Enhancing Stroke Prediction in Atrial Fibrillation: Integrating Polygenic Risk Scores with Artificial Intelligence-Guided Clinical Models

Tae-Hoon Kim

The Graduate School
Yonsei University
Department of Medicine

Enhancing Stroke Prediction in Atrial Fibrillation: Integrating Polygenic Risk Scores with Artificial Intelligence-Guided Clinical Models

A Dissertation Submitted
to the Department of Medicine
and the Graduate School of Yonsei University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Medical Science

Tae-Hoon Kim

December 2024

This certifies that the Dissertation of Tae-Hoon Kim is approved

Thesis Supervisor [Hui-Nam Pak]

Thesis Committee Member [Young Dae Kim]

Thesis Committee Member [Sang-Hak Lee]

Thesis Committee Member [Eue-Keun Choi]

Thesis Committee Member [Hyo Sup Shim]

The Graduate School
Yonsei University
December 2024

ACKNOWLEDGEMENTS

I have received tremendous support throughout my journey from medical school to my work at Severance Hospital. I would like to express my deepest gratitude to my advisor, Professor Hui-Nam Pak, as well as the professors in the arrhythmia department. I am also sincerely thankful to the professors who kindly served as members of my review committee. Lastly, I want to convey my heartfelt thanks to my wife and son, who stood by me and gave me strength during my illness.

October 2024

Tae-Hoon Kim

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Enhancing Stroke Prediction in Atrial Fibrillation: Integrating Polygenic Risk Scores with Artificial Intelligence-Guided Clinical Models

Background: Atrial fibrillation (AF) significantly increases the risk of ischemic stroke, and current risk stratification models like CHA_2DS_2-VASc have limitations. This study aimed to develop a predictive model combining polygenic risk scores (PRS) with Artificial Intelligence (AI)-guided analysis of age and sex to improve stroke risk prediction in AF patients.

Methods: We included 3,190 AF patients who undergoing AF catheter ablation (AFCA) from the Yonsei AF Ablation Cohort. Patients were categorized into two groups: 320 with a history of early ischemic stroke before AFCA, and 2,870 stroke-free controls. We developed a polygenic risk score (PRS) based on genome-wide association studies (GWAS) and combined it with AI-guided analysis of age and sex derived from ECG data. The predictive performance of the integrated model for the prediction of ischemic stroke history before AFCA was compared to the CHA_2DS_2-VASc score (excluding the stroke/TIA component) using logistic regression analysis and area under the receiver operating characteristic curve (AUC) metrics.

Results: The integrated model, combining multi-ethnic AF PRS with AI-guided age and sex predictions, significantly improved stroke risk stratification compared to the CHA_2DS_2-VASc score alone. The inclusion of PRS increased the AUC from 0.615 to 0.621, while the addition of AI-guided analysis further increased the AUC to 0.632. The integrated multi-ethnic AF PRS and AI-guided CHA_2DS_2-VASc risk model showed a significantly improved statistical fit ($\chi 2$ $P<0.001$) and modestly improved discrimination.

Conclusions: Integrating PRS with AI-guided analysis of age and sex improves risk stratification in AF patients, enabling more precise identification of individuals at higher risk for ischemic stroke. This approach enhances the early detection model, allowing for more proactive preventive measures.

Key words : Atrial fibrillation, Stroke, Polygenic Risk Score, Artificial Intelligence

I. INTRODUCTION

Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia with a nearly five-fold increased risk of ischemic stroke. While advances in diagnostic techniques have led to earlier detection of AF and the adoption of advanced rhythm control strategies including catheter ablation have improved patient outcomes, stroke remains a persistent complication. Moreover, many patients only begin rhythm control after the occurrence of a stroke, highlighting the need for more effective early risk stratification to prevent such events.

Traditional clinical risk stratification models, such as the CHA_2DS_2-VASc score, have been instrumental in guiding stroke prevention strategies in AF patients. However, these models have limitations in their ability to accurately predict stroke risk, particularly in patients with early-onset and late-diagnosed AF. As a result, there is a growing need for more precise predictive tools that can identify high-risk patients before a stroke occurs.

Recent advances in genomics and artificial intelligence (AI) can offer an opportunity to enhance stroke risk prediction beyond the capabilities of previous conventional models. Genome-wide association studies (GWAS) have identified numerous genetic variants associated with AF and stroke risk, leading to the development of polygenic risk scores (PRS) that aggregate the effects of these variants. PRS holds potential as a powerful tool for refining stroke risk stratification, especially when integrated with traditional clinical variables.

Concurrently, the application of AI in medicine has shown remarkable potential in analyzing complex datasets, uncovering patterns, and improving predictive models. By incorporating AI-guided models that consider age, sex, and other demographic factors, alongside genetic information, it is possible to generate a more personalized and accurate risk assessment tool. Such a tool could surpass the predictive performance of existing clinical scoring systems, thereby improving patient outcomes by enabling more targeted and timely interventions. This study aims to develop and validate a novel predictive model that integrates polygenic risk scores with AI-guided analysis of age and sex to predict early onset ischemic stroke in patients with AF who undergoing catheter ablation. We hypothesize that this model will outperform traditional clinical scoring systems, offering a more precise stratification of stroke risk, which is crucial for optimizing preventive strategies in this high-risk population.

II. METHODS

2.1. Study population

The study protocol adhered to the principles of the Declaration of Helsinki and was approved by the Institutional Review Board of the Yonsei University Health System. All the patients provided written informed consent for inclusion in the Yonsei AF Ablation Cohort Database (ClinicalTrials.gov Identifier: NCT02138695). Among the 7,058 patients who underwent AF catheter ablation (AFCA) in the Yonsei AF Ablation Cohort from 2009 to 2024, 3,190 patients with AI ECG data and available PRS data were enrolled in the study. The case group consisted of 320 patients with ischemic stroke history before AFCA, and stroke-free 2,870 controls who undergoing AFCA. Study protocol is presented in Figure 1. The diagnosis of ischemic stroke was defined as a clinical diagnosis confirmed by a neurologist at either our institution or at a referring hospital. An ischemic stroke was considered confirmed if the neurologist documented the diagnosis based on standard clinical assessments, neuroimaging (such as brain computed tomography [CT] or magnetic resonance imaging [MRI]), and other relevant diagnostic criteria. Only cases explicitly labeled as "ischemic stroke" in the patient's medical record by the attending neurologist were included in this study. Cases of transient ischemic attack (TIA) were specifically excluded from the analysis.

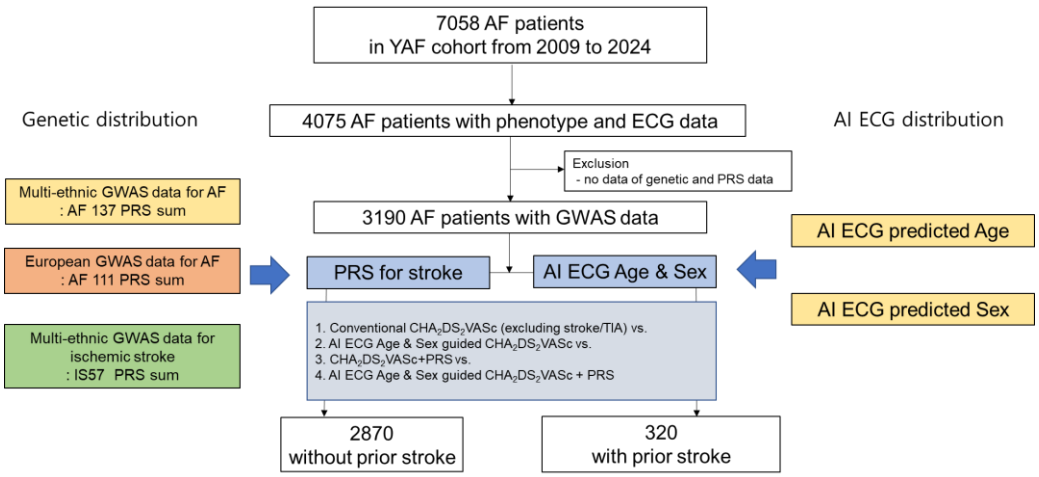

Figure 1. Study Flow Chart

2.2. Genome-wide association study

Genomic deoxyribonucleic acid (DNA) was obtained from blood samples using the QuckGene DNA Whole Blood Kit S with a QuickGene mini 80 (KURABO, Osaka, Japan). DNA genotyping data were obtained using the Axiom Precision Medicine Research Array (PMRA; Thermo Fisher Scientific, MA, USA). We searched a specific phenotype -associated single nucleotide polymorphism (SNP) in a gene, the most common AF-associated genome, using PMRA data. Using AF-associated SNP, we calculated polygenic risk score (PRS) and investigated the association between history of ischemic stroke before catheter ablation and PRS among the included patients.

2.3. Derivation of Polygenic risk score

We adopted a study design similar to previous published PRS studies, following the recommended methodological and reporting guidance. Briefly, the process involved four key steps; (1) curation of previously published genome-wide association study (GWAS) summary statistics, (2) accounting for linkage disequilibrium in GWAS summary statistics using the R package lassosum, (3) constructing PRS within our YAF cohort with eighty different PRS were constructed across the lassosum hyperparameters (λ and s), and (4) identifying the most accurate PRS in the YAF cohort. Detailed methods were presented in our previous studies.

2.4. Artificial Intelligence-guided Age and Sex prediction

We developed a ResNet-based model for ECG sex prediction. This framework has been demonstrated to successfully create models to predict age, sex, as well as mortality from a standard 12-lead ECG. The ResNet architecture includes residual blocks composed of 1D convolutional layers each followed by batch normalization, ReLU activation, and dropout layers. Detailed information on 12-lead ECG pre-processing and ResNet architecture is provided in the prior studies. We developed ECG sex prediction model using 500,000 ECGs from the development set of MIMIC-IV cohort. Following development, the model was validated on four independent multinational

datasets that were not used the development process. These validation sets included a total of 410,816 ECGs from the MIMIC-IV hold out set (n=13,628), CODE-15% (n=345,779), UK Biobank (n=45,595), Yonsei AF registry (n=5,814) datasets.

For predicting AI-ECG age, we utilized a pre-trained ResNet-based model. Previous research has demonstrated the ResNet's proficiency in identifying ECG abnormalities and estimating ECG-based age.[27, 28] Our AI model was trained using 85% of the ECG data (n=1,340,246) from the CODE study cohort. We then validated the model using four distinct multinational datasets that were not part of the training phase. These validation datasets comprised a total of 414,804 ECGs from CODE-15% (n=345,779), Physionet (n=21,799), Sami-Trop (n=1,631), and the UK Biobank (n=45,595). Further details regarding each validation cohort can be found in other sources.

For ECG pattern recognition, we employed Gradient-weighted Class Activation Mapping (Grad-CAM) to emphasize the ECG signatures identified by our age and sex prediction model. Additionally, we developed and implemented a black-box interpretation method to further analyze the ECG data (Figure 2). This method quantifies the significance of ECG components in 4,000 samples using calculations derived from Grad-CAM. It was designed to identify patterns associated with the prediction of repetitive PQRST rhythms observed in the ECG data over 10 seconds. We extracted ECG beats and identified the R peaks to isolate the PQRST rhythms utilizing the Neurokit2 library. Subsequently, we calculated the median beat, standardized it to a length of 600 samples, and extracted the ECG components (P waves, PR intervals, PR segments, QRS complexes, ST segments, QT intervals, and T waves). Finally, we visualized the importance of each ECG component by averaging the Grad-CAM importance scores for the respective intervals.

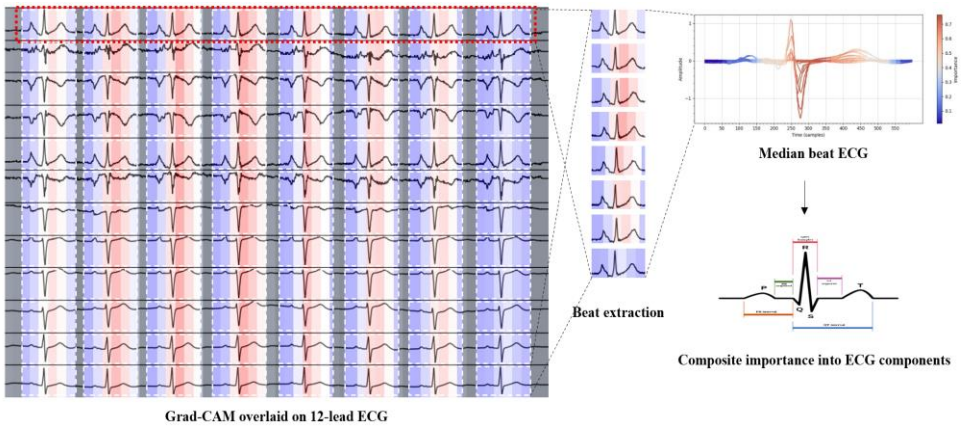

Figure 2. Black-box AI interpretation algorithm for long-term 12-lead ECG analysis

2.5. Statistical Analysis

Descriptive statistics were used to characterize baseline characteristics. Categorical variables are reported as numbers (percentages) and compared using the Chi-square or Fisher's exact test. Continuous variables were tested for a normal distribution using the Shapiro-Wilk or Kolmogorov-Smirnov tests. Continuous variables without a normal distribution are presented as medians with the interquartile range (IQR), while those with a normal distribution are presented as the mean ± standard deviation. Normally distributed continuous data were compared using unpaired Student's t-test. The Mann-Whitney U test was used to compare the continuous variables without a normal distribution between the two groups.

The association of the CHA2DS2-VASc score excluding history of stroke or transient ischemic attack (TIA), with the presence of stroke history at the time of catheter ablation was assessed by logistic regression. We used the area under curve (AUC) of receiver operating characteristics (ROC) and likelihood ratio test to assess the discrimination power and model fit of the logistic regression model. To assess amount of risk in PRS, multivariate logistic regression was performed adjusting for CHA_2DS_2-VASc score and AI-guided CHA_2DS_2-VASc score. Multiple subgroup analyses according to chronological age, biological sex, and underlying comorbid conditions were performed. All analyses were performed using R statistics, version 4.0.2 software (R Foundation for Statistical Computing); and a two-sided p-value < 0.05 was considered statistically significant.

III. RESULTS

3.1. Patient characteristics

The baseline clinical characteristics of the study population are listed in Table 1. There were 3,190 total study population with AF who undergoing AFCA, of which 320 had a history of ischemic stroke, and 2,870 did not. After reweighting, 530933 single nucleotide variants (SNVs) had a nonzero effect size and were included in our PRS.

Table 1. Baseline characteristics according to occurrence of early ischemic stroke before AF ablation.

	Overall (n=3190)	**No IS (n=2870)**	**IS (n=320)**	**p-value**
Age, years (median)	61 [53.0, 67.0]	60.0 [52.0, 67.0]	65.0 [59.0, 71.0]	<0.001
AI-ECG predicted age (median)	63.9 [52.5, 73.3]	63.1 [52.0, 72.5]	70.1 [59.9, 78.1]	<0.001
Male	2320 (72.7%)	2092 (72.9%)	228 (71.2%)	0.576
AI-ECG predicted male	2487 (78%)	2238 (78.0%)	249 (77.8%)	>0.999
Hypertension	1500 (47.0%)	1322 (46.1%)	178 (55.6%)	0.001
Diabetes mellitus	476 (14.9%)	402 (14.0%)	74 (23.1%)	<0.001
Heart failure	446 (14.0%)	394 (13.7%)	52 (16.2%)	0.251
Vascular disease	272 (8.5%)	222 (7.7%)	50 (15.6%)	<0.001
Age gap (mean)	3.05 ± 12.16	2.93 ± 12.27	4.15 ± 11.05	0.063
The aged ECG group (Age_gap≥10)	1367 (43.85%)	1239 (43.17%)	128 (40.00%)	0.304
AI sex mismatch	379 (11.88%)	338 (11.78%)	41 (12.81%)	0.651
CHA_2DS_2VASc score	1.0 [1.0, 2.0]	1.0 [0.0, 2.0]	2.0 [1.0, 3.0]	<0.001
AI-guided CHA_2DS_2VASc score	2.0 [1.0, 3.0]	1.0 [1.0, 3.0]	2.0 [1.0, 3.0]	<0.001
Multi-ethnic AF PRS, sum	6.1 [5.8, 6.4]	6.1 [5.8, 6.4]	6.1 [5.9, 6.4]	0.103
European AF PRS, sum	4.22 ± 0.31	4.22 ± 0.31	4.26 ± 0.32	0.041
Stroke PRS, sum	1.84 ± 0.15	1.84 ± 0.15	1.85 ± 0.15	0.403

† AF indicated atrial fibrillation; AI, artificial intelligence; ECG, electrocardiogram; PRS, polygenic risk score.

3.2. Stroke prediction

The logistic regression analyses for stroke were presented in Table 2. The analyses using AF PRS (multi-ethnic AF PRS sum) revealed a PRS odds ratio (OR), 1.33 per SD (95% CI, 1.01–1.77; Table 2 and Figure 3).

Table 2. Logistic regression for early ischemic stroke before AF ablation

Variables	**Odds ratio (95% CI)**	**P-value**
Age	1.05 (1.04-1.06)	<0.001
AI-ECG predicted age	1.04 (1.03-1.05)	<0.001
Male	0.92 (0.71-1.19)	0.532
AI-ECG predicted male	0.99 (0.75-1.31)	0.946
Hypertension	1.47 (1.16-1.85)	0.001
Diabetes mellitus	1.85 (1.40-2.44)	<0.001
Heart failure	1.22 (0.89-1.67)	0.218
Vascular disease	2.21 (1.59-3.08)	<0.001
CHA_2DS_2VASc score	1.35 (1.25-1.47)	<0.001
Multi-ethnic AF PRS, sum	1.33 (1.01-1.77)	0.044
European AF PRS, sum	1.48 (1.02-2.15)	0.041
Stroke PRS, sum	1.39 (0.64-2.99)	0.403

† AF indicated atrial fibrillation; AI, artificial intelligence; ECG, electrocardiogram; PRS, polygenic risk score.

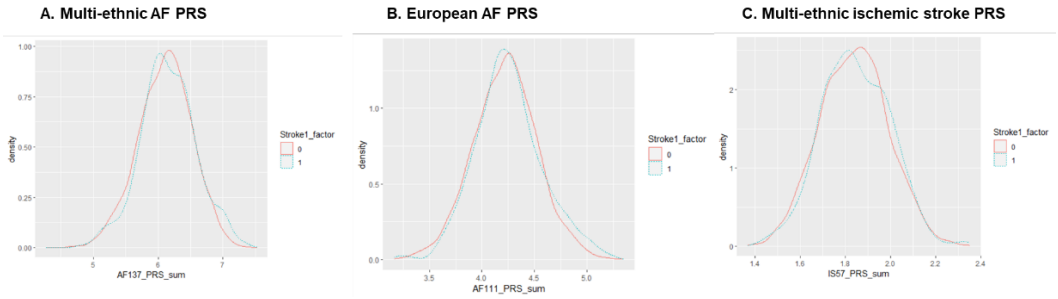

Figure 3. Polygenic risk score distribution. Histogram of participants with AF, color representing those that had an iscvhemic stroke (green) or not (red). PRS with multi-ethnic AF SNP (A), European AF SNP (B), and multi-ethnic ischemic stroke SNP (C).

Performance power analyses demonstrated an area under the receiver operating characteristics for CHA_2DS_2-VASc score (excluding the stroke/TIA component) of: 0.615 (95% CI, 0.583–0.647), with the addition of multi-ethnic AF PRS this rose to 0.621 (95% CI, 0.588–0.655) and corresponded to a PRS odds ratio of 1.44 per SD (95% CI, 1.09–1.91) (Table 3).

Table 3. Performance power of risk model for early ischemic stroke before AF catheter ablation

Risk model	Odd ratio (95% CI)	AUC (95% CI)	p-value in LR test
Overall AF patients			
CHA_2DS_2VASc score	1.35 (1.25-1.47) of CVS, p<0.001	0.615 (0.583-0.647)	Reference
CHA_2DS_2VASc score + Multi-ethnic AF PRS, sum	1.44 (1.09-1.91) of PRS, p=0.011	0.621 (0.588-0.655)	0.011
CHA_2DS_2VASc score + European AF PRS, sum	1.54 (1.06-2.25) of PRS, p=0.023	0.617 (0.583-0.650)	0.023
CHA_2DS_2VASc score + stroke PRS, sum	1.37 (0.63-2.97) of PRS, p=0.424	0.617 (0.583-0.650)	0.424
AI-guided CHA_2DS_2VASc	1.37 (1.26-1.48) of CVS, p<0.001	0.625 (0.592-0.658)	Reference
AI-guided CHA_2DS_2VASc + Multi-ethnic AF PRS, sum	1.43 (1.07-1.90) of PRS, p=0.014	0.632 (0.598-0.665)	0.014
AI-guided CHA_2DS_2VASc + European AF PRS, sum	1.52 (1.04-2.21) of PRS, p=0.029	0.626 (0.592-0.660)	0.029
AI-guided CHA_2DS_2VASc + Stroke PRS, sum	1.36 (0.63-2.95) of PRS, p=0.434	0.627 (0.593-0.660)	0.434

LR=likelihood ratio

The same analysis adjusting for age and sex by AI guided ECG showed a higher C statistic (0.625 [95% CI, 0.592–0.658]), and this rose to 0.632 (95% CI, 0.598–0.665) with the addition of multi-ethnic AF PRS. The integrated multi-ethnic AF PRS and AI-guided CHA_2DS_2-VASc risk model showed a significantly improved statistical fit ($\chi 2$ P<0.001) and modestly improved discrimination (Figure 4 and Table 4). Subgroup analyses were conducted based on age groups and the number of comorbidities, but no statistically significant differences were observed in any of the groups (Table 5).

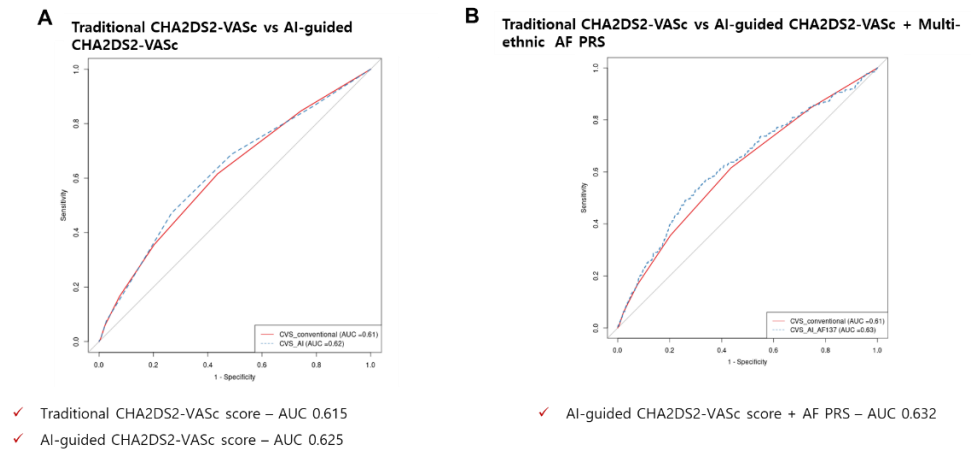

Figure 4. AUC-ROC curve of each risk model

Table 4. Model-Fit-LR test for the comparison between two models

Model1	Model2	LR_Chi	LR_P value
CVS_traditional_model	CVS_traditional _AF111_model	5.184573668	0.022788263
CVS_traditional_model	CVS_traditional _AF137_model	6.477965127	0.010921995
CVS_traditional_model	CVS_traditional _IS57_model	0.638124859	0.42439063
CVS_traditional_model	CVS_AI_AF111_model	12.29295633	0.000454671
CVS_traditional_model	CVS_AI_AF137_model	13.5956169	0.000226714
CVS_traditional_model	CVS_AI_IS57_model	8.129308168	0.004355536
CVS_traditional _AF111_model	CVS_AI_model	2.333040514	0.126654275
CVS_traditional _AF137_model	CVS_AI_model	1.039649055	0.307903106
CVS_traditional _IS57_model	CVS_AI_model	6.879489323	0.008719047
CVS_AI_model	CVS_AI_AF111_model	4.775342153	0.028870109
CVS_AI_model	CVS_AI_AF137_model	6.07800272	0.013687544
CVS_AI_model	CVS_AI_IS57_model	0.611693986	0.434150707

LR=likelihood ratio

Table 5. Subgroup analysis for performance power of risk model for early ischemic stroke before AF ablation

Risk model	Odd ratio (95% CI)	AUC (95% CI)	p-value in LR test
Low cormobidities AF patients (n=1294, stroke n=97)*			
CHA_2DS_2VASc score	1.39 (1.08-1.80), p=0.012	0.562	Reference
CHA_2DS_2VASc score + Multi-ethnic AF PRS, sum	0.96 (0.58-1.59) of PRS, p=0.881	0.568	0.881
CHA_2DS_2VASc score + European AF PRS, sum	0.97 (0.49-1.92) of PRS, p=0.937	0.566	0.937
CHA_2DS_2VASc score + Stroke PRS, sum	0.87 (0.22-3.49) of PRS, p=0.841	0.570	0.841
AI-guided CHA_2DS_2VASc	1.33 (1.05-1.68), p=0.016	0.553	Reference
AI-guided CHA_2DS_2VASc + Multi-ethnic AF PRS, sum	0.96 (0.58-1.58) of PRS, p=0.862	0.559	0.862
AI-guided CHA_2DS_2VASc + European AF PRS, sum	0.98 (0.50-1.92) of PRS, p=0.955	0.556	0.955
AI-guided CHA_2DS_2VASc + Stroke PRS, sum	0.81 (0.20-3.27) of PRS, p=0.767	0.560	0.767
Young AF patients (<65 years) (n=2082, stroke n=156)			
CHA_2DS_2VASc score	1.28 (1.07-1.52), p=0.007	0.558 (0.512-0.603)	Reference
CHA_2DS_2VASc score + Multi-ethnic AF PRS, sum	1.15 (0.77-1.72) of PRS, p=0.501	0.560 (0.511-0.609)	0.500
CHA_2DS_2VASc score + European AF PRS, sum	0.91 (0.53-1.56) of PRS, p=0.730	0.558 (0.510-0.605)	0.730
CHA_2DS_2VASc score + Stroke PRS, sum	1.13 (0.38-3.35) of PRS, p=0.821	0.559 (0.509-0.608)	0.821
AI-guided CHA_2DS_2VASc	1.33 (1.17-1.50), p<0.001	0.595 (0.547-0.643)	Reference
AI-guided CHA_2DS_2VASc + Multi-ethnic AF PRS, sum	1.18 (0.78-1.77) of PRS, p=0.429	0.596 (0.546-0.645)	0.428
AI-guided CHA_2DS_2VASc + European AF PRS, sum	0.93 (0.54-1.60) of PRS, p=0.788	0.598 (0.550-0.647)	0.788
AI-guided CHA_2DS_2VASc + Stroke PRS, sum	1.13 (0.38-3.33) of PRS, p=0.824	0.596 (0.546-0.646)	0.824

LR=likelihood ratio. *The AF patients after excluding hypertension, diabetes mellitus, heart failure, and vascular disease which are consisted of CHA2DS2VASc score.

IV. DISCUSSION

4.1. Main findings

In this study, we developed and validated a novel predictive model that integrates polygenic risk scores (PRS) with AI-guided analysis of age and sex to predict early-onset ischemic stroke in patients with atrial fibrillation (AF) undergoing catheter ablation. Our findings suggest that this combined approach offers modest but statistically significant improvements in stroke risk stratification over the conventional CHA_2DS_2-VASc scoring system. The integration of PRS into the risk model provided a significant enhancement in predictive performance, as indicated by the improvement in the area under the receiver operating characteristic curve (AUC). Specifically, the inclusion of the multi-ethnic AF PRS increased the AUC from 0.615 to 0.621, with a corresponding odds ratio of 1.44 per SD. These findings suggest the potential of PRS to capture genetic predispositions that are not accounted for by traditional clinical risk factors. Moreover, the AI-guided analysis, which incorporated age and sex predictions derived from ECG data, further improved the model's discriminative power. The final integrated model, which combined PRS with AI-guided age and sex adjustments, achieved the highest AUC of 0.632 and demonstrated a significantly better fit compared to the CHA_2DS_2-VASc score alone.

4.2. Genetic predisposition for ischemic stroke in AF

Although AF itself is a well-established risk factor for ischemic stroke, the variability in stroke occurrence among individuals with similar clinical profiles suggests that genetic factors may play a significant role in modulating this risk. However, the CHA_2DS_2-VASc score, the current risk stratification system, does not account for genetic risk of ischemic stroke, despite its significant heritability (about 40%). Genome-wide association studies (GWAS) have identified several genetic loci associated with AF, many of which are also implicated in stroke risk. In this study, the PRS for AF shows a stronger statistical association with ischemic stroke in AF patients compared to the stroke-specific PRS. These findings indicate a shared genetic basis between AF and ischemic stroke. Common variants in genes related to atrial remodeling, ion channel function, and inflammation may be associated with the pathophysiology of both conditions, and these genetic associations suggest that certain individuals may be predisposed to an elevated stroke risk due to their genetic background, independent of traditional clinical risk factors.

4.3. AI-ECG guided Age and Sex

In the epidemiology of AF,[37, 38] each sex is linked to distinct risk factors for both AF and stroke. For instance, men are more prone to coronary artery disease, left atrial enlargement, and thicker left ventricular walls, whereas women are more likely to have hypertension, valvular heart disease, and heart failure with preserved ejection fraction. These sex-specific risk factors can be identified from a single raw ECG tracing, and misclassifying the sex on an ECG could reveal risk factors typically linked to the opposite sex, possibly changing the ischemic stroke risk profile. Chronological aging leads to electroanatomic changes in the heart, and both AF and ischemic stroke are especially common in older populations. This suggests that some individuals may experience either accelerated or delayed cardiac aging relative to their chronological age. In this study, adjusting for age and sex—typically seen as non-modifiable factors—shows that these often overlooked variables can improve the accuracy of risk assessment. This highlights the importance of reevaluating their role in predictive models, which could enhance risk stratification and enable more personalized patient care.

While the use of AI-ECG has shown promise in various cardiovascular risk assessments, directly predicting ischemic stroke in AF patients based solely on ECG data presents significant challenges. Stroke risk in AF is influenced by a wide range of clinical variables, including comorbidities like hypertension, diabetes, and previous history of stroke, as well as structural heart changes that cannot be fully captured by ECG alone. These multifactorial influences make the direct prediction of ischemic stroke from ECG data a highly complex task. Moreover, in previous attempts to leverage AI-ECG for direct stroke prediction, the predictive accuracy was found to be lower than anticipated. Despite the potential of AI to identify subtle ECG changes, the inherent complexity of stroke risk—where multiple, non-ECG factors play a critical role—limited the model's ability to reliably predict stroke events. As a result, the focus of this study shifted toward more fundamental and well-established risk factors, namely age and sex. Both of these factors, though traditionally considered non-modifiable, have a profound impact on stroke risk in AF and are easily integrated into predictive models.

4.4. Clinical implications

Aggressive rhythm control in atrial fibrillation (AF) has been shown to reduce stroke incidence, improve various morbidities, and confer a mortality benefit. However, some patients still

experience a stroke before the initiation of rhythm control and catheter ablation. Identifying these patients early is crucial, as it not only improves individual outcomes but also has significant implications for reducing long-term healthcare costs. Early prediction and intervention could, therefore, play a pivotal role in both patient care and the broader socioeconomic impact of AF management.

4.5. Limitations

This study had several limitations. First, this was an observational cohort study from a single center that included a limited number of highly selected patients referred for AF ablation. Morevoer, ischemic stroke events were defined retrospectively as those that had already occurred at the time of the procedure, rather than prospectively. Second, the polygenic risk scores (PRS) were developed based on genetic data predominantly from populations of Japan, European, and multiethnic ancestry. The generalizability of these scores to other ethnicities may be limited, potentially leading to less accurate risk stratification in Korean populations. Third, while AI-guided models have demonstrated potential in improving predictive accuracy, the "black box" nature of AI algorithms may pose challenges in clinical practice. Lastly, the effectiveness of AI-guided models and PRS depends heavily on the quality and completeness of the input data. Missing or inaccurate data can significantly compromise the accuracy and reliability of the predictions made by the model.

V. CONCLUSIONS

This study developed and validated a novel predictive model that combines PRS with AI-guided analysis of age and sex to enhance ischemic stroke prediction in atrial fibrillation (AF) patients undergoing catheter ablation. Our results demonstrate that this integrated model significantly improves stroke risk stratification compared to the conventional CHA_2DS_2-VASc score. This approach provides a more refined tool for identifying patients at heightened risk, facilitating earlier recognition of potential stroke risk and enabling more proactive intervention strategies. Rather than predicting stroke events directly, this model supports pre-detection efforts aimed at better anticipating the development of stroke in high-risk individuals. Further research is needed to validate these findings across diverse populations and assess their practical application in routine care.

References

1. Abraham, G., Malik, R., Yonova-Doing, E., Salim, A., Wang, T., Danesh, J., Butterworth, A. S., Howson, J. M., Inouye, M., & Dichgans, M. (2019). Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. Nature communications, 10(1), 5819.
2. Attia, Z. I., Friedman, P. A., Noseworthy, P. A., Lopez-Jimenez, F., Ladewig, D. J., Satam, G., Pellikka, P. A., Munger, T. M., Asirvatham, S. J., Scott, C. G., Carter, R. E., & Kapa, S. (2019). Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs. Circ Arrhythm Electrophysiol, 12(9), e007284. https://doi.org/10.1161/circep.119.007284
3. Attia, Z. I., Noseworthy, P. A., Lopez-Jimenez, F., Asirvatham, S. J., Deshmukh, A. J., Gersh, B. J., Carter, R. E., Yao, X., Rabinstein, A. A., Erickson, B. J., Kapa, S., & Friedman, P. A. (2019). An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet, 394(10201), 861-867. https://doi.org/10.1016/s0140-6736(19)31721-0
4. Awasthi, S., Sachdeva, N., Gupta, Y., Anto, A. G., Asfahan, S., Abbou, R., Bade, S., Sood, S., Hegstrom, L., Vellanki, N., Alger, H. M., Babu, M., Medina-Inojosa, J. R., McCully, R. B., Lerman, A., Stampehl, M., Barve, R., Attia, Z. I., Friedman, P. A., . . . Lopez-Jimenez, F. (2023). Identification and risk stratification of coronary disease by artificial intelligence-enabled ECG. EClinicalMedicine, 65, 102259. https://doi.org/10.1016/j.eclinm.2023.102259
5. Beale, A. L., Meyer, P., Marwick, T. H., Lam, C. S. P., & Kaye, D. M. (2018). Sex Differences in Cardiovascular Pathophysiology: Why Women Are Overrepresented in Heart Failure With Preserved Ejection Fraction. Circulation, 138(2), 198-205. https://doi.org/10.1161/CIRCULATIONAHA.118.034271
6. Bevan, S., Traylor, M., Adib-Samii, P., Malik, R., Paul, N. L., Jackson, C., Farrall, M., Rothwell, P. M., Sudlow, C., Dichgans, M., & Markus, H. S. (2012). Genetic heritability of ischemic stroke and the contribution of previously reported candidate gene and genomewide associations. Stroke, 43(12), 3161-3167.

https://doi.org/10.1161/STROKEAHA.112.665760

7. Bumgarner, J. M., Lambert, C. T., Hussein, A. A., Cantillon, D. J., Baranowski, B., Wolski, K., Lindsay, B. D., Wazni, O. M., & Tarakji, K. G. (2018). Smartwatch Algorithm for Automated Detection of Atrial Fibrillation. J Am Coll Cardiol, 71(21), 2381-2388. https://doi.org/10.1016/j.jacc.2018.03.003
8. Camm, A. J., Naccarelli, G. V., Mittal, S., Crijns, H., Hohnloser, S. H., Ma, C. S., Natale, A., Turakhia, M. P., & Kirchhof, P. (2022). The Increasing Role of Rhythm Control in Patients With Atrial Fibrillation: JACC State-of-the-Art Review. J Am Coll Cardiol, 79(19), 1932-1948. https://doi.org/10.1016/j.jacc.2022.03.337
9. Cardoso, C. S., Sabino, E. C., Oliveira, C. D., de Oliveira, L. C., Ferreira, A. M., Cunha-Neto, E., Bierrenbach, A. L., Ferreira, J. E., Haikal, D. S., Reingold, A. L., & Ribeiro, A. L. (2016). Longitudinal study of patients with chronic Chagas cardiomyopathy in Brazil (SaMi-Trop project): a cohort profile. BMJ Open, 6(5), e011181. https://doi.org/10.1136/bmjopen-2016-011181
10. Chiang, C. E., Wang, T. D., Li, Y. H., Lin, T. H., Chien, K. L., Yeh, H. I., Shyu, K. G., Tsai, W. C., Chao, T. H., Hwang, J. J., Chiang, F. T., & Chen, J. H. (2010). 2010 guidelines of the Taiwan Society of Cardiology for the management of hypertension. J Formos Med Assoc, 109(10), 740-773. https://doi.org/S0929-6646(10)60120-9 [pii]
11. 10.1016/S0929-6646(10)60120-9
12. Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. Nature protocols, 15(9), 2759-2772.
13. Deo, R. C. (2015). Machine Learning in Medicine. Circulation, 132(20), 1920-1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593
14. Elias, P., Poterucha, T. J., Rajaram, V., Moller, L. M., Rodriguez, V., Bhave, S., Hahn, R. T., Tison, G., Abreau, S. A., Barrios, J., Torres, J. N., Hughes, J. W., Perez, M. V., Finer, J., Kodali, S., Khalique, O., Hamid, N., Schwartz, A., Homma, S., . . . Perotte, A. J. (2022). Deep Learning Electrocardiographic Analysis for Detection of Left-Sided Valvular Heart Disease. J Am Coll Cardiol, 80(6), 613-626. https://doi.org/10.1016/j.jacc.2022.05.029
15. Ellinor, P. T., Lunetta, K. L., Albert, C. M., Glazer, N. L., Ritchie, M. D., Smith, A. V.,

Arking, D. E., Muller-Nurasyid, M., Krijthe, B. P., Lubitz, S. A., Bis, J. C., Chung, M. K., Dorr, M., Ozaki, K., Roberts, J. D., Smith, J. G., Pfeufer, A., Sinner, M. F., Lohman, K., . . . Kaab, S. (2012). Meta-analysis identifies six new susceptibility loci for atrial fibrillation. Nat Genet, 44(6), 670-675. https://doi.org/10.1038/ng.2261

16. Elliott, J., Bodinier, B., Bond, T. A., Chadeau-Hyam, M., Evangelou, E., Moons, K. G., Dehghan, A., Muller, D. C., Elliott, P., & Tzoulaki, I. (2020). Predictive accuracy of a polygenic risk score–enhanced prediction model vs a clinical risk score for coronary artery disease. Jama, 323(7), 636-645.
17. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118. https://doi.org/10.1038/nature21056
18. Gage, B. F., Waterman, A. D., Shannon, W., Boechler, M., Rich, M. W., & Radford, M. J. (2001). Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. Jama, 285(22), 2864-2870.
19. Ganesan, A. N., Chew, D. P., Hartshorne, T., Selvanayagam, J. B., Aylward, P. E., Sanders, P., & McGavigan, A. D. (2016). The impact of atrial fibrillation type on the risk of thromboembolism, mortality, and bleeding: a systematic review and meta-analysis. Eur Heart J, 37(20), 1591-1602. https://doi.org/10.1093/eurheartj/ehw007
20. Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C. K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation, 101(23), E215-220. https://doi.org/10.1161/01.cir.101.23.e215
21. Hindricks, G., Potpara, T., Dagres, N., Arbelo, E., Bax, J. J., Blomstrom-Lundqvist, C., Boriani, G., Castella, M., Dan, G. A., Dilaveris, P. E., Fauchier, L., Filippatos, G., Kalman, J. M., La Meir, M., Lane, D. A., Lebeau, J. P., Lettino, M., Lip, G. Y. H., Pinto, F. J., . . . Group, E. S. C. S. D. (2021). 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, 42(5), 373-498. https://doi.org/10.1093/eurheartj/ehaa612

22. Hirota, N., Suzuki, S., Motogi, J., Nakai, H., Matsuzawa, W., Takayanagi, T., Umemoto, T., Hyodo, A., Satoh, K., Arita, T., Yagi, N., Otsuka, T., & Yamashita, T. (2023). Cardiovascular events and artificial intelligence-predicted age using 12-lead electrocardiograms. Int J Cardiol Heart Vasc, 44, 101172. https://doi.org/10.1016/j.ijcha.2023.101172
23. Inouye, M., Abraham, G., Nelson, C. P., Wood, A. M., Sweeting, M. J., Dudbridge, F., Lai, F. Y., Kaptoge, S., Brozynska, M., Wang, T., Ye, S., Webb, T. R., Rutter, M. K., Tzoulaki, I., Patel, R. S., Loos, R. J. F., Keavney, B., Hemingway, H., Thompson, J., . . . Group, U. K. B. C. C. C. W. (2018). Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol, 72(16), 1883-1893. https://doi.org/10.1016/j.jacc.2018.07.079
24. January, C. T., Wann, L. S., Alpert, J. S., Calkins, H., Cigarroa, J. E., Cleveland, J. C., Jr., Conti, J. B., Ellinor, P. T., Ezekowitz, M. D., Field, M. E., Murray, K. T., Sacco, R. L., Stevenson, W. G., Tchou, P. J., Tracy, C. M., Yancy, C. W., & American College of Cardiology/American Heart Association Task Force on Practice, G. (2014). 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. J Am Coll Cardiol, 64(21), e1-76. https://doi.org/10.1016/j.jacc.2014.03.022
25. Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., & Ellinor, P. T. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature genetics, 50(9), 1219-1224.
26. Khera, A. V., Chaffin, M., Aragam, K. G., Haas, M. E., Roselli, C., Choi, S. H., Natarajan, P., Lander, E. S., Lubitz, S. A., Ellinor, P. T., & Kathiresan, S. (2018). Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet, 50(9), 1219-1224.

https://doi.org/10.1038/s41588-018-0183-z

27. Kirchhof, P., Benussi, S., Kotecha, D., Ahlsson, A., Atar, D., Casadei, B., Castella, M., Diener, H. C., Heidbuchel, H., Hendriks, J., Hindricks, G., Manolis, A. S., Oldgren, J., Popescu, B. A., Schotten, U., Van Putte, B., Vardas, P., Agewall, S., Camm, J., . . . Zeppenfeld, K. (2016). 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. Europace, 18(11), 1609-1678. https://doi.org/10.1093/europace/euw295
28. Kirchhof, P., Camm, A. J., Goette, A., Brandes, A., Eckardt, L., Elvan, A., Fetsch, T., van Gelder, I. C., Haase, D., Haegeli, L. M., Hamann, F., Heidbuchel, H., Hindricks, G., Kautzner, J., Kuck, K. H., Mont, L., Ng, G. A., Rekosz, J., Schoen, N., . . . Investigators, E.-A. T. (2020). Early Rhythm-Control Therapy in Patients with Atrial Fibrillation. N Engl J Med, 383(14), 1305-1316. https://doi.org/10.1056/NEJMoa2019422
29. Ko, D., Rahman, F., Schnabel, R. B., Yin, X., Benjamin, E. J., & Christophersen, I. E. (2016). Atrial fibrillation in women: epidemiology, pathophysiology, presentation, and prognosis. Nat Rev Cardiol, 13(6), 321-332. https://doi.org/10.1038/nrcardio.2016.45
30. Ko, W. Y., Siontis, K. C., Attia, Z. I., Carter, R. E., Kapa, S., Ommen, S. R., Demuth, S. J., Ackerman, M. J., Gersh, B. J., Arruda-Olson, A. M., Geske, J. B., Asirvatham, S. J., Lopez-Jimenez, F., Nishimura, R. A., Friedman, P. A., & Noseworthy, P. A. (2020). Detection of Hypertrophic Cardiomyopathy Using a Convolutional Neural Network-Enabled Electrocardiogram. J Am Coll Cardiol, 75(7), 722-733. https://doi.org/10.1016/j.jacc.2019.12.030
31. Lee, C. H., & Yoon, H. J. (2017). Medical big data: promise and challenges. Kidney Res Clin Pract, 36(1), 3-11. https://doi.org/10.23876/j.krcp.2017.36.1.3
32. Lima, E. M., Ribeiro, A. H., Paixão, G. M. M., Ribeiro, M. H., Pinto-Filho, M. M., Gomes, P. R., Oliveira, D. M., Sabino, E. C., Duncan, B. B., Giatti, L., Barreto, S. M., Meira Jr, W., Schön, T. B., & Ribeiro, A. L. P. (2021). Deep neural network-estimated electrocardiographic age as a mortality predictor. Nature Communications, 12(1), 5117. https://doi.org/10.1038/s41467-021-25351-7
33. Lima, E. M., Ribeiro, A. H., Paixão, G. M. M., Ribeiro, M. H., Pinto-Filho, M. M., Gomes, P. R., Oliveira, D. M., Sabino, E. C., Duncan, B. B., Giatti, L., Barreto, S. M.,

Meira, W., Jr., Schön, T. B., & Ribeiro, A. L. P. (2021). Deep neural network-estimated electrocardiographic age as a mortality predictor. Nat Commun, 12(1), 5117. https://doi.org/10.1038/s41467-021-25351-7

34. Lip, G. Y., Nieuwlaat, R., Pisters, R., Lane, D. A., & Crijns, H. J. (2010). Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. Chest, 137(2), 263-272. https://doi.org/10.1378/chest.09-1584
35. Miyasaka, Y., Barnes, M. E., Gersh, B. J., Cha, S. S., Bailey, K. R., Abhayaratna, W. P., Seward, J. B., & Tsang, T. S. (2006). Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. Circulation, 114(2), 119-125. https://doi.org/10.1161/circulationaha.105.595140
36. National-Institute-for-Health-and-Care-Excellence. (2014). Atrial fibrillation: the management of atrial fibrillation. (Clinical guideline 180.) 2014. http://guidance.nice.org.uk/CG180., http://guidance.nice.org.uk/CG180. http://guidance.nice.org.uk/CG180.
37. O'Sullivan, J. W., Shcherbina, A., Justesen, J. M., Turakhia, M., Perez, M., Wand, H., Tcheandjieu, C., Clarke, S. L., Rivas, M. A., & Ashley, E. A. (2021). Combining Clinical and Polygenic Risk Improves Stroke Prediction Among Individuals With Atrial Fibrillation. Circ Genom Precis Med, 14(3), e003168. https://doi.org/10.1161/CIRCGEN.120.003168
38. Ribeiro, A. H., Ribeiro, M. H., Paixão, G. M. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., Ferreira, M. P. S., Andersson, C. R., Macfarlane, P. W., Meira Jr, W., Schön, T. B., & Ribeiro, A. L. P. (2020). Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications, 11(1), 1760. https://doi.org/10.1038/s41467-020-15432-4
39. Ribeiro, A. L. P., Paixão, G. M. M., Gomes, P. R., Ribeiro, M. H., Ribeiro, A. H., Canazart, J. A., Oliveira, D. M., Ferreira, M. P., Lima, E. M., Moraes, J. L., Castro, N., Ribeiro, L. B., & Macfarlane, P. W. (2019). Tele-electrocardiography and bigdata: The CODE (Clinical Outcomes in Digital Electrocardiography) study. J Electrocardiol, 57s,

S75-s78. https://doi.org/10.1016/j.jelectrocard.2019.09.008

40. Roselli, C., Chaffin, M. D., Weng, L. C., Aeschbacher, S., Ahlberg, G., Albert, C. M., Almgren, P., Alonso, A., Anderson, C. D., Aragam, K. G., Arking, D. E., Barnard, J., Bartz, T. M., Benjamin, E. J., Bihlmeyer, N. A., Bis, J. C., Bloom, H. L., Boerwinkle, E., Bottinger, E. B., . . . Ellinor, P. T. (2018). Multi-ethnic genome-wide association study for atrial fibrillation. Nat Genet, 50(9), 1225-1233. https://doi.org/10.1038/s41588-018-0133-9
41. Sangha, V., Nargesi, A. A., Dhingra, L. S., Khunte, A., Mortazavi, B. J., Ribeiro, A. H., Banina, E., Adeola, O., Garg, N., Brandt, C. A., Miller, E. J., Ribeiro, A. L. P., Velazquez, E. J., Giatti, L., Barreto, S. M., Foppa, M., Yuan, N., Ouyang, D., Krumholz, H. M., & Khera, R. (2023). Detection of Left Ventricular Systolic Dysfunction From Electrocardiographic Images. Circulation, 148(9), 765-777. https://doi.org/10.1161/circulationaha.122.062646
42. Schnabel, R. B., Yin, X., Gona, P., Larson, M. G., Beiser, A. S., McManus, D. D., Newton-Cheh, C., Lubitz, S. A., Magnani, J. W., Ellinor, P. T., Seshadri, S., Wolf, P. A., Vasan, R. S., Benjamin, E. J., & Levy, D. (2015). 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study. Lancet, 386(9989), 154-162. https://doi.org/10.1016/s0140-6736(14)61774-8
43. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. International Journal of Computer Vision, 128(2), 336-359. https://doi.org/10.1007/s11263-019-01228-7
44. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis. IEEE J Biomed Health Inform, 22(5), 1589-1604. https://doi.org/10.1109/JBHI.2017.2767063
45. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK biobank: an open access resource for

identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med, 12(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

46. Svennberg, E., Engdahl, J., Al-Khalili, F., Friberg, L., Frykman, V., & Rosenqvist, M. (2015). Mass Screening for Untreated Atrial Fibrillation: The STROKESTOP Study. Circulation, 131(25), 2176-2184. https://doi.org/10.1161/CIRCULATIONAHA.114.014343
47. van Doorn, S., Debray, T. P. A., Kaasenbrood, F., Hoes, A. W., Rutten, F. H., Moons, K. G. M., & Geersing, G. J. (2017). Predictive performance of the CHA2DS2-VASc rule in atrial fibrillation: a systematic review and meta-analysis. J Thromb Haemost, 15(6), 1065-1077. https://doi.org/10.1111/jth.13690
48. Vaziri, S. M., Larson, M. G., Benjamin, E. J., & Levy, D. (1994). Echocardiographic predictors of nonrheumatic atrial fibrillation. The Framingham Heart Study. Circulation, 89(2), 724-730. https://doi.org/10.1161/01.cir.89.2.724
49. Wand, H., Lambert, S. A., Tamburro, C., Iacocca, M. A., O'Sullivan, J. W., Sillari, C., Kullo, I. J., Rowley, R., Dron, J. S., & Brockman, D. (2021). Improving reporting standards for polygenic scores in risk prediction studies. Nature, 591(7849), 211-219.
50. Wolf, P. A., Abbott, R. D., & Kannel, W. B. (1991). Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. Stroke, 22(8), 983-988.

Abstract in Korean

심방세동에서 뇌졸중 예측 향상: 다유전자 위험 점수와 인공지능 기반 임상 모델의 통합

배경: 심방세동은 허혈성 뇌졸중의 위험을 크게 증가시키며, 현재 사용되는 위험 계층화 모델인 CHA_2DS_2-VASc 점수체계는 정확한 예측에는 뚜렷한 한계가 있다. 본 연구는 다유전자 위험 점수(Polygenic Risk Score, PRS)와 나이 및 성별에 대한 인공지능(AI) 기반 분석을 결합하여 심방세동 환자의 뇌졸중 위험 예측을 개선하는 예측 모델을 개발하는 것을 목표로 하였다.

방법: 본 연구는 연세의대 세브란스병원에서 심방세동 절제술을 시행받은 환자 코호트(Yonsei AF Ablation Cohort)에 속한 3,190명의 심방세동 환자를 대상으로 했다. 환자들은 심방세동 절제술 전 조기 허혈성 뇌졸중 병력이 있는 320명의 환자와 뇌졸중 병력이 없는 2,870명의 대조군으로 나누어졌다. 우리는 유전체 연관 연구(GWAS)에 기반한 PRS를 개발하고 이를 심전도 데이터를 기반으로 한 나이 및 성별에 대한 AI 기반 분석과 결합했다. 통합된 모델의 허혈성 뇌졸중 병력 예측 성능은 CHA_2DS_2-VASc 점수(뇌졸중/TIA 구성 요소 제외)와 비교하여 로지스틱 회귀 분석 및 수신자 조작 특성 곡선(AUC) 지표를 사용해 평가했다.

결과: 다인종 심방세동 PRS와 AI 기반 나이 및 성별 예측을 결합한 통합 모델은 CHA_2DS_2-VASc 점수만 사용한 것보다 뇌졸중 위험 계층화에 있어 유의미한 개선을 보였다. PRS를 포함했을 때 AUC가 0.615에서 0.621로 증가했으며, AI 기반 분석을 추가했을 때 AUC는 0.632로 더 증가하였다. 다인종 심방세동 PRS와 AI 기반 CHA_2DS_2-VASc 위험 모델은 통계적으로 유의미한 개선($\chi 2$ $P<0.001$)과 소폭의 판별력 향상을 보였다.

결론: PRS와 AI 기반 나이 및 성별 분석을 통합함으로써 심방세동 환자의 위험 계층화가 개선되었으며, 허혈성 뇌졸중의 고위험군을 더 정확하게 식별할 수 있게 되었다. 이러한 접근 방식은 조기 발견 모델을 향상시키며, 더 적극적인 예방 조치를 가능하게 할 것으로 기대된다.

핵심되는 말 : 심방세동, 뇌졸중, 인공지능, 다유전자 위험점수