrial remodeling.
- Deep learning models outperformed our single ECG ML model, but our proposed serial ECG ML model outperformed all existing AI models.
- A difference in P-wave voltage between serial ECGs was the most potent predictive AF prediction parameter.
- The findings of this study can help advance explainable AI by providing a pathophysiological rationale for ECG change used to predict new-onset AF.
- Further clinical investigation is necessary to confirm the performance of this serial ECG model.

Thank you for your attention