Korean Circ J. 2023 Oct;53(10):677-689
https://doi.org/10.4070/kcj.2023.0012
pISSN 1738-5520·eISSN 1738-5555

kcj
Korean Circulation Journal

Original Research

Check for updates

Machine Learning Prediction for the Recurrence After Electrical Cardioversion of Patients With Persistent Atrial Fibrillation

Soonil Kwon, MD[1,*], Eunjung Lee, PhD[2,*,†], Hojin Ju, MD[1], Hyo-Jeong Ahn, MD[1], So-Ryoung Lee, MD, PhD[1], Eue-Keun Choi, MD, PhD[1,3,‡], Jangwon Suh, BA[4], Seil Oh, MD, PhD[1,3], and Wonjong Rhee, PhD, FIEEE[4,‡]

[1]Department of Internal Medicine, Seoul National University Hospital, Seoul, Korea
[2]Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, USA
[3]Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea
[4]Department of Intelligence and Information, Seoul National University, Seoul, Korea

OPEN ACCESS

Received: Jan 9, 2023
Revised: May 25, 2023
Accepted: Jun 13, 2023
Published online: Jul 24, 2023

Correspondence to
Eue-Keun Choi, MD, PhD
Department of Internal Medicine, Seoul National University College of Medicine and Seoul National University Hospital, 101, Daehak-ro, Jongno-gu, Seoul 03080, Korea.
Email: choiek17@snu.ac.kr

Wonjong Rhee, PhD, FIEEE
Department of Intelligence and Information, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Korea.
Email: wonjong@gmail.com

*Soonil Kwon and Eunjung Lee contributed equally.

†This work was performed while the author was with Seoul National University.

‡Eue-Keun Choi and Wonjong Rhee are co-corresponding authors.

ORCID iDs
Soonil Kwon
https://orcid.org/0000-0002-4791-6855

AUTHOR'S SUMMARY

There is limited evidence regarding machine-learning prediction for recurrence after electrical cardioversion (ECV) among patients with persistent atrial fibrillation (AF). We aimed to predict the recurrence after ECV in persistent AF with a machine-learning approach. Machine-learning models were evaluated with various training datasets. Compared to the training of clinical features alone, that of electrocardiogram (ECG) features or both features showed significantly higher areas under the receiver operating characteristic curves (0.57 vs. 0.60 and 0.63, respectively; both p<0.001). Machine learning of both clinical features and ECG showed a synergistic impact in predicting AF recurrence after ECV in persistent AF patients.

ABSTRACT

Background and Objectives: There is limited evidence regarding machine-learning prediction for the recurrence of atrial fibrillation (AF) after electrical cardioversion (ECV). This study aimed to predict the recurrence of AF after ECV using machine learning of clinical features and electrocardiograms (ECGs) in persistent AF patients.
Methods: We analyzed patients who underwent successful ECV for persistent AF. Machine learning was designed to predict patients with 1-month recurrence. Individual 12-lead ECGs were collected before and after ECV. Various clinical features were collected and trained the extreme gradient boost (XGBoost)-based model. Ten-fold cross-validation was used to evaluate the performance of the model. The performance was compared to the C-statistics of the selected clinical features.
Results: Among 718 patients (mean age 63.5±9.3 years, men 78.8%), AF recurred in 435 (60.6%) patients after 1 month. With the XGBoost-based model, the areas under the receiver operating characteristic curves (AUROCs) were 0.57, 0.60, and 0.63 if the model was trained by clinical features, ECGs, and both (the final model), respectively. For the final model, the sensitivity, specificity, and F1-score were 84.7%, 28.2%, and 0.73, respectively. Although the AF duration showed the best predictive performance (AUROC, 0.58) among the clinical

Eunjung Lee
https://orcid.org/0000-0002-6476-9527
Hojin Ju
https://orcid.org/0000-0002-9072-7611
Hyo-Jeong Ahn
https://orcid.org/0000-0003-3590-0378
So-Ryoung Lee
https://orcid.org/0000-0002-6351-5015
Eue-Keun Choi
https://orcid.org/0000-0002-0411-6372
Jangwon Suh
https://orcid.org/0000-0002-6293-0444
Seil Oh
https://orcid.org/0000-0002-2504-9615
Wonjong Rhee
https://orcid.org/0000-0002-2590-8774

Funding

This research was supported by the Korean Cardiac Research Foundation (No. 201901-01), the SNUH Research Fund (No. 0320202040), and the Korea Medical Device Development Fund by the Korean government (Ministry of Science and Information and Communications Technology, Ministry of Trade, Industry and Energy, Ministry of Health and Welfare, Republic of Korea, Ministry of Food and Drug Safety; project number 202013B14).

Conflict of Interest

Eue-Keun Choi received research grants from Bayer, BMS/Pfizer, Biosense Webster, Chong Kun Dang, Daiichi-Sankyo, Samjinpharm, Sanofi-Aventis, Seers Technology, Skylabs, and Yuhan. No personal fees were received. Other authors have no financial conflicts of interest.

Data Sharing Statement

The data generated in this study is available from the corresponding author upon reasonable request.

Author Contributions

Conceptualization: Kwon S, Lee E, Ju H, Choi EK; Data curation: Kwon S, Lee E, Ju H, Suh J; Formal analysis: Kwon S, Lee E, Ju H; Funding acquisition: Choi EK; Investigation: Kwon S, Lee E, Ju H, Ahn HJ, Lee SR, Choi EK, Suh J; Methodology: Kwon S, Lee E, Ju H, Ahn HJ, Lee SR, Choi EK, Suh J; Project administration: Choi EK, Oh S, Rhee W; Supervision: Choi EK, Oh S, Rhee W; Validation: Kwon S, Lee E, Ju H, Ahn HJ, Lee SR, Suh J; Visualization: Kwon S, Lee E; Writing - original draft: Kwon S, Lee E; Writing - review & editing: Choi EK, Oh S, Rhee W.

features, it was significantly lower than that of the final machine-learning model (p<0.001). Additional training of extended monitoring data of 15-minute single-lead ECG and photoplethysmography in available patients (n=261) did not significantly improve the model's performance.

Conclusions: Machine learning showed modest performance in predicting AF recurrence after ECV in persistent AF patients, warranting further validation studies.

Keywords: Atrial fibrillation; Electric countershock; Machine learning; Recurrence

INTRODUCTION

Electrical cardioversion (ECV) is a common strategy for restoring sinus rhythm in atrial fibrillation (AF). It is useful for treating highly symptomatic or hemodynamically unstable patients with AF. Although ECV has a high success rate (up to 90%) in restoring sinus rhythm from AF,[1] a frequent recurrence rate has been regarded as one of its main limitations. The 1-year recurrence rate of AF after ECV may range from 40% to 50% despite antiarrhythmic drug use.[2-4] Additionally, especially for patients with persistent AF, ECV becomes less effective because AF triggers induce immediate recurrence of atrial fibrillation (IRAF).[5] Therefore, predicting ineffective ECV among AF patients is crucial.

Accordingly, various predictors, including longer AF durations, older age,[6] electrocardiogram (ECG)[7] or echocardiography features,[8] and serological biomarkers,[9][10] have been suggested as potential predictors of AF recurrence after ECV. However, every predictor needs more predictive performance; thus, their clinical applicability remains limited.[11] Recently, machine learning has been widely studied in cardiovascular medicine.[12][13] It has an advantage in solving difficult classifications or predictions using nonlinear computations of given features, which remains limited in traditional statistical methods. Therefore, a machine-learning approach might be necessary to improve predictive performance by analyzing multiple predictors.

In this study, we aimed to investigate the feasibility of a machine-learning approach to predict AF recurrence after ECV by using various clinical features and ECGs among patients with persistent AF.

METHODS

Study population and study outcome

This study used data from a registry of patients with AF who underwent ECV for persistent AF between 2010 to 2021. The enrollment flow is illustrated in **Figure 1**. From the registry, we excluded the following: 1) patients who failed to restore sinus rhythm on ECG or had IRAF; 2) those who underwent ECV for conditions other than AF, such as atrial flutter or ventricular arrhythmia; 3) those who had missing data for any of the 12-lead ECGs before or after ECV; and 4) those who were lost-to-follow-up at 1 month. Finally, a total of 718 patients was included in this study. The study outcome was defined as AF recurrence at 1 month following a successful electrical cardioversion. AF recurrence was determined based on documented episodes of AF from a 12-lead ECG or ambulatory Holter test conducted at 30±7 days after the successful electrical cardioversion.

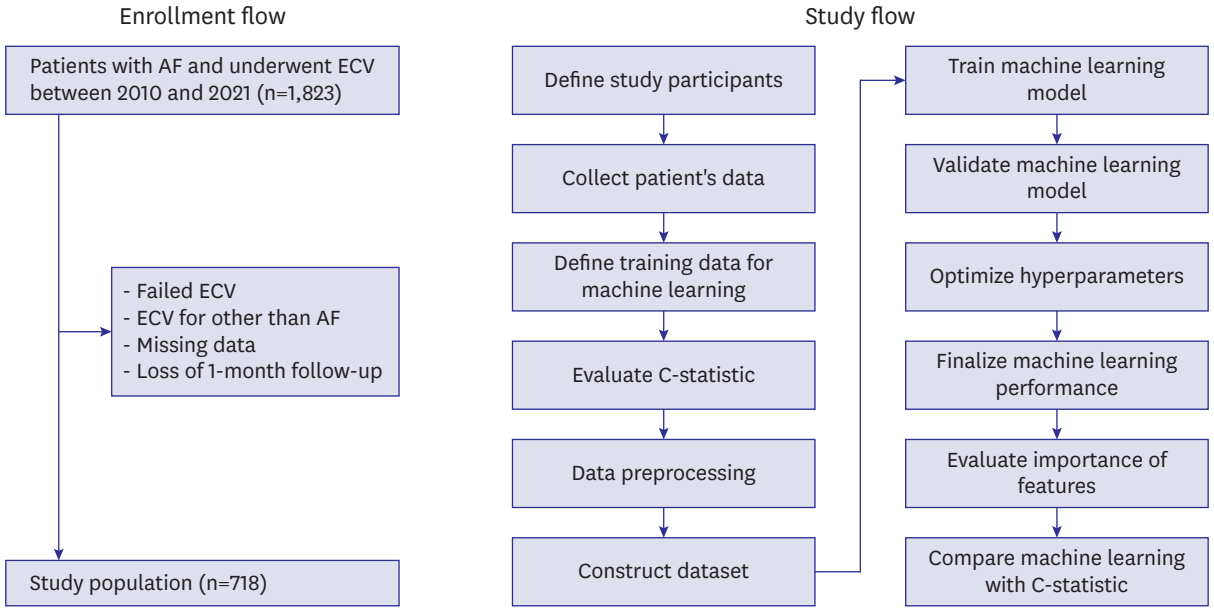

Figure 1. The enrollment and study flow.
AF = atrial fibrillation; ECV = electrical cardioversion.

Data acquisition and dataset construction

To predict 1-month recurrence after ECV, we collected all available clinical features, including demographic information, blood tests, ECG features, echocardiographic features, comorbidities, concomitant medications, and detailed information on AF history; a complete list of features is presented in **Supplementary Table 1**. Among the initially collected features, those with missing values in >30% of the study population were excluded. We then performed a univariate logistic regression analysis for each feature to predict a 1-month recurrence and excluded those with a p value >0.90 from machine learning; they were regarded as irrelevant in predicting the recurrence. Finally, eight clinical features, including age, sex, body mass index, AF duration, CHA_2DS_2-VASc score, left ventricular ejection fraction, left atrial diameter, and types of antiarrhythmic drugs, were included in machine learning. We also collected 12-lead ECGs before and after ECV, and their ECG features were used for machine learning. Because we included only patients who underwent successful ECV, all 12-lead ECGs before ECV showed AF, whereas those after ECV showed sinus rhythm. Among the numerous available ECGs, we chose the closest possible ECGs from the date of the ECV. Consequently, we constructed two datasets i.e., clinical features and 12-lead ECGs. The data construction flow is illustrated in **Figure 2**.

Preprocessing

For clinical features, continuous variables were normalized using each maximum value, and empty values were substituted with –1. For 12-lead ECG features, we used the ventricular rate, QT interval, QRS duration, and R-axis from the raw ECG data. The R peak mean value of each lead (except aVR and aVL) and root mean square of the successive differences (RMSSD) in the RR interval was calculated using preprocessed ECG signals. To obtain these signals, they were filtered using a finite impulse response filter and a band-pass filter (22–333 Hz). The location of the R peaks was then determined using the BioSPPy Python library. For 15-minute ECG and photoplethysmography (PPG) signals, we employed a band-pass filter (0.2–8 Hz) for preprocessing. Features from the 12-lead ECG, 15-min single-lead ECG, and 15-minute PPG were also normalized by their maximum values.

https://doi.org/10.4070/kcj.2023.0012

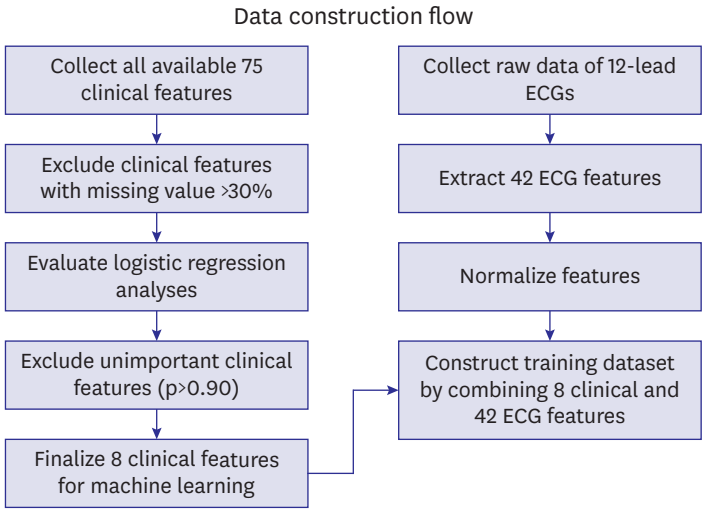

Figure 2. The flow of data construction for machine learning.
Eight clinical features and 42 ECG features were used for machine learning.
ECG = electrocardiogram.

Machine-learning framework

In this study, logistic regression, gradient boosting, and random forest were implemented. We set logistic regression with L2 penalty and the number of maximum iterations to be 1,000 using the Scikit-learn library from Python. The Scikit-learn library was also used for the random forest method, with 50 trees and the maximum depth of the tree as 3. For gradient boosting, we trained gradient boosting models from the extreme gradient boost (XGBoost) Python library with 50 boosting iterations, a learning rate of 0.05, and the learning task as logistic regression for binary classification. The area under the precision-recall curve (AUPRC) was used for evaluation during gradient-boosting training. We trained all the models using the default parameter settings of each Python library, except for what we described.

Evaluation of machine-learning performance

To evaluate the machine-learning performance, we performed a 10-fold cross-validation. In addition, each split for cross-validation was divided by considering the ratio of labels. After cross-validation, we gathered the results of each test split and measured the sensitivity, specificity, area under the receiver operating characteristic curve (AUROC), AUPRC, and F1-score. We then repeated the cross-validation with 5 different seed numbers and reported their average performance. Finally, the machine-learning performances with three different training datasets i.e., clinical features, 12-lead ECGs, and both, were compared.

Feature of importance

In addition to measuring the overall performance, we investigated the feature importance of the three machine-learning methods. Each method used different metrics because of the nature of machine-learning methods. The coefficients of the features were used to determine the importance of the features in the logistic regression. The accuracy improvement when each feature was added to an existing branch to create a new split was used to calculate the feature importance of the gradient-boosting method. The mean decrease in impurity, also known as the *gini importance*,[14] was computed for the feature importance of the random forest method.

https://doi.org/10.4070/kcj.2023.0012

Exploratory analysis: extended monitoring of single-lead electrocardiogram and photoplethysmography

In selected patients, we measured single-lead ECG (lead II) and PPG data simultaneously over 15 minutes before and after ECV. PPG signals were recorded using CART (Sky Labs, Inc., Seongnam, Korea), which is a ring-type wearable device that can monitor PPG signals from a user's finger and has a performance comparable to that of a medical-grade pulse oximeter.[15] After preprocessing, the 15-minute ECG and PPG data were constructed as another training dataset. We investigated whether there was a performance improvement by adding 15-minute ECG and PPG datasets to the 12-lead ECG dataset.

Statistical analyses

The data were presented as either number (%), mean ± standard deviation, or mean (95% confidence interval; CI) according to the data type. The diagnostic performance of machine learning was evaluated using sensitivity, specificity, AUROC, AUPRC, and F1-score. The AUROC and AUPRC were calculated with 95% CIs. For clinical features, C-statistics were performed to predict the 1-month recurrence. The best AUROC with C-statistics was then compared with the machine-learning performance. The AUROCs of the different training datasets were compared using the DeLong test. In all statistical tests, p values <0.05 were used to reject the null hypothesis. The data were analyzed using SPSS version 22.0 (IBM Corp., Armonk, NY, USA).

RESULTS

A total of 718 patients who underwent successful ECV for persistent AF were analyzed (**Table 1**). The population's mean age was 63.5±9.3 years; the male proportion was 78.8%; and the mean CHA_2DS_2-VASc score was 1.9±1.4. The most used antiarrhythmic drug was amiodarone (53.8%). The mean left atrial size was 49.0±6.6 mm.

Predictive performances of clinical features using the C-statistics

Supplementary Figure 1 illustrates the predictive performances of selected clinical features using C-statistics. Among the clinical features, AF duration showed the highest AUROC (0.58 [95% CI, 0.54–0.62]). Left atrial size, body mass index, and age did not show significant predictive performance (AUROC, 0.53 [95% CI, 0.48–0.57]; 0.51 [0.47–0.55]; and 0.50 [0.46–0.54], respectively).

Predictive performance of machine-learning models

Among the three machine-learning models (XGBoost, logistic regression, and random forest), the XGBoost model improved both AUROC and AUPRC when clinical features and ECGs were trained together (AUROC, 0.57, 0.60, and 0.63; AUPRC, 0.66, 0.68, and 0.71 for training clinical features, ECGs, and both, respectively) (**Table 2**). For the logistic regression model, no synergistic effect was observed in AUPRC when both datasets were used together (AUPRC, 0.62, 0.61, and 0.62 for training clinical features, ECGs, and both datasets, respectively). For the random forest model, there was no improvement in both AUROC and AUPRC by training both datasets (AUROC and AUPRC were 0.55 and 0.63, respectively).

Figure 3 illustrates the receiver operating characteristic curves of the XGBoost model based on the trained datasets. Compared to the training of clinical features, both the training of 12-lead ECGs and all datasets showed significantly higher AUROCs (0.60 [95% CI, 0.58–0.62]

https://doi.org/10.4070/kcj.2023.0012

kcj Korean Circulation Journal

Table 1. Baseline characteristics of the study population (n =718)

Characteristics	Value
Demographic factors	
Age (years)	63.5±9.3
Male (%)	566 (78.8)
Body weight (kg)	70.6±14.1
Body mass index (kg/m^2)	25.0±3.7
CHA_2DS_2-VASc scores	1.9±1.4
AF duration (months)	24 (6–66)
Types of atrial fibrillation (%)	
Persistent	718 (100.0)
Comorbidities (%)	
Hypertension	401 (55.8)
Diabetes mellitus	147 (20.5)
Heart failure	98 (13.6)
Ischemic heart disease	49 (6.8)
Stroke	53 (7.4)
Peripheral artery disease	0 (0)
Valvular heart disease	26 (3.6)
Dyslipidemia	204 (28.4)
Chronic kidney disease	45 (6.3)
Chronic obstructive pulmonary disease	0 (0)
Types and doses of antiarrhythmic drugs (%)*	
None	50 (7.0)
Flecainide 25 mg bid or 50 mg qd	5 (0.7)
Flecainide 50 mg bid	45 (6.3)
Flecainide 75 mg bid or higher doses	10 (1.4)
Propafenone 225 mg qd	10 (1.4)
Propafenone 225 mg bid	190 (26.5)
Propafenone 325 mg bid	8 (1.1)
Pilsicainide 25 mg qd	3 (0.4)
Pilsicainide 50 mg bid or higher doses	8 (1.1)
Amiodarone 100 mg qd	77 (10.7)
Amiodarone 200 mg qd	300 (41.8)
Amiodarone 200 mg bid	9 (1.3)
Sotalol 40 mg bid	2 (0.3)
Dronedarone 400 mg qd	1 (0.1)
Cardiovascular medications (%)	
β blockers	242 (33.7)
Calcium channel blockers	132 (18.4)
Renin-angiotensin-aldosterone system blockades	181 (25.2)
Statins	238 (33.1)
Echocardiographic features	
Left ventricular ejection fraction (%)	57.1±7.8
Left ventricular end-diastolic diameter (mm)	48.9±4.5
Left ventricular end-systolic diameter (mm)	31.2±4.9
Interventricular septal thickness (mm)	9.4±1.4
Posterior wall thickness (mm)	9.5±2.2
Aorta diameter (mm)	34.8±5.6
Left atrial size (mm)	49.0±6.6
Left atrial volume (cc)	110.2±36.9
Pulmonary arterial systolic pressure (mmHg)	30.9±5.8
Left ventricular mass (g)	164±45

Data are number (%), mean ± standard deviation, or median (interquartile range).
AF = atrial fibrillation; bid = twice a day; qd = once a day.
*The sum of proportions may not add up to 100% due to rounding.

and 0.63 [0.61–0.64], respectively; both p<0.001). The XGBoost model training all datasets achieved a significantly higher AUROC than the C-statistic of the AF duration (AUROC, 0.63 [95% CI, 0.61–0.64] vs. 0.58 [0.54–0.62], p<0.001).

https://doi.org/10.4070/kcj.2023.0012

kcj
Korean Circulation Journal

Table 2. Predictive performances of machine learning models according to algorithms and trained datasets

Algorithms	Datasets	AUROC	AUPRC	F1-score	Sensitivity	Specificity
XGBoost	Clinical features	0.57	0.66	0.73	89.2	17.9
	ECGs	0.60	0.68	0.72	82.2	29.0
	All datasets	0.63	0.71	0.73	84.7	28.2
Logistic regression	Clinical features	0.53	0.62	0.72	85.8	19.9
	ECGs	0.51	0.61	0.72	87.3	13.8
	All datasets	0.54	0.62	0.70	79.1	28.4
Random forest	Clinical features	0.55	0.63	0.74	90.0	20.6
	ECGs	0.55	0.63	0.74	88.7	20.7
	All datasets	0.55	0.63	0.74	90.2	20.1

The machine learning model used the XGBoost algorithm. Data for sensitivity and specificity are shown as %.
AUPRC = area under the precision-recall curve; AUROC = area under the receiver operating characteristic curve; ECG = electrocardiogram; XGBoost = extreme gradient boost.

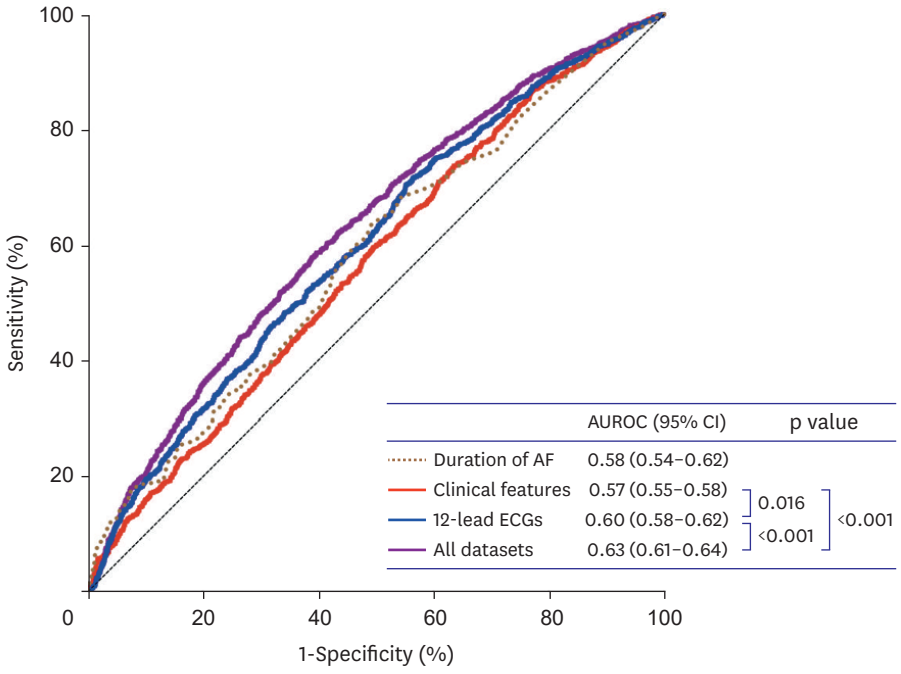

Figure 3. Predictive performance of the machine learning model for 1-month AF recurrence after electrical cardioversion according to trained datasets.
The XGBoost model training both ECGs and clinical features improved the predictive performance compared with training either ECGs or clinical features alone. The dotted line represents the best AUROC based on the C-statistics of the AF durations.
AF = atrial fibrillation; AUROC = area under the receiver operating characteristic curve; ECG = electrocardiogram; XGBoost = extreme gradient boost.

Impact of features on the machine-learning model

The features of importance evaluated using the XGBoost model are summarized in **Figure 4**. Among the clinical features, the AF duration was the most important feature for the machine-learning model to predict the 1-month recurrence of AF, followed by the type of concomitant antiarrhythmic drug, CHA_2DS_2-VASc score, and age. Among the ECGs, features from AF ECG are regarded as more important than those from sinus rhythm ECG. Atrial rate, R peak, and QRS duration are some of the most important ECG features. If we included all datasets, the ECG features were at a higher priority than the clinical features in predicting AF recurrence.

Impact of training extended monitoring datasets on predictive performance

Among the study population, 261 patients successfully recorded extended monitoring

https://doi.org/10.4070/kcj.2023.0012

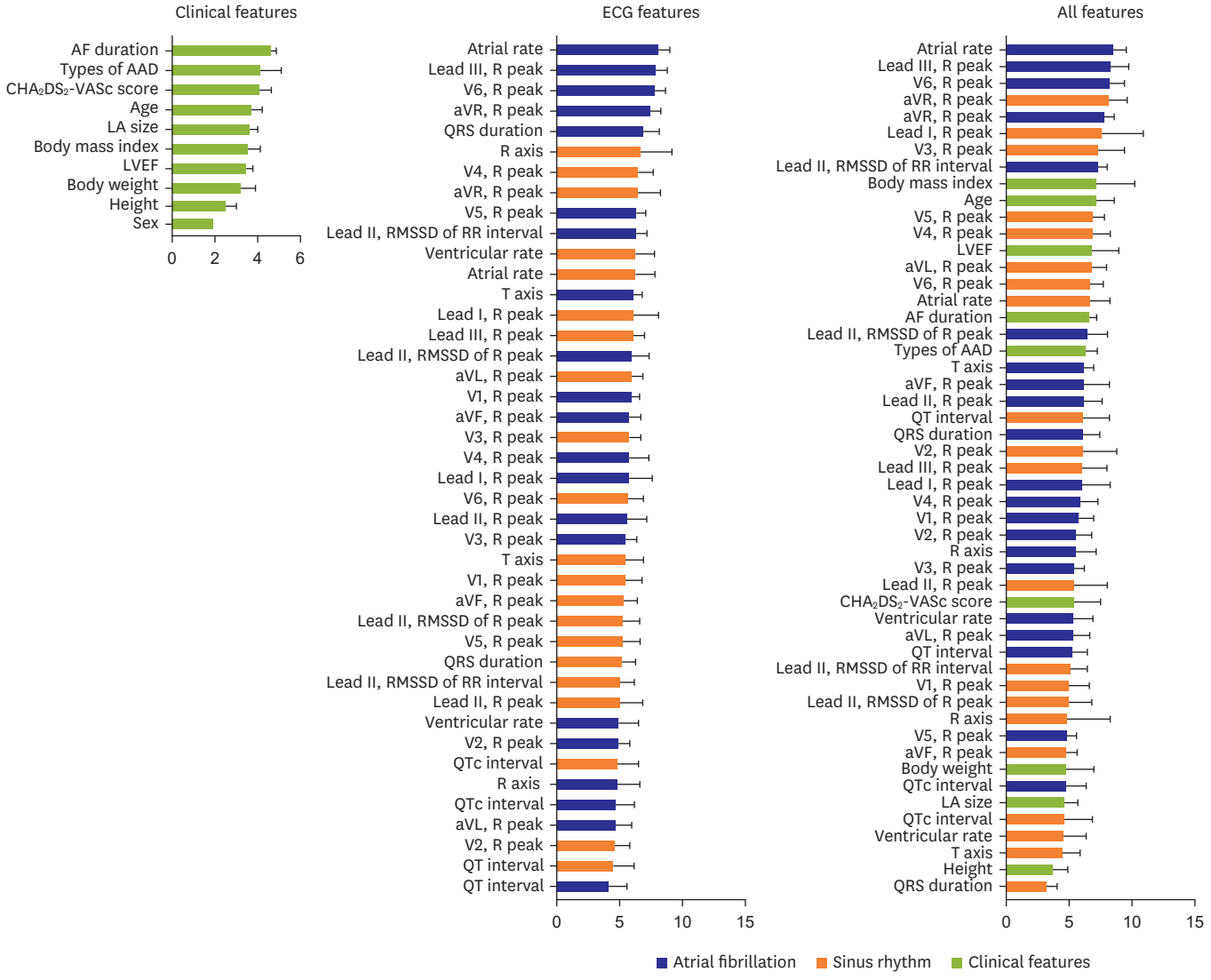

Figure 4. The visualization of the feature of importance for the machine learning model.
The AF duration was the most important feature among the clinical parameters. Among all the features, AF ECG features were of high priority for the machine learning model to predict 1-month recurrence after ECV. The x-axis represents the frequency of a feature being utilized to split the data across all trees in the machine learning model. A higher value on the x-axis indicates greater importance of the feature in classifying patients with and without 1-month recurrence of AF.
AAD = antiarrhythmic drug; AF = atrial fibrillation; ECG = electrocardiogram; ECV = electrical cardioversion; LA = left atrial; LVEF = left ventricular ejection fraction.

i.e., simultaneous single-lead ECG and PPG over 15 minutes both before and after ECV. The XGBoost model showed AUROC of 0.58 (95% CI, 0.55–0.61) if it trained the clinical features and the extended monitoring dataset. Even though the extended monitoring dataset was added to the 12-lead ECG dataset and the clinical features, there was no significant improvement in AUROC (0.57 [95% CI, 0.54–0.60], p=0.618) (**Supplementary Figure 2**). In addition, the other diagnostic parameters showed no significant improvement (**Table 3**).

DISCUSSION

In this study, we investigated a machine-learning approach to predict 1-month AF recurrence after ECV in patients with persistent AF. Among the algorithms, XGBoost exhibited the

https://doi.org/10.4070/kcj.2023.0012

Table 3. Predictive performances of the machine learning model after additional training of extended monitoring datasets in selected patients (n=261)

Datasets	AUROC	AUPRC	F1-score	Sensitivity	Specificity
Clinical features and extended monitoring dataset	0.58	0.71	0.77	89.1	18.4
12-lead ECGs and extended monitoring dataset	0.58	0.72	0.77	87.9	18.9
All datasets	0.57	0.71	0.77	88.4	18.9

The machine learning model used the XGBoost algorithm. The extended monitoring dataset refers to the simultaneous 15-minute single-lead ECG (lead II) and PPG datasets recorded before and after the ECV. Data for sensitivity and specificity are shown as %.
AUPRC = area under the precision-recall curve; AUROC = area under the receiver operating characteristic curve; ECG = electrocardiogram; ECV = electrical cardioversion; PPG = photoplethysmography; XGBoost = extreme gradient boost.

best predictive performance. Training both clinical and ECG features improved the model's performance as compared to training each dataset alone, thereby suggesting a synergistic impact in improving the predictability of AF recurrence by learning multiple types of information on a given patient. Although the machine-learning model only achieved modest performance (AUROC, 0.63), it outperformed the C-statistics of the selected clinical features. Compared with previous reports, the main difference in our study is that we used not only clinical features but also ECGs and extended monitoring data for machine learning.

ECV is an effective and safe method for converting sinus rhythm in patients with AF. Although it risks periprocedural thromboembolic events, the event rate could be lowered to approximately 0.28–0.8%, if treated appropriately with anticoagulants.[16)17)] When ECV is considered as an elective procedure for AF management, the prediction of successful and effective ECV is important. If patients with AF fail ECV or suffer from IRAF, choosing ineffective patients may save costs, time, and medical supplies for both the patients and hospitals.

Because AF affects electrical, structural, and mechanical remodeling of the atria,[11)] various characteristics of the heart e.g., ECG, echocardiography, and serologic biomarkers, may reflect the severity of AF and thus, can be used to predict recurrence after ECV. For ECG, an atrial substrate, a known risk factor of AF recurrence,[18)] is not only related to atrial fibrosis and mechanical dysfunction, but also affects P waves resulting in low amplitudes and dispersed widths. Therefore, P-wave signal-averaged ECG may be used to predict AF recurrence after ECV.[7)] Besides laboratory test results and clinical features associated with AF progression e.g., longer AF duration, older age, or higher body mass index, are well-known predictors of AF recurrence because they affect atrial remodeling.[11)] However, the clinical utility of such predictors remains limited because of their suboptimal predictive performance.

Although we used machine learning to improve the predictive performance, our model may not seem to be significantly superior to those of previous studies.[7)8)] We hypothesized that AF characteristics differed across studies, thus leading to heterogeneous results. However, such a modest performance may be insignificant when the study focuses on *persistent* AF alone. For example, the studies using persistent AF patients tended to have modest performance despite machine learning; our data showed the best AUROC (0.63 [95% CI, 0.61–0.64]), whereas Vinter et al.[19)] showed AUROC, 0.59 (95% CI 0.51–0.68). We assume that the improved performance could be explained by the fact that we used 12-lead ECG data together with other clinical features. Our analysis showed that ECG features were more important than clinical features in machine learning (**Figure 4**). Another possibility is that predicting AF recurrence after ECV may be a difficult task in machine learning. According to a recent report, machine-learning prediction of AF recurrence after ECV was not superior to that of CHA_2DS_2-VASc scores, which was inconsistent with our findings (**Figure 3**, **Supplementary**

https://doi.org/10.4070/kcj.2023.0012

Figure 1).[20] However, this inconsistency might be explained by the fact that the previous study also lacked 12-lead ECG data for machine learning.

In our analysis, the XGBoost model trained with all clinical features showed a similar AUROC performance when compared to the C-statistic of AF duration; AUROC, 0.57 (95% CI, 0.55–0.58) and 0.58 (0.54–0.62), respectively. The slightly worse performance of XGBoost can be due to the suboptimal model tuning and/or the small size of the training dataset. Further tuning might allow a slightly better performance for the XGBoost model with clinical features, but it is unlikely to affect our overall conclusion on whether ECG features can be useful when used in addition to the clinical features.

In our study, a total of 42 features (21 features each for ECGs before and after electrical cardioversion) were utilized in the machine learning analysis of ECG features. These features included variables such as rates and amplitudes of R-wave peaks, which were averaged over the 10-second ECG recordings. Additionally, to emphasize the importance of variable RR intervals, we incorporated features such as RMSSD of RR intervals using lead II. This allowed the machine learning algorithm to learn the irregularity of RR intervals during AF. As a result, the model assigned higher importance to RMSSD of RR intervals during AF compared to those during sinus rhythm (**Figure 4**). Among the ECG features, the amplitudes of R-wave peaks and QRS duration were ranked as highly important (**Figure 4**). This observation can be partly explained by the association between advanced heart failure, diseased myocardium, decreased amplitudes of R-wave peaks, and widened QRS durations.

Instead of relying on predefined ECG features, there is a possibility that learning the raw data of ECGs could lead to improved performance. However, deep learning with raw ECG data requires a sufficiently large sample size to learn features effectively. In our study, we aimed to determine if a deep learning model utilizing raw ECG signal data could outperform machine learning using XGBoost. When we trained a few general architectures such as multilayer perceptron, ResNet-18, and ResNet-34 for deep learning analysis from scratch with raw ECG signal data, we obtained inferior results to those achieved using XGBoost (**Table 2**).

The inferior performance of deep learning with raw ECG data compared to machine learning with predefined ECG features can be attributed to several factors. One hypothesis is that the limited sample size hindered the capture of intricate ECG features that were predictive of AF recurrence after ECV. Our study population comprised only 718 patients, and each ECG sample contained 500 Hz (sampling rate) × 10 seconds (duration of ECG recording) = 5,000 dimensions. As the dimensionality of the training data increases, a larger number of samples becomes necessary to adequately learn important features. Therefore, the relatively limited size of our training samples may have contributed to the inferior performance of deep learning in our experiments. Our study did not exclude the possibility that deep learning of a sufficiently larger dataset may improve the performance. Additionally, utilizing a pretrained model with transfer learning or self-supervised learning approaches could allow us to improve the performance. In our study, however, we aimed to develop machine learning models solely based on our current dataset.

Based on the finding that ECG may be more beneficial than clinical features, we hypothesized that extended monitoring of biosignals i.e., single-lead ECG or PPG, could further improve the predictive performance. However, in our study, the small number of participants (n=261) seemed to limit appropriate feature learning and thus, led to no significant improvement

https://doi.org/10.4070/kcj.2023.0012

in performance (**Table 3**, **Supplementary Figure 2**). Because the extended monitoring data require a larger dimension of the neural network for feature learning, we assume that a larger participant number is needed for sufficient learning and feature exploitation to effectively predict AF recurrence.

Some limitations of this study should be noted. First, the study population might need to be increased to effectively learn the appropriate features for predicting AF recurrence after ECV. Transfer learning[21] could potentially be applied in our study if a relevant large-volume dataset including both ECGs and information on AF recurrence after direct current cardioversion were available. Unfortunately, to our knowledge, no such dataset currently exists. As a result, the utilization of transfer learning in our study is not feasible at present. Second, our study did not perform external validation, which may yield different results in other populations. While our analyses enabled internal validation of our machine learning model, we recognize the limitation of not having external validation, which hampers the generalizability of our model to new patients. To mitigate this issue, potential strategies like transfer learning or increasing the dataset size could be explored. However, both approaches necessitate a larger dataset, and regrettably, we have not encountered any suitable databases to fulfill this requirement to the best of our knowledge. Third, the detection of AF recurrence relied on either a 12-lead ECG or 24-hour Holter test; some patients with underlying low-burden AF might have been undetected. AF recurrence is a stochastic phenomenon, and undetected recurrences might adversely affect the machine-learning model. Fourth, sinus rhythm ECG after ECV may be less effective in predicting AF recurrence. Because sinus rhythm ECG was obtained while the atrium was in post-shock vulnerability, its information may be nonspecific and less beneficial to the machine-learning model. Consequently, its data may differ from those obtained after the relapse gap or electrical modeling.[22] This may partly explain why the features of sinus rhythm ECGs were less important than those of AF ECGs (**Figure 4**). Fifth, some of the known predictors e.g., natriuretic peptides[9] and high-sensitivity C-reactive proteins,[10] were not included in our model because of their unavailability.

In conclusion, a machine-learning approach was feasible for predicting AF recurrence after ECV in patients with persistent AF. Although machine learning predicts AF recurrence better than traditional statistical analysis, its clinical utility is limited owing to its modest predictive performance. Currently, given the modest performance of our model, relying solely on the model to predict AF recurrence may have limited utility. Therefore, it should be used as an adjunct tool to physician's evaluation. Prediction of AF recurrence after ECV seems difficult, partly because of the stochastic and multifactorial nature of AF. However, as machine learning gathered more information from AF patients, we observed that its performance improved. Therefore, further studies are warranted to improve predictive performance and validate the utility of machine learning in predicting patients with ineffective ECV.

SUPPLEMENTARY MATERIALS

Supplementary Table 1

A complete list of investigated clinical features

Click here to view

https://doi.org/10.4070/kcj.2023.0012

Supplementary Figure 1

The C-statistics of clinical features for 1-month AF recurrence after electrical cardioversion.

Click here to view

Supplementary Figure 2

The exploratory analysis – The impact of adding the 15-minute single-lead ECG and PPG dataset to the 12-lead ECG datasets on predictive performance.

Click here to view

REFERENCES

1. Kirchhof P, Benussi S, Kotecha D, et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Europace* 2016;18:1609-78.
PUBMED | CROSSREF
2. Juul-Möller S, Edvardsson N, Rehnqvist-Ahlberg N. Sotalol versus quinidine for the maintenance of sinus rhythm after direct current conversion of atrial fibrillation. *Circulation* 1990;82:1932-9.
PUBMED | CROSSREF
3. Van Gelder IC, Crijns HJ, Van Gilst WH, Van Wijk LM, Hamer HP, Lie KI. Efficacy and safety of flecainide acetate in the maintenance of sinus rhythm after electrical cardioversion of chronic atrial fibrillation or atrial flutter. *Am J Cardiol* 1989;64:1317-21.
PUBMED | CROSSREF
4. Van Gelder IC, Crijns HJ. Cardioversion of atrial fibrillation and subsequent maintenance of sinus rhythm. *Pacing Clin Electrophysiol* 1997;20:2675-83.
PUBMED | CROSSREF
5. Inoue K, Kurotobi T, Kimura R, et al. Trigger-based mechanism of the persistence of atrial fibrillation and its impact on the efficacy of catheter ablation. *Circ Arrhythm Electrophysiol* 2012;5:295-301.
PUBMED | CROSSREF
6. Brandes A, Crijns HJ, Rienstra M, et al. Cardioversion of atrial fibrillation and atrial flutter revisited: current evidence and practical guidance for a common procedure. *Europace* 2020;22:1149-61.
PUBMED | CROSSREF
7. Ehrlich JR, Schadow K, Steul K, Zhang GQ, Israel CW, Hohnloser SH. Prediction of early recurrence of atrial fibrillation after external cardioversion by means of P wave signal-averaged electrocardiogram. *Z Kardiol* 2003;92:540-6.
PUBMED | CROSSREF
8. Wałek P, Sielski J, Starzyk K, Gorczyca I, Roskal-Wałek J, Wożakowska-Kapłon B. Echocardiographic assessment of left atrial morphology and function to predict maintenance of sinus rhythm after electrical cardioversion in patients with non-valvular persistent atrial fibrillation and normal function or mild dysfunction of left ventricle. *Cardiol J* 2020;27:246-53.
PUBMED | CROSSREF
9. Andersson J, Rosenqvist M, Tornvall P, Boman K. NT-proBNP predicts maintenance of sinus rhythm after electrical cardioversion. *Thromb Res* 2015;135:289-91.
PUBMED | CROSSREF
10. Liu T, Li G, Li L, Korantzopoulos P. Association between C-reactive protein and recurrence of atrial fibrillation after successful electrical cardioversion: a meta-analysis. *J Am Coll Cardiol* 2007;49:1642-8.
PUBMED | CROSSREF
11. Vizzardi E, Curnis A, Latini MG, et al. Risk factors for atrial fibrillation recurrence: a literature review. *J Cardiovasc Med (Hagerstown)* 2014;15:235-53.
PUBMED | CROSSREF
12. Al'Aref SJ, Anchouche K, Singh G, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *Eur Heart J* 2019;40:1975-86.
PUBMED | CROSSREF
13. Feeny AK, Chung MK, Madabhushi A, et al. Artificial intelligence and machine learning in arrhythmias and cardiac electrophysiology. *Circ Arrhythm Electrophysiol* 2020;13:e007952.
PUBMED | CROSSREF

14. Breiman L. Random forests. *Mach Learn* 2001;45:5-32.
CROSSREF

15. Kwon S, Hong J, Choi EK, et al. Detection of atrial fibrillation using a ring-type wearable device (CardioTracker) and deep learning analysis of photoplethysmography signals: prospective observational proof-of-concept study. *J Med Internet Res* 2020;22:e16443.
PUBMED | CROSSREF

16. Hansen ML, Jepsen RM, Olesen JB, et al. Thromboembolic risk in 16 274 atrial fibrillation patients undergoing direct current cardioversion with and without oral anticoagulant therapy. *Europace* 2015;17:18-23.
PUBMED | CROSSREF

17. Apostolakis S, Haeusler KG, Oeff M, et al. Low stroke risk after elective cardioversion of atrial fibrillation: an analysis of the Flec-SL trial. *Int J Cardiol* 2013;168:3977-81.
PUBMED | CROSSREF

18. Hindricks G, Potpara T, Dagres N, et al. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur Heart J* 2021;42:373-498.
PUBMED | CROSSREF

19. Vinter N, Frederiksen AS, Albertsen AE, et al. Role for machine learning in sex-specific prediction of successful electrical cardioversion in atrial fibrillation? *Open Heart* 2020;7:e001297.
PUBMED | CROSSREF

20. Nuñez-Garcia JC, Sánchez-Puente A, Sampedro-Gómez J, et al. Outcome analysis in elective electrical cardioversion of atrial fibrillation patients: development and validation of a machine learning prognostic model. *J Clin Med* 2022;11:2636.
PUBMED | CROSSREF

21. Weimann K, Conrad TOF. Transfer learning for ECG classification. *Sci Rep* 2021;11:5251.
PUBMED | CROSSREF

22. Weijs B, Limantoro I, Delhaas T, et al. Cardioversion of persistent atrial fibrillation is associated with a 24-hour relapse gap: observations from prolonged postcardioversion rhythm monitoring. *Clin Cardiol* 2018;41:366-71.
PUBMED | CROSSREF