ORIGINAL ARTICLE

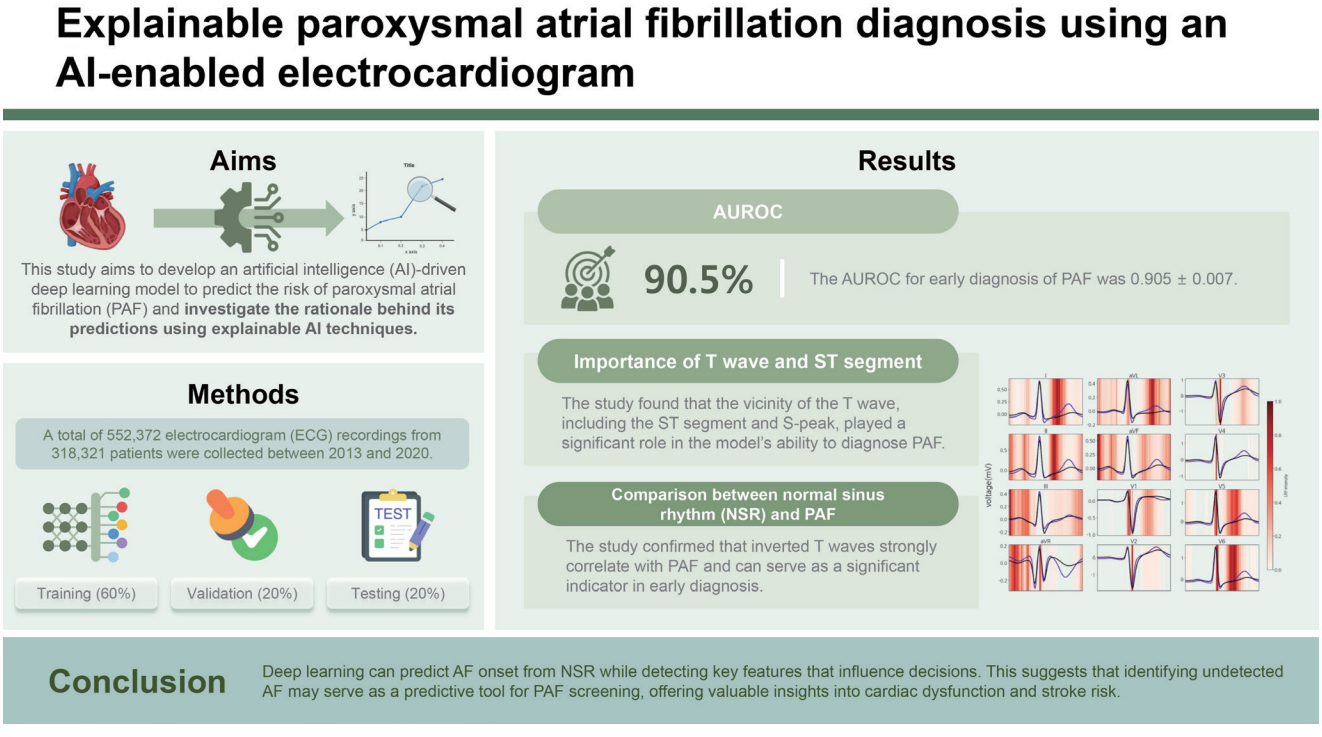
Korean J Intern Med 2025;40:251-261
https://doi.org/10.3904/kjim.2024.130

Explainable paroxysmal atrial fibrillation diagnosis using an artificial intelligence-enabled electrocardiogram

Yeongbong Jin[1,*], Bonggyun Ko[2,3,*], Woojin Chang[1], Kang-Ho Choi[4,5], and Ki Hong Lee[6,7]

[1]Department of Industrial Engineering, Seoul National University, Seoul; [2]Department of Mathematics and Statistics, Chonnam National University, Gwangju; [3]XRAI, Gwangju; [4]Department of Neurology, Chonnam National University Hospital, Gwangju; [5]Department of Neurology, Chonnam National University Medical School, Gwangju; [6]Department of Internal Medicine, Chonnam National University Hospital, Gwangju; [7]Department of Internal Medicine, Chonnam National University Medical School, Gwangju, Korea

*These authors contributed equally to this manuscript.



Background/Aims: Atrial fibrillation (AF) significantly contributes to global morbidity and mortality. Paroxysmal atrial fibrillation (PAF) is particularly common among patients with cryptogenic strokes or transient ischemic attacks and has a silent nature. This study aims to develop reliable artificial intelligence (AI) algorithms to detect early signs of AF in patients with normal sinus rhythm (NSR) using a 12-lead electrocardiogram (ECG).

Methods: Between 2013 and 2020, 552,372 ECG traces from 318,321 patients were collected and split into training (n = 331,422), validation (n = 110,475), and test sets (n = 110,475). Deep neural networks were then trained to predict AF onset within one month of NSR. Model performance was evaluated using the area under the receiver operating characteristic curve

pISSN 1226-3303
eISSN 2005-6648
http://www.kjim.org

(AUROC). An explainable AI technique was employed to identify the inference evidence underlying the predictions of deep learning models.

Results: The AUROC for early diagnosis of PAF was 0.905 ± 0.007. The findings reveal that the vicinity of the T wave, including the ST segment and S-peak, significantly influences the ability of the trained neural network to diagnose PAF. Additionally, comparing the summarized ECG in NSR with those in PAF revealed that nonspecific ST-T abnormalities and inverted T waves were associated with PAF.

Conclusions: Deep learning can predict AF onset from NSR while detecting key features that influence decisions. This suggests that identifying undetected AF may serve as a predictive tool for PAF screening, offering valuable insights into cardiac dysfunction and stroke risk.

Keywords: Atrial fibrillation; Deep learning; Electrocardiography; Artificial intelligence; Paroxysmal atrial fibrillation

INTRODUCTION

Atrial fibrillation (AF), a prevalent form of arrhythmia, is associated with severe cardiovascular conditions and stands as a leading cause of mortality and morbidity [1,2]. AF significantly contributes to ischemic stroke associated with thromboembolism, a risk that anticoagulation can effectively prevent [3-8]. Nonetheless, AF often progresses without any clinical manifestation. Therefore, early detection and diagnosis of AF from normal conditions could support a comprehensive management system, potentially improving survival rates and alleviating disease burden.

Electrocardiogram (ECG) is the most widely used method for cardiovascular diagnostics and can offer significant prognostic insights [9]. However, screening for AF remains challenging, as many patients exhibit paroxysmal and asymptomatic features [1,2,10-12]. Although intermittent ECG screening or opportunistic pulse palpation provides a common, cost-effective approach to detect AF, several cases go undetected, and identifying the prevalence of AF poses a more fundamental issue than the choice of a screening strategy [13-15] Certain features of ECG, especially P waves, are often used to diagnose early AF [16-18]. However, they lack sufficient probability to be clinically useful in statistical models [7]. Machine learning algorithms, such as deep neural networks, can address these limitations by uncovering intricate patterns within large-scale datasets, and they demonstrate effectiveness in tasks including early AF detection and ECG classification [19-23].

Deep learning approaches advance beyond traditional pattern-based methods to detect paroxysmal atrial fibrillation (PAF) [6-8]. When these models are trained, probabilistically distinguishing PAF by learning data features from large ECG datasets is possible. However, studies show that the current deep learning systems prioritize reducing prediction errors over providing the significance of features or explaining what drives the networks [7,8]. This focus limits their operational usefulness and reduces the reliability of deep learning outputs in healthcare.

Therefore, this study aims to develop an artificial intelligence (AI) model to distinguish subtle patterns in a standard 12-lead ECG that are imperceptible to the human eye. To test this, we trained a deep neural network using a large cohort of patients from a tertiary hospital. Additionally, we interpreted the inference engine of the model to uncover the basis of its decisions.

METHODS

Data sources and study population

We extracted standard 12-lead ECGs for patients with at least one instance of normal sinus rhythm (NSR) recorded between May 16, 2013 and December 31, 2020. Each ECG was captured at a 500 Hz sampling rate, with raw data stored in the MUSE cardiology information system (GE Healthcare, Chicago, IL, USA).

The extracted ECG data included 10-s recordings. The quantitative measurements from ECG clinical reports were analyzed to identify diagnostic class and 18 patient features. These features were age, sex, ventricular/atrial rate, QRS duration/count, QTc (Bazett/Fridericia correction), QT interval, PR interval, the axes of P, R, and T waves, T-offset and on/offset of P and Q. Each variable had a missing value rate

https://doi.org/10.3904/kjim.2024.130

of 0–24.7%, with the highest percentage being sex. For cases with missing values, due to reasons such as privacy constraints, we imputed values to denote the absence of information.

In the prepared data, we excluded 95,398 ECGs where patient ID was not tracked and 287,247 ECGs where diagnostic classification was unavailable for AF or NSR. Rhythm diagnosis and labeling were carefully performed by clinical experts to ensure reliability before further analysis. These annotations were further validated using the electronic medical records of the patients, cross-referencing for AF diagnosis codes or documented history. Only AF diagnosis codes or previous history without a documented 12-lead ECG were not classified as PAF.

In the ECG clinical report of a patient, NSR recorded within 31 days following the first AF episode was labeled as "PAF." This broader definition, extending beyond the standard definition of PAF (AF episodes lasting under 7 days without intervention), enabled a comprehensive assessment of early AF episodes. The ECGs recorded more than 31 days after the first AF episode and those taken before the first onset were excluded. ECGs with consistent NSR across clinical records were annotated as 'Normal.' After annotation, we categorized patients into PAF and control groups and then deidentified the ECG data to remove personal ID. This approach ensured that subtle ECG patterns related to PAF were robustly identified, independent of individual characteristics.

We initially extracted 1,014,617 raw ECGs from 422,664 patients. From these, 95,398 ECGs with incomplete PID tracking or missing data were excluded, as well as 287,247 ECGs without AF records or NSR. Additionally, sinus rhythm ECGs immediately preceding the first AF event were excluded to avoid recording transitional ECG patterns that may not accurately reflect the early characteristics of PAF [7], thereby reducing noise and enhancing model prediction accuracy. ECGs were annotated as normal if all serial recordings for a patient showed NSR, while ECG records within the target window were labeled as PAF. After applying certain exclusions (Fig. 1), the dataset included 552,372 ECGs from 318,321 patients. Applying predefined labeling protocol, 26,541 ECGs (4.8%) were annotated as PAF, reflecting similar prevalence rates in the general population [24]. The processed ECGs were divided into training (60%), validation (20%), and test datasets (20%) for model training.

Algorithm development and evaluation for PAF early detection

The convolutional neural network (CNN)-based statistical model for early PAF detection was developed using patient demographics and raw ECG data as input. The CNN model was designed to identify patterns and extract local spatial features from global maps using filters that apply each input subregion through dot product operations [25]. Given the seasonality and fixed length of ECG signals, 1d-CNN is the most appropriate to apply [26]. The network architec-

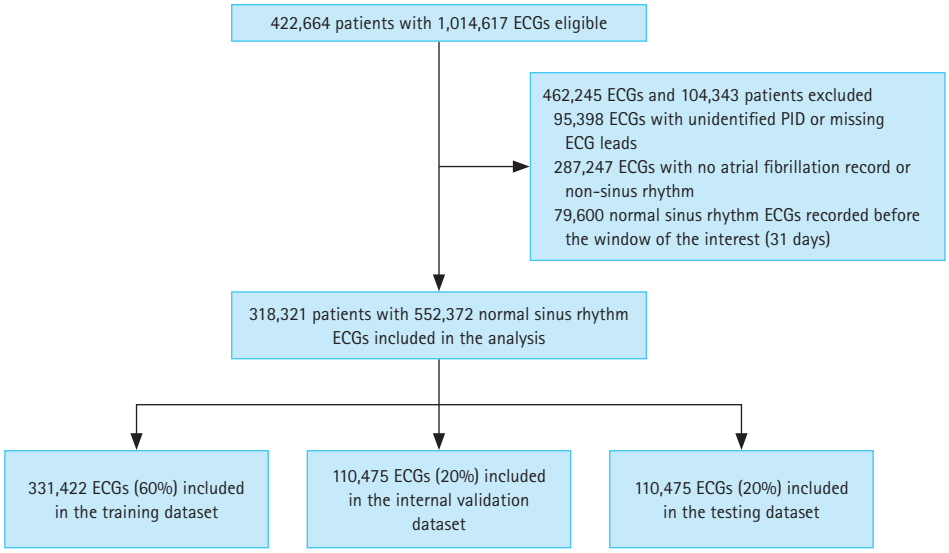


Figure 1. Summary of data used in the study. Diagram summarizing patient selection for training, validation, and testing cohorts. ECG, electrocardiogram; PID, patient identification.

ture for analyzing standard 12-lead ECGs (sampled at 500 Hz) included 50 convolutional layers, using the skip-connection method of the residual network [27] to ensure effective training. The model we used consisted of 16 residual blocks, each containing three convolution layers, followed by a Batch Normalization and rectified linear activation function (ReLU) applied after each layer [28,29]. Where the output feature map dimensions decrease, a convolutional layer with a stride of 2 was applied to ensure alignment between input and output dimensions. Pooling was conducted at the first and last layers after the nonlinear activation function. A feature vector was then extracted from the ECG by applying two fully connected (FC) layers with ReLU activations. Finally, this feature vector was concatenated with a 12-lead ECG feature vector, which produced class probabilities via a final FC layer and softmax function. The network weights were initialized following the He method, and the model was optimized using the Adam algorithm with default parameters $\beta1 = 0.9$ and $\beta2 = 0.999$ [30,31]. The batch size was 32, with a learning rate of 1×10^{-4}, halved every 10 epochs. Figure 2 depicts the overall architect of our model.

Additionally, focal loss was used to address data imbalance. This loss function—a variation of cross entropy—counteracts extreme interclass imbalances by assigning lower weights to easily classified negatives to reduce their contribution to learning while increasing weights for challenging positives that are harder to classify [32]. Hyperparameters for the network structure and loss function used were set through grid search and manual tuning.

We trained 30 models using different dataset configurations generated through pseudo-random number sampling. The model performance was evaluated using the area under the curve (AUC) metric, a key performance indicator in binary classification where a higher AUC indicates better performance. To further evaluate the generalizability and stability of the model, we recorded the AUC of the developed models.

Identify contributors to PAF early detection through LRP

The interpretability of complex deep-learning models remains a major concern in the medical field. Enhancing model transparency can improve acceptance in clinical decision-making and justify specific diagnoses and treatment recommendations [33]. Layer-wise relevance propagation (LRP) is a prominent method for interpreting complex

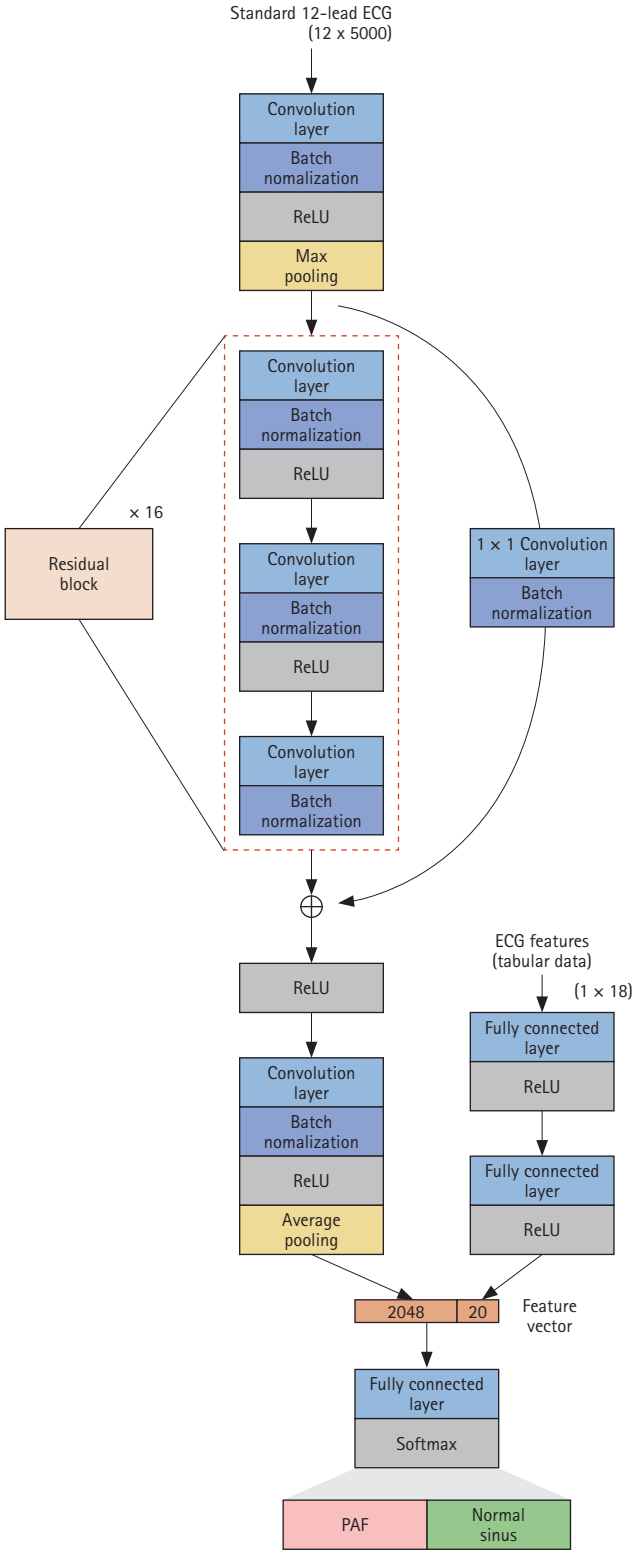


Figure 2. Network architecture. Network architecture used in the study. Our network processes raw ECG data (sampled at 500 Hz) and ECG features to predict AF onset within a month. ECG, electrocardiogram; ReLU, rectified linear activation function; PAF, paroxysmal atrial fibrillation; AF, atrial fibrillation.

https://doi.org/10.3904/kjim.2024.130

Table 1. Comparison of patient characteristics

Variable	Total dataset			Training dataset			Validation dataset			Test dataset		
	Normal (n = 525,831)	PAF (n = 26,541)	*p* value	Normal (n = 315,535)	PAF (n = 15,887)	*p* value	Normal (n = 105,122)	PAF (n = 5,353)	*p* value	Normal (n = 105,174)	PAF (n = 5,301)	*p* value
Female	193,890 (48.8)	6,919 (36.5)	< 0.001	116,393 (48.8)	4,103 (36.3)	< 0.001	38,589 (48.7)	1,429 (37.5)	< 0.001	38,908 (49.0)	1,387 (36.0)	< 0.001
Age (yr)	55. 4 ± 19.3	67.6 ± 12.4	< 0.001	55.5 ± 19.3	67.7 ± 12.5	< 0.001	55.5 ± 19.3	67.7 ± 12.3	< 0.001	55.4 ± 19.3	67.2 ± 12.5	< 0.001
Ventricular rate (bpm)	75.7 ± 12.7	75.7 ± 10.7	0.751	75.7 ± 12.7	75.7 ± 10.7	0.674	75.7 ± 12.8	75.8 ± 10.7	0.542	75.7 ± 12.7	75.6 ± 10.7	0.553
Atrial rate (bpm)	75.7 ± 12.8	75.7 ± 10.7	0.702	75.7 ± 12.7	75.7 ± 10.7	0.632	75.7 ± 12.8	75.8 ± 10.7	0.568	75.7 ± 12.7	75.6 ± 10.7	0.549
QRS duration (ms)	92.5 ± 15.2	99.5 ± 21.0	< 0.001	92.5 ± 15.3	99.6 ± 21.2	< 0.001	92.5 ± 15.1	99.4 ± 20.7	< 0.001	92.4 ± 15.1	99.1 ± 20.4	< 0.001
QT interval	390.9 ± 34.7	417.2 ± 45.4	< 0.001	390.9 ± 34.7	417.3 ± 45.3	< 0.001	390.9 ± 34.6	417.8 ± 45.3	< 0.001	390.8 ± 34.6	417.0 ± 45.9	< 0.001
QT corrected	434.8 ± 30.3	464.6 ± 42.9	< 0.001	434.8 ± 30.4	464.8 ± 43.1	< 0.001	434.9 ± 30.2	464.6 ± 42.9	< 0.001	434.8 ± 30.2	464.1 ± 42.6	< 0.001
PR interval	158.6 ± 21.6	166.9 ± 22.3	< 0.001	158.7 ± 21.6	166.9 ± 22.3	< 0.001	158.9 ± 21.6	166.8 ± 22.1	< 0.001	158.6 ± 21.6	166.8 ± 22.6	< 0.001
P axis	49.6 ± 21.3	49.1 ± 25.5	0.0019	49.6 ± 21.3	49.3 ± 25.5	0.100	49.7 ± 21.2	49.0 ± 25.5	0.081	49.6 ± 21.3	48.8 ± 25.4	0.0189
R axis	40.2 ± 38.6	31.8 ± 45.4	< 0.001	40.0 ± 38.6	31.7 ± 45.7	< 0.001	40.1 ± 38.6	32.4 ± 45.1	< 0.001	40.3 ± 38.6	31.4 ± 44.8	< 0.001
T axis	48.0 ± 37.3	66.9 ± 64.7	< 0.001	48.0 ± 37.2	66.9 ± 64.8	< 0.001	48.0 ± 37.4	66.0 ± 64.8	< 0.001	47.9 ± 37.3	67.8 ± 64.4	< 0.001
Q onset	219.2 ± 6.2	217.6 ± 6.4	< 0.001	219.2 ± 6.2	217.5 ± 6.5	< 0.001	219.2 ± 6.2	217.6 ± 6.5	< 0.001	219.2 ± 6.2	217.5 ± 6.3	< 0.001
Q offset	265.4 ± 8.0	267.3 ± 9.9	< 0.001	265.4 ± 8.0	267.4 ± 9.9	< 0.001	265.5 ± 7.9	267.3 ± 9.8	< 0.001	265.4 ± 8.0	267.1 ± 9.7	< 0.001
P onset	139.9 ± 13.1	134.1 ± 13.6	< 0.001	139.9 ± 13.1	134.1 ± 13.6	< 0.001	139.9 ± 13.11	134.2 ± 13.48	< 0.001	139.9 ± 13.1	134.1 ± 13.5	< 0.001
P offset	191.3 ± 12.1	186.7 ± 13.9	< 0.001	191.3 ± 12.1	186.6 ± 13.9	< 0.001	191.3 ± 12.1	186.9 ± 14.0	< 0.001	191.3 ± 12.1	186.8 ± 13.9	< 0.001
T offset	414.6 ± 17.6	426.1 ± 22.8	< 0.001	414.6 ± 17.6	426.2 ± 22.8	< 0.001	414.7 ± 17.5	426.0 ± 22.8	< 0.001	414.6 ± 17.5	426.0 ± 23.0	< 0.001
QT cFrederica	419.6 ± 28.7	448.2 ± 41.5	< 0.001	419.6 ± 28.8	448.4 ± 41.6	< 0.001	419.6 ± 28.6	448.1 ± 41.5	< 0.001	419.5 ± 28.6	447.8 ± 41.5	< 0.001

Values are presented as number (%) or mean ± standard deviation.
PAF, paroxysmal atrial fibrillation.
The patient characteristics are excluded when they are not available.

deep-learning models by measuring the contribution of each input to the output of the model [34]. It functions by estimating and decomposing the layer-level contributions presenting relevance scores for input features as a heatmap. We calculated the relevance score of each input value by backpropagating from the class score of the output node to the input layer, scaling values between 0 and 1 across all leads. This produced a scaled relevance score for each input dimension. To elucidate factors influencing early PAF diagnosis, we interpreted LRP from the following perspectives: leads and intervals. Lead contributions were identified by averaging relevance scores for each lead across classifications. To maintain consistency and facilitate model interpretation, the ECG summary process was referenced and applied to relevance scores. Next, the ECG intervals were confirmed by summarizing key ECG measurements obtaining the average value of the relevance score for each interval.

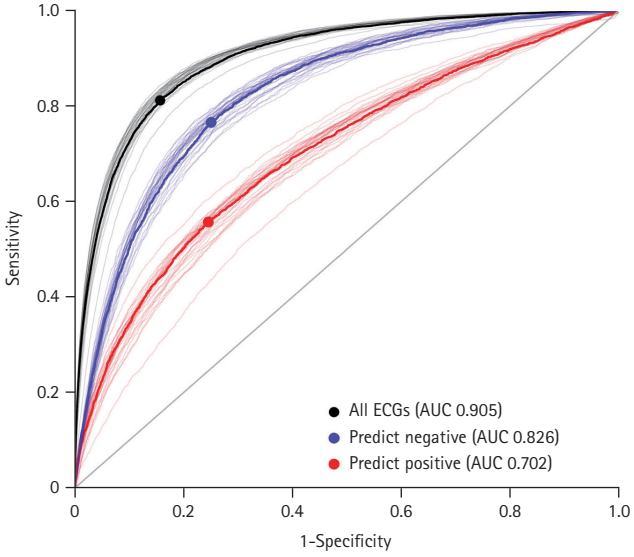


Figure 3. Model performance. ROC curves of the trained model for all data (black), negative prediction ECG subset (blue), and positive prediction ECG subset (red). ECG, electrocardiogram; AUC, area under the curve; ROC, receiver operating characteristic.

Ethics statement

The Institutional Review Board of Chonnam National University Hospital approved this study with a waiver of consent (CNUH-2021-158), adhering to institutional patient privacy policies.

RESULTS

In the training, validation, and test datasets, age and sex distributions were similar between the PAF and normal groups. Table 1 presents the patient characteristics for PAF and normal groups across each dataset. Categorical variables are shown as absolute numbers or percentages, while contin-

Table 2. Average of relevance scores for each ECG lead in the confusion matrix

Lead	True negative	True positive	False negative	False positive
I	0.3410	0.4319	0.4320	0.3765
II	0.3193	0.4228	0.4200	0.3488
III	0.3696	0.4470	0.4429	0.3994
aVR	0.5525	0.5042	0.5083	0.5123
aVL	0.3574	0.4397	0.4378	0.3950
aVF	0.2980	0.4134	0.4097	0.3313
V1	0.4287	0.4710	0.4693	0.4400
V2	0.3758	0.4518	0.4503	0.3979
V3	0.3624	0.4483	0.4453	0.3887
V4	0.3237	0.4262	0.4242	0.3468
V5	0.3332	0.4258	0.4264	0.3529
V6	0.3730	0.4398	0.4438	0.3906
ECG features	0.0000	0.0000	0.0000	0.0000

ECG, electrocardiogram.
Based on the classification results of the test dataset, the average relevance score for each lead was calculated, with scores ranging from 0 to 1; higher values indicate a greater influence on classification results.

https://doi.org/10.3904/kjim.2024.130

uous variables are reported as mean ± standard deviation. Categorical and continuous variables were compared using the chi-square test and Student's t-test, respectively.

Model training involved 331,422 ECGs, with a mean patient age of 55.9 ±19.2 years at the date of the ECG recording. Of these, 129,342 (51.8%) ECGs were from male patients, and 15,887 (4.8%) were labeled as PAF. The internal validation set included 110,475 ECGs, with a mean age of 55.9 ±19.2 years; 43,098 (51.9%) were male, and 5,353 (4.8%) were PAF cases. The test dataset also included 110,475 ECGs, with a mean age of 55.7 ±19.2 years; 42,880 (51.6%) were male, and 5,301 (4.8%) were PAF cases. Cases lacking age (n = 64,587; 11.7%) and sex information (n = 136,243; 24.7%) were treated as noise during model training.

Evaluation of model performance

The limited number of positive cases required for effective training was a critical barrier when applying deep learning to diagnose PAF. To address this class imbalance, we employed focal loss as a loss function of CNN. This technique reduces the influence of easy negative cases while assigning greater weight to hard positive cases, thereby enhancing model accuracy. Here, we implemented a CNN model that inputs ECG data with features such as age, sex, and ECG parameters. Model performance was evaluated using the area under the curve of the receiver operating characteristic curve (AUROC). To understand the generalizability and stability of the model, 30 datasets were generated through pseudo-random sampling and trained individually against the corresponding test dataset. The performance results

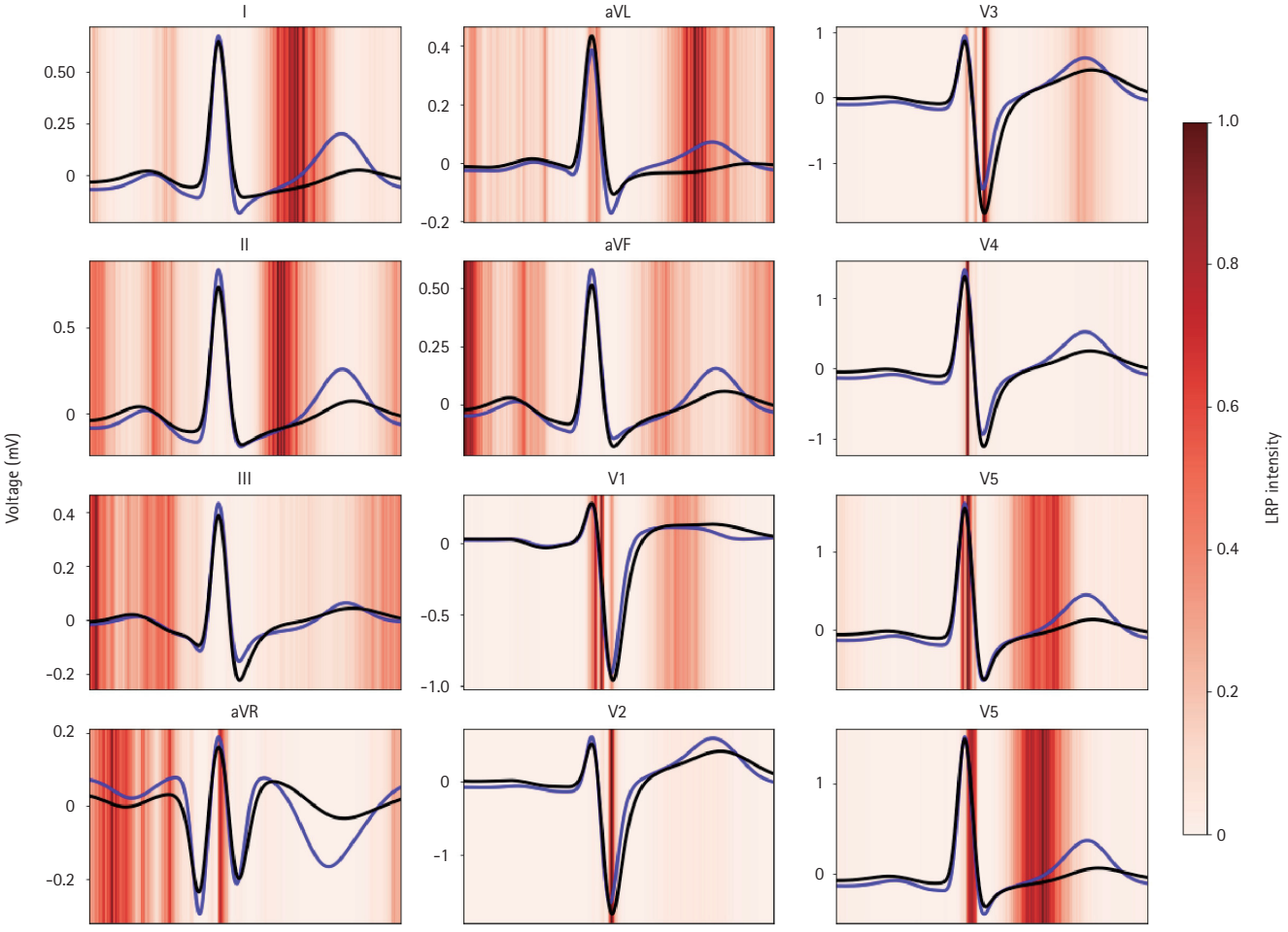


Figure 4. Identification of major contributors to PAF diagnosis through model explainability with the LRP method. Overlay of summarized ECG signals for all true positive (black) and negative (blue) cases. The ECG signals for each lead are aligned and interpolated into one signal. The LRP projection map for true positives is highlighted in red, with the darker red regions indicating a stronger contribution to PAF diagnosis. LRP, layer-wise relevance propagation; ECG, electrocardiogram; PAF, paroxysmal atrial fibrillation.

showed a sensitivity, specificity, and F1 score of 0.722 ± 0.015, 0.906 ± 0.006, and 0.542 ± 0.011, respectively. The AUC for early PAF diagnosis was 0.905 ± 0.007. Class-specific predictive accuracy was 0.826 ± 0.011 for negative cases (Fig. 3, blue) and 0.702 ± 0.019 for the positive cases (Fig. 3, red).

Interpreting the inference process of the deep learning model

Although our model showed strong performance in diagnosing PAF using large-scale data, interpreting the inference process of the deep learning model remains challenging. We used the LRP method to explore the ECG leads and intervals that affected PAF prediction. The P wave is clinically known to affect the diagnosis of PAF [18,21,35,36]. We hypothesized that critical factors could influence prediction more subtly, which could be revealed by analyzing ECG contribution.

To explain the deep learning explanation technique on ECG records, we used the LRP method to identify key ECG leads and intervals. First, the average relevance scores of each ECG lead in the confusion matrix were calculated (Table 2). Regardless of prediction results, the aVR lead had the highest relevance score, indicating it as the most referenced by the model in PAF diagnosis. In contrast, other ECG features showed low relevance scores, suggesting limited predictive effect. To identify the most significant interval for PAF diagnosis and minimize mutual cancellation of ECG signals, we constructed an averaged ECG by aligning, sorting, and interpolating each QRS complex based on the R-peaks. Figure 4 displays this summarized ECG along with LRP intensities for the true positive cases. We summarized ECGs for true positives (Fig. 4, black) and negatives (Fig. 4, blue) for each lead. These summaries reveal a significant difference in the ST segment and T wave regions. In particular, the true positives showed T wave depression, suggesting that an inverted (or flattened) T wave may influence the PAF diagnosis. The T waves, indicating ventricular repolarization [37], are commonly classified as nonspecific ST-T abnormalities (NSTTA) when flattened. To identify the most significant ECG interval in PAF diagnosis, the average LRP weights across intervals were compared (Table 3). The ST segment and QRS complex emerged as the strongest predictors for PAF. We concluded that NSTTA could be a characteristic associated with PAF. Studies show that ST segment depression and T wave inversion are significant markers for cardiovascular disease, with T wave inversion linked to increased mortality risk [38,39]. Additionally, transient ST segment depression has been observed during PAF episodes and linked to underlying coronary artery disease [40]. These findings indicate the association between NSTTA and PAF, which is consistent with our results, underscoring the significance of NSTTA and inverted T waves in predicting early PAF diagnosis, further emphasizing their clinical relevance in arrhythmia detection. However, the relationship between NSTTA and PAF has received limited attention, and our analysis confirms that a flattened T wave influences PAF diagnosis. This suggests that T waves could serve as novel predictors for early PAF detection and that deep learning models can effectively reveal complex mechanisms in PAF diagnosis.

Table 3. Relevance scores for each ECG interval

Interval	Alert Index
PR interval	0.2078
PR segment	0.1909
QRS complex	0.3254
ST segment	0.3973
ST interval	0.3109
QT interval	0.3249

ECG, electrocardiogram.
The average relevance score according to major ECG intervals for true-positive cases, scaled to account for the influence of all leads.

DISCUSSION

In this study, we demonstrate the development and analysis of an explainable deep learning algorithm applied to ECG data for early PAF detection. The model showed strong classification performance, even with relatively few positive cases, reflecting the prevalence of PAF in the general population. Across 30 different, nonoverlapping datasets, the model consistently performed well, with an AUC of 0.905 ± 0.007. These findings suggest the potential for clinical tests to prescreen patients at risk of onset PAF during NSR.

The explainability of deep learning closely relates to the reliability of the model output. Identifying potential patterns of AF is crucial, as many PAF cases are asymptomatic. We aim to uncover patterns of AF onset in ECGs primarily classified as NSR. Direct analysis is challenging due to the variabil-

https://doi.org/10.3904/kjim.2024.130

ity in individual ECG characteristics. Using the classification results of the AI model, the outputs were summarized based on the QRS complex, key contributors to model inference were identified for each major interval, and the ECGs were compared between the control and PAF groups. Our deep learning inference analysis revealed that the model referenced the aVR lead most frequently in PAF diagnosis, with the ST segment exerting the greatest influence among ECG intervals. Additionally, ECG features such as age, sex, and PR interval contributed minimally, suggesting that ECG signals may interact nonlinearly in some cases of early PAF detection, which traditional methods cannot fully explain. These findings suggest that the proposed model predicts potential ventricular dysfunction, indicating structural changes that preidentify the disease before AF onset. We identified a previously undescribed and significant role of NSTTA in PAF diagnosis. However, further research into the role of NSTTA and T wave variations is essential to understand their clinical significance in diagnosing PAF. Additional studies are needed to examine how these ECG features might enhance diagnostic accuracy and improve AI model performance in predicting PAF.

Screening strategies under atypical conditions, such as PAF, face inherent limitations owing to false-positive or low-positive cases. To address this, we trained the model with higher weighting on positive cases. Consequently, the model demonstrated high negative predictive value, supporting the feasibility of a low-cost screening test. We believe the described methods may benefit numerous clinical situations. For instance, the output of the model could serve as an alert index. Primary cardiologists can leverage early detection tools to proactively assess the safety of surgical procedures or pacing modalities. This study has some limitations, including its single-center design, which requires validation across diverse healthcare systems. Data imbalance may influence specificity and AUROC, and the absence of personalized information extraction hinders individualized analysis. Additionally, ECG summarization based on true positive and negative cases may reduce detail on specific ECG intervals. In conclusion, an AI model based on standard 12-lead ECG data can predict future AF onset in NSR, with model inference rationale analyzed. Through external validation in more varied cohorts, our model can enhance PAF screening strategies and serve as a proactive clinical tool.

KEY MESSAGE

1. Developed a deep learning model to diagnose AF onset in potential patients exhibiting NSR.
2. Employed a deep learning explanation method to verify the decision basis of the model and investigated ECG patterns influencing the results.
3. Indicating a potential association between NSTTA and PAF, highlighting the need for future validation across various systems.

REFERENCES

1. Sanna T, Diener HC, Passman RS, et al. Cryptogenic stroke and underlying atrial fibrillation. N Engl J Med 2014;370:2478-2486.
2. Healey JS, Connolly SJ, Gold MR, et al. Subclinical atrial fibrillation and the risk of stroke. N Engl J Med 2012;366:120-129.
3. Gladstone DJ, Spring M, Dorian P, et al. Atrial fibrillation in patients with cryptogenic stroke. N Engl J Med 2014;370:2467-2477.
4. Seet RC, Friedman PA, Rabinstein AA. Prolonged rhythm monitoring for the detection of occult paroxysmal atrial fibrillation in ischemic stroke of unknown cause. Circulation 2011;124:477-486.
5. Ziegler PD, Glotzer TV, Daoud EG, et al. Detection of previously undiagnosed atrial fibrillation in patients with stroke risk factors and usefulness of continuous monitoring in primary stroke prevention. Am J Cardiol 2012;110:1309-1314.
6. Pourbabaee B, Roshtkhari MJ, Khorasani K. Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients. IEEE Trans Syst Man Cybern Syst 2018;48:2095-2104.
7. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet 2019;394:861-867.
8. Baek YS, Lee SC, Choi W, Kim DH. A new deep learning algorithm of 12-lead electrocardiogram for identifying atrial fibrillation during sinus rhythm. Sci Rep 2021;11:12818.
9. Diederichsen SZ, Haugan KJ, Kronborg C, et al. Comprehensive evaluation of rhythm monitoring strategies in screening for

atrial fibrillation: insights from patients at risk monitored long term with an implantable loop recorder. Circulation 2020;141:1510-1522.
10. Steinberg BA, Piccini JP. Anticoagulation in atrial fibrillation. BMJ 2014;348:g2116.
11. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med 2009;361:1139-1151.
12. Meschia JF, Bushnell C, Boden-Albala B, et al. Guidelines for the primary prevention of stroke: a statement for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2014;45:3754-3832.
13. Hart RG, Pearce LA, Aguilar MI. Meta-analysis: antithrombotic therapy to prevent stroke in patients who have nonvalvular atrial fibrillation. Ann Intern Med 2007;146:857-867.
14. Granger CB, Alexander JH, McMurray JJ, et al. Apixaban versus warfarin in patients with atrial fibrillation. N Engl J Med 2011;365:981-992.
15. Lanza GA. The electrocardiogram as a prognostic tool for predicting major cardiac events. Prog Cardiovasc Dis 2007;50:87-111.
16. Svennberg E, Engdahl J, Al-Khalili F, Friberg L, Frykman V, Rosenqvist M. Mass screening for untreated atrial fibrillation: the STROKESTOP study. Circulation 2015;131:2176-2184.
17. Ponikowski P, Voors AA, Anker SD, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016;37:2129-2200.
18. Reiffel JA, Verma A, Kowey PR, et al. Incidence of previously undiagnosed atrial fibrillation using insertable cardiac monitors in a high-risk population: the REVEAL AF study. JAMA Cardiol 2017;2:1120-1127.
19. Steinberg JS, Zelenkofske S, Wong SC, Gelernt M, Sciacca R, Menchavez E. Value of the P-wave signal-averaged ECG for predicting atrial fibrillation after cardiac surgery. Circulation 1993;88:2618-2622.
20. Dilaveris PE, Gialafos JE. P-wave dispersion: a novel predictor of paroxysmal atrial fibrillation. Ann Noninvasive Electrocardiol 2001;6:159-165.
21. Aytemir K, Ozer N, Atalar E, et al. P wave dispersion on 12-lead electrocardiography in patients with paroxysmal atrial fibrillation. Pacing Clin Electrophysiol 2000;23:1109-1112.
22. Hannun AY, Rajpurkar P, Haghpanahi M, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019;25:65-69.
23. Ribeiro AH, Ribeiro MH, Paixão GMM, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nat Commun 2020;11:1760.
24. Norberg J, Bäckström S, Jansson JH, Johansson L. Estimating the prevalence of atrial fibrillation in a general population using validated electronic health data. Clin Epidemiol 2013;5:475-481.
25. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-444.
26. Kiranyaz, S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman DJ. 1D convolutional neural networks and applications: a survey. Mech Syst Signal Process 2021;151:107398.
27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 26-30; Las Vegas, NV. IEEE, 2016: 770-778.
28. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on International Conference on Machine Learning; 2015 Jul 6-11; Lille, France. ICML, 2015: 448-456.
29. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th International Conference on International Conference on Machine Learning; 2010 Jun 21-24; Haifa, Israel. ICML, 2010: 807-814.
30. He K, Zhang X, Ren S, Sun, J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV); 2015 Dec 7-13; Santiago, Chile. IEEE, 2015: 1026-1034.
31. Kingma DP, Ba J. Adam: a method for stochastic optimization. Proceedings of the International Conference on Learning Representations 2015; 2015 May 7-9; San diego, CA. ICLR, 2015: 1-15.
32. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22-29; Venice, Italy. ICCV, 2017: 2999-3007.
33. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019;25:30-36.
34. Bach S, Binder A, Montavon G, Klauschen F, Müller KR, Samek W. On pixel-wise explanations for non-linear classifier deci-

https://doi.org/10.3904/kjim.2024.130

sions by layer-wise relevance propagation. PLoS One 2015;10:e0130140.
35. Zhang D, Yang S, Yuan X, Zhang P. Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram. iScience 2021;24:102373.
36. Moraes DN, Nascimento BR, Beaton AZ, et al. Value of the electrocardiographic (P wave, T wave, QRS) axis as a predictor of mortality in 14 years in a population with a high prevalence of chagas disease from the Bambuí Cohort Study of Aging. Am J Cardiol 2018;121:364-369.
37. Burgess MJ. Relation of ventricular repolarization to electrocardiographic T wave-form and arrhythmia vulnerability. Am J Physiol 1979;236:H391-H402.
38. Hanna EB, Glancy DL. ST-segment depression and T-wave inversion: classification, differential diagnosis, and caveats. Cleve Clin J Med 2011;78:404-414.
39. Kurl S, Mäkikallio TH, Laukkanen JA. T-wave inversion and mortality risk. Ann Med 2015;47:69-73.
40. Androulakis A, Aznaouridis KA, Aggeli CJ, R et al. Transient ST-segment depression during paroxysms of atrial fibrillation in otherwise normal individuals: relation with underlying coronary artery disease. J Am Coll Cardiol 2007;50:1909-1911.

Received : April 16, 2024
Revised : October 14, 2024
Accepted : October 28, 2024

Correspondence to
Ki Hong Lee, M.D., Ph.D.
Department of Internal Medicine, Chonnam National University Hospital, 42 Jebong-ro, Dong-gu, Gwangju 61469, Korea
Tel: +82-62-220-6256, Fax: +82-62-225-8578
E-mail: drgood2@naver.com
https://orcid.org/0000-0002-9938-3464

Acknowledgments
Special thanks to Dr. Seong Won Jeon, Dr. Changhyun Kim, and Dr. Dong Kyun Kim for the diagnosis and labeling of ECGs.

CRedit authorship contributions
Yeongbong Jin: conceptualization, investigation, writing - original draft, visualization; Bonggyun Ko: conceptualization, investigation, formal analysis, writing - original draft, supervision; Woojin Chang: formal analysis, supervision; Kang-Ho Choi: investigation, formal analysis, validation; Ki Hong Lee: investigation, writing - review & editing, supervision

Conflicts of interest
The authors disclose no conflicts.

Funding
This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1G1A1100704, No. 2021R1F1A1060049 and No. 2022M3A9E4017151), a grant (BCRI23054) of Chonnam National University Hospital Biomedical Research Institute, a grant of Establishment of K-Health National Medical Care Service and Industrial Ecosystem funded by the Ministry of Science and ICT (MSIT, Korea)(No. ITAH060323011001000100010100) and also by the BK21 FOUR (Fostering Outstanding Universities for Research, NO.5120200913674) funded by the Ministry of Education (MOE, Korea) and National Research Foundation of Korea (NRF).