Deep Learning-Based Recurrence Prediction of Atrial Fibrillation After Catheter Ablation

Xue Zhou, BSc; Keijiro Nakamura, MD, PhD; Naohiko Sahara, MD; Takahito Takagi, MD; Yasutake Toyoda, MD; Yoshinari Enomoto, MD, PhD; Hidehiko Hara, MD, PhD; Mahito Noro, MD, PhD; Kaoru Sugi, MD, PhD; Masao Moroi, MD, PhD; Masato Nakamura, MD, PhD; Xin Zhu, PhD

Background: Radiofrequency catheter ablation (RFCA) is an effective therapy for atrial fibrillation (AF). However, the problem of AF recurrence remains. This study investigates whether a deep convolutional neural network (CNN) can accurately predict AF recurrence in patients with AF who underwent RFCA, and compares CNN with conventional statistical analysis.

Methods and Results: Three-hundred and ten patients with AF after RFCA treatment, including 94 patients with AF recurrence, were enrolled. Nine variables are identified as candidate predictors by univariate Cox proportional hazards regression (CPH). A CNNSurv model for AF recurrence prediction was proposed. The model's discrimination ability is validated by a 10-fold cross validation method and measured by C-index. After back elimination, 4 predictors are used for model development, they are N-terminal pro-BNP (NT-proBNP), paroxysmal AF (PAF), left atrial appendage volume (LAAV) and left atrial volume (LAV). The average testing C-index is 0.76 (0.72–0.79). The corresponding calibration plot appears to fit well to a diagonal, and the P value of the Hosmer-Lemeshow test also indicates the proposed model has good calibration ability. The proposed model has superior performance compared with the DeepSurv and multivariate CPH. The result of risk stratification indicates that patients with non-PAF, higher NT-proBNP, larger LAAV and LAV would have higher risks of AF recurrence.

Conclusions: The proposed CNNSurv model has better performance than conventional statistical analysis, which may provide valuable guidance for clinical practice.

Key Words: Atrial fibrillation; Catheter ablation; Deep learning; Electronic health record; Recurrence prediction

Atrial fibrillation (AF) is a common cardiac rhythm disturbance associated with low quality of life, and may increase the morbidity of other diseases such as stroke. AF is also closely related to heart failure and the prediction of heart failure.¹ Radiofrequency catheter ablation (RFCA) is an effective therapy for AF and can better restore sinus rhythm compared with drug therapy.² However, relatively high AF recurrence rates after RFCA constitute a persistent problem.³ After a successful completion of RFCA, it remains difficult to predict what would contribute to the radical cure or recurrence of AF. According to 2020 European Society of Cardiology Guidelines, AF duration, age, left atrial (LA) size, the abundance of epicardial fat tissues, and the presence of atrial substrate, are the risk factors of AF recurrence.⁴ Increased chronic diastolic

atrial pressure and overload tend to contribute to the creation of AF substrates, and atrial remodeling can affect the risk of recurrence.⁵

Factors related to AF recurrence are usually strongly correlated with one another. Although conventional statistical approaches, such as multivariate survival analysis, are commonly used to identify predictors of AF recurrence, missing values may affect the performance of these methods.⁶ Kornej et al built an APPLE score [one point for age >65 years, persistent AF, impaired eGFR (<60 mL/min/1.73 m²), LA diameter ≥43 mm, EF <50%] on 1,145 patients for the prediction of AF recurrence; the score showed an area under the curve (AUC) of 0.634.⁷ Potpara et al developed a MB-LATER score [assigning 1 point each for male gender, bundle branch block (i.e., QRS

Received July 12, 2021; revised manuscript received August 29, 2021; accepted September 1, 2021; J-STAGE Advance Publication released online October 8, 2021 Time for primary review: 20 days

Biomedical Information Engineering Lab, The University of Aizu, Aizu-Wakamatsu (X. Zhou, X. Zhu); Division of Cardiovascular Medicine, Toho University Ohashi Medical Center, Tokyo (K.N., N.S., T.T., Y.T., Y.E., H.H., M.M., M. Nakamura); and Division of Cardiovascular Medicine, Odawara Cardiovascular Hospital, Odawara (M. Noro, K.S.), Japan

The first two authors contributed equally to this work (X. Zhou, K.N.).

Mailing address: Xin Zhu, PhD, Biomedical Information Engineering Lab, The University of Aizu, Tsuruga, Ikkimachi, Aizu-Wakamatsu 965-8580, Japan. E-mail: zhuxin@u-aizu.ac.jp and Keijiro Nakamura, MD, PhD, Division of Cardiovascular Medicine, Toho University Ohashi Medical Center, Meguro-ku, Tokyo 153-8515, Japan. E-mail: nakamura1-04@live.jp

All rights are reserved to the Japanese Circulation Society. For permissions, please e-mail: cj@j-circ.or.jp

ISSN-1346-9843

Figure 1. Whole experimental framework. (A) The experiment process. (B) Details of the proposed CNNSurv.

complex duration of ≥ 120 ms), LA diameter ≥ 47 mm, persistent AF and early recurrence of AF during the 3-month post-ablation blanking period, and 2 points for pre-ablation history of long-standing persistent AF] for late AF recurrence prediction; the AUC was 0.62 for 226

patients.⁸ The prediction performance of the 2 scores are not satisfactory. In recent years, deep learning (DL) used for medical information analysis has achieved a considerable performance. Rajkomar et al used electronic health records (EHR) data from 2 US academic medical centers

with 216,221 adult patients and used DL to predict in-hospital death, readmission, length of stay, and discharge diagnoses. It was confirmed that all the prediction accuracy of DL models was better than the statistical results obtained using conventional methods.⁹ Katzman et al proposed a state-of-the-art non-linear survival method referred to as DeepSurv; it is a multilayer perceptron (MLP) and has ability to learn complex interactions rather than simple linear association between variables and event of failure.¹⁰ DeepSurv shows a considerable performance for predicting prognosis of various diseases such as survival prediction of oral cancer,¹¹ non-small cell lung cancer¹² and early triage of critically ill COVID-19 patients.¹³ Therefore, considering the successful usage of DL techniques in such study fields, we use DL to predict AF recurrence. However, MLP needs to train lots of parameters due to them being fully connected between layers, and the number of parameters will increase significantly as the number of input variables increases. Therefore, DeepSurv may be susceptible to overfitting when trained on a limited sample size. In this study, we propose a convolutional neural network-based prediction model cooperating with cox proportional hazards analysis (CNNSurv) to perform survival analysis to predict the recurrence of AF in patients after catheter ablation. Compared with MLP, the number of parameters in the CNN is lower, which benefits from parameter sharing. The proposed model demonstrates a better performance than the traditional Cox proportional hazard model (CPH)¹⁴ and DeepSurv.¹⁰

Methods

Study Population and Protocol

We enrolled 398 patients who received an initial catheter ablation for symptomatic drug-refractory paroxysmal AF (PAF) or non-PAF at Toho University Ohashi Medical Center between June 2016 and October 2019. PAF was defined as AF that terminated spontaneously within 7 days, and non-PAF was persistent and longstanding AF; these followed the 2013 guidelines of the Japan Circulation Society.¹⁵ The diagnosis of 2 patients could not gain a consensus. Four patients were excluded from the study because left atrial (LA) thrombus was detected on transthoracic and transesophageal echocardiography. Four patients were also removed because they could not receive a computed tomography (CT) scan. Eight patients were excluded due to hemodialysis and 3 patients were excluded because they had a history of cardiac surgery. In addition, 2 patients with atrial septal defects and 65 patients who could not be followed up or lacked an event indicator were also excluded. A total of 310 patients were finally enrolled in this study. **Figure 1A** illustrates the subjects' inclusion criteria and the whole experiment design.

The study protocol was approved by Institutional Review Board of Toho University Ohashi Medical Center (Approval No.: H21049), and informed consent was obtained from patients before participating in the study and releasing of the study data. Enhanced CT imaging and transthoracic echocardiography were performed before catheter ablation.

Data Collection

The data collection procedure is the same as that in outlined in the study by Takagi et al¹⁶ and is briefly described as follows.

Echocardiography Transthoracic echocardiography was

carried out before ablation only on patients with non-PAF to rule out the existence of intracardiac thrombi. We measured LA dimensions, left ventricular ejection fractions (LVEF), the values and ratio of early (E) and late (A) diastolic mitral inflow velocity, early diastolic velocity (e'), E/e', mitral regurgitation (MR) and tricuspid regurgitation (TR) according to American Society of Echocardiography guidelines. MR and TR were scored according to a previous report¹⁷ as follows: none, 0; none-mild, 0.5; mild, 1; mild-moderate, 1.5; moderate, 2; moderate-severe, 2.5; and severe, 3.

CT Images All patients underwent single-breath-holding contrast-enhanced electrocardiogram (ECG)-gated CT imaging with an 80-slice multidetector CT scanner (Aquilion; Canon Medical Systems, Japan) before catheter ablation. The LA volume (LAV), RA volume (RAV), and LA appendage volume (LAAV) were calculated during the atrial end-diastolic phase using semi-automated 3D reconstruction. Atrial volumes were determined by CT as previously reported in the study by Fuchs et al.¹⁸

Electro-Anatomical Mapping In patients with non-PAF, voltage mapping was obtained at sinus rhythms after ablation using a 7-F decapolar circular catheter (REFLEXION HD; Abbott, Inc.). Mapping points were acquired to fill all color gaps on the voltage map using an electro-anatomical mapping system. Each acquired point was classified according to the peak-to-peak electrogram as follows: >0.5 mV, healthy; 0.2–0.5 mV, diseased; and <0.1 mV, scarred. Low voltage area was defined as a site with ≥ 3 adjacent low voltage points of <0.5 mV, similar to what was reported in a previous study.¹⁹

Radiofrequency Catheter Ablation Pulmonary vein isolation (PVI) was performed using one 7-F decapolar circular catheter (Optima; St. Jude Medical, Inc.) positioned at the ostia of the ipsilateral PVs. We created bilateral circular lesions with wide-area circumferential ablation encircling the ipsilateral PV. Abbott FlexAbility ablation catheter or Abbott TactiCath contact force (CF) sensing catheters were used. A point-by-point ablation was performed. Energy was delivered at 25–35 W power with the temperature limited to 45°C.

A non-contact ablation protocol has been previously published.¹⁶ Contact guide ablation using a TactiCath CF sensing catheter targeted to 10–40 g, catheter stability and CF waveforms were monitored visually, and the lesion size was target to the standard value (from 5 to 6 points). Esophageal temperature monitoring was performed in all patients. No patient underwent extensive complex fractionated atrial electrogram ablation. Only patients with clinical or induced right atrial isthmus flutter underwent cavotricuspid isthmus (CTI) ablation. All non-CTI flutters and other atrial tachycardias (AT) were mapped and ablated using activation and entrainment mapping, if induced easily. Additionally, mitral isthmus ablation and/or roof-line ablation were rarely performed for the treatment of LA flutter and/or roof-dependent AT. Ablation in the superior vena cava (SVC) isolation were also rarely performed. Biphasic direct-current cardioversion (defibrillation) restored sinus rhythm if AF did not terminate spontaneously after successful PVI. The endpoint of the PVI was the creation of a bidirectional conduction block between the LA and the PVs.²⁰ All PVs were successfully isolated following the procedure.

Follow up Antiarrhythmic drugs (AADs) before the ablation were prescribed only if early recurrences of AF

were observed prior to discharge in patients with AF. Administration of AADs was stopped 3 months after ablation (blanking period) in patients who did not experience a recurrence of atrial tachyarrhythmia. All patients were followed up with a 12-lead ECG after 1 month and every 2 months. And 24h, Holter monitoring was performed every 6 and 12 months during follow up. If suggestive symptoms occurred, additional Holter monitoring and/or external loop recorder (SPIDERFLASH-t AFib: Sorin, France) would be performed to minimize the risk of missing the recurrence of AF after ablation. Recurrence of arrhythmia was defined as any atrial tachyarrhythmia lasting >30s documented by a 12-lead ECG or 24-h Holter monitoring after a 3-month blanking period from the ablation procedure.

Statistical Analysis

We select candidate variables for model development in 4 steps, as shown in the “variable selection” part of **Figure 1A**. The EHR data record 126 clinical variables; we first check the percentage of missing values of each variable and exclude variables that have >20% missing values. Second, numerical variables with missing values are imputed with mean value and the missing values of categorical variables are imputed with mode value. Third, we use univariate CPH analysis to check the association between each variable and AF recurrence; variables with statistically significant hazard ratios ($P < 0.05$) are selected. Then, multicollinearity is assessed by using Pearson’s correlation coefficient between pairs of variables (all ≤ 0.60), as was done in the study by Mesquita et al.²¹ The statistical analysis is performed by Python-3.7.7 on Windows X64.

Model Development

After variable selection, we use a 10-fold cross validation to split training and testing sets with stratified sampling for model development and evaluation, respectively. All numerical candidate variables are normalized by z-score normalization.²² The proposed CNNSurv has 3 convolutional branches, each with different size of 1D convolutional filters or stride sizes. In this way, the model is expected to learn from multiple variable sets consisting of different variable combinations. The detailed model structure for an input of 4 variables as an example is illustrated in **Figure 1B**. We set the number of convolutional kernels to be the same as the number of input variables. The kernel size in branch1 is the same as the input dimension to learn from all the input variables for global information capturing. Branch2 and branch3 extract local features with smaller kernel size and set stride to get different local combinations of input variables to increase feature diversity. We use batch normalization to the convolutional layer to speed up the training of a neural network²³ and use dropout to reduce overfitting and generalization error.²⁴ We use a rectified linear unit (ReLU) activation function for non-linearity learning. Huang et al suggested that for DL approaches, simpler models can have similar or better performances compared with more complex models in biological data analysis.²⁵ Additionally, complex models with more parameters may be prone to overfitting in small sample datasets. Therefore, we try to simplify our model as much as possible by directly connecting the output layer after convolutional layers. The output layer is a fully connected layer with a linear activation and outputs recurrence-free probability. The model is trained with an Adam optimizer, a learning rate of 0.05,

and a training batch size of 64. We set an early stopping if the loss function does not decrease in 20 consecutively epochs, otherwise the training process will continue to 1,000 epoch.

For comparison, we also build a DeepSurv model.¹⁰ It is a shallow and single branch model with 2 FC layers; the number of units in each layer is the same as the number of input variables. The first layer of DeepSurv is also followed by normalization, ReLU activation function and Dropout. Its training configurations are the same as those used in CNNSurv. Additionally, a multivariate CPH model is developed because it is the most used conventional statistical method for survival analysis.

Fewer and easily collected predictors are preferred, which enable a predictive model to be easily used in a clinical setting.²⁶ Therefore, we employ a backward elimination procedure for further variable selection and model simplification after the above initial model development. We set the testing C-index²⁷ as the deletion criterion. We delete the variable that has the greatest negative impact on testing performance and repeat this process until no further variable would reduce testing performance.

Model Evaluation

We use C-index, time-dependent AUC²⁸ and calibration plot, as used in the study by Miyagawa et al,²⁹ for the discrimination and calibration measurement of the 3 models. We use the Hosmer-Lemeshow test to check the statistical difference between actual and predicted recurrence-free probabilities in the calibration plot. We further calculate the recurrence-free function of each patient in the entire dataset and divided all patients into three groups: low-, medium- and high-risk groups, based on the first quartile, median and third quartile of predicted recurrence-free probability. Then, we fit Kaplan-Meier curves for the 3 risk groups and use a log-rank test to compare the recurrence-free curves among groups. In addition to the evaluation of model performance, we also assess the importance of predictors based on their effects on testing performance.

Results

Baseline Characteristics and Variable Selection

A total of 310 consecutive patients were recruited, with a mean follow up of 13.5 months (13.5 ± 6.9 months), during which 94 patients encountered AF recurrence. The clinical characteristics of patients with or without AF recurrence included in this study are summarized in **Supplementary Table 1**. We present variables as mean $\pm$ standard deviations or number and percentage. Univariate CPH analysis reveals that non-PAF (hazard ratio 2.92, 95% confidence interval (CI) (1.95–4.39), $P < 0.005$), AF duration (1.10, (1.05–1.16), $P < 0.005$), LAV (1.02, (1.01–1.02), $P < 0.005$), LA AV (1.04, (1.02–1.06), $P < 0.005$), LA diameter (LAD) (1.06, (1.02–1.10), $P < 0.005$), MR (1.39, (1.01–1.92), $P = 0.04$), TR (1.50, (1.08–2.08), $P = 0.02$), RAV (1.01, (1.01–1.01), $P < 0.005$) and log-transformed NT-proBNP (logNT-proBNP) (2.07, (1.42–3.02), $P < 0.005$) are the statistically significant risk factors of AF recurrence. Then, we use Pearson’s correlation coefficient (all ≤ 0.60) to check the multicollinearity among the above risk factors. Finally, all of the statistically significant risk factors are selected as candidate predictors for further modelling.

Model	Before back elimination	After back elimination	Best variables order
CNNSurv	Train: 0.73 (0.72–0.75); Test: 0.73 (0.68–0.78)	Train: 0.75 (0.74–0.75)*; Test: 0.75 (0.72–0.79)*	Train: 0.75 (0.74–0.76)**; Test: 0.76 (0.72–0.79)**
DeepSurv	Train: 0.73 (0.70–0.75); Test: 0.71 (0.64–0.78)	Train: 0.73 (0.71–0.75)*; Test: 0.74 (0.68–0.80)*	–
CPH	Train: 0.72 (0.72–0.73); Test: 0.70 (0.63–0.76)	Train: 0.72 (0.71–0.72)*; Test: 0.71 (0.65–0.77)*	–

Data are presented as mean (95% CI). *After back elimination, the models' performance becomes better than before.

**The best C-index.

Figure 2. Discrimination ability of the models (CNNSurv, DeepSurv, CPH). (A,B) the C-index of models before and after back elimination. (C,D) Time-dependent AUC of models.

Performance of Prediction Models and Predictor Importance

The proposed model is trained with the 9 predictors described above. Through the validation using the testing set, the average C-index of the proposed model is 0.73 (95% CI (0.68–0.78)), whereas the average C-indexes of DeepSurv and CPH are 0.71 (0.64–0.78) and 0.70 (0.63–0.76), respectively. After back elimination, the C-index of CNNSurv increases to 0.75 (0.72–0.79) when trained with 4 predictors: LAV, LAAB, PAF, and logNT-proBNP.

Due to sparse connectivity in the CNN, its results would slightly vary depending on the order of predictors fed into the model. Thus, we check all C-indexes when inputting the 4 predictors in a different order. All the results are listed in **Supplementary Table 2**. When the input is with the order of logNT-proBNP, LAAB, LAV, PAF, CNNSurv obtains the best C-index of 0.76 (0.72–0.79). We performed the same elimination procedure for DeepSurv and CPH, and the C-index of DeepSurv increases to 0.74 (0.68–0.80) with predictors: PAF, duration, LAV, LAD, RAV,

Figure 3. Calibration plots of CNNSurv (A), DeepSurv (B), and CPH (C) for 1- and 2-year AF recurrence prediction.

proNT_logBNP. CPH needs only 3 predictors: PAF, duration, LAV, to obtain its best C-index 0.71 (0.65–0.77). As summarized in **Table**, with or without back elimination, CNNSurv always outperforms DeepSurv and CPH. **Figure 2A** and **2B** show that the testing C-index of CNNSurv has a shorter interquartile range (IQR) than that of DeepSurv and CPH, indicating that CNNSurv may have better robustness. Furthermore, when the 3 models trained on fewer variables determined by back elimination, their C-indexes have less variance.

In addition to the C-index, we also evaluated time-dependent AUC of models, as shown in **Figure 2C,D**. We set the upper boundary to 24 months, because the censoring rate is high at later time points. During follow up, the average time-dependent AUCs of CNNSurv, DeepSurv and CPH are 0.76 (0.75–0.78), 0.75 (0.72–0.78) and 0.69 (0.64–0.73), respectively. CNNSurv has more stable AUCs with shorter IQR and its AUC is always >0.70 at each time point. However, DeepSurv and CPH may have extremely low AUCs at some points, as shown in **Figure 2C**. DL methods always have better AUCs than CPH for the prediction of recurrence after 1 year.

We evaluate the calibration of the proposed model using

the calibration plot shown in **Figure 3A**. It fits well to a diagonal line ($y=x$), and the P values for goodness of fit are 1.00 for both 1- and 2-year recurrence prediction, respectively. The calibration plots of DeepSurv and CPH are shown in **Figure 3B,C**, respectively. For 1- and 2-year recurrence prediction, P values for goodness of fit are 1.00 and 0.9998, respectively, in DeepSurv, and 0.9991 and 0.9987, respectively, in CPH. CNNSurv has a reliable alignment with a diagonal line for 1-year recurrence prediction, especially when recurrence-free probability is <0.9. The calibration plots of DeepSurv and CPH are more dispersive. DeepSurv tends to underestimate recurrence-free probability when it is >0.8, whereas in most cases, CPH tends to overestimate the survival probability.

Figure 4A–C shows the risk stratification conducted by CNNSurv, DeepSurv and CPH, respectively. For all 3 models, the log-rank test indicates that the recurrence rates of low- and high-risk groups have significant difference. However, low- and medium-risk groups stratified by CPH do not show significant difference. This indicates CPH fails to identify patients with low- or medium-risk of recurrence. Although DeepSurv has the ability to identify low-risk or medium-risk recurrence patients, CNNSurv provides a more

Figure 4. Kaplan-Meier curve analysis of the 3 patients risk groups (low-, medium- and high-risk). **(A)** Risk stratified by CNNSurv. The survival probabilities of low-, medium and high-risk groups show obvious boundary. **(B)** Risk stratified by DeepSurv. **(C)** Risk stratified by CPH. CPH has no satisfactory ability to distinguish between low-risk and medium-risk groups.

obvious difference between the 2 groups, which shows a better prediction ability. **Supplementary Table 3** describes statistics of baseline characteristics of each risk group stratified by CNNSurv. The result reveals that patient with non-PAF, higher NT-proBNP, larger LAAV and LAV would have higher risk of recurrence. The recurrence rate in low-, medium- and high-risk groups is 7.7%, 24.0% and 65.4%, respectively, showing good risk identification ability.

In addition to evaluating performances of models, we also explore the importance of predictors, as shown in **Figure 5**. LAV has the most important contribution to all the 3 models. PAF also has a positive effect on all the 3 models. In CNNSurv, duration is not a key predictor, but it is an important predictor for DeepSurv and CPH. LAAV and logNT-proBNP are important for CNNSurv, but they are not the key predictors for CPH. DeepSurv needs more key predictors to support decision-making, such as RAV

and LAD. These predictors' importance varies with each model; the exception is LAV, which is the most important predictor in all 3 models.

Discussion

Catheter ablation, mainly via PVI, is an effective treatment to restore sinus rhythm for patients with AF,² but the prognosis after therapy is complex, this includes recurrence or recurrence-free, short-term or long-term recurrence. Predictors for the prognosis are also miscellaneous.^{7,8,21,30-33} In this study, we propose a CNNSurv for AF recurrence prediction. Its discrimination and calibration abilities are evaluated using C-index, time-dependent AUC, and calibration plot. They are better than those of DeepSurv and CPH.

Previous studies have demonstrated that DL can be

effectively applied for cancer prognosis.²⁵ In this study, the performance of DL models is better than that of CPH. This may be attributed to the fact that the linearity in CPH is too simple to fit the relationship between predictors and outcome. In contrast, DL has no linear fitting assumption, and it can learn more complex relationships between EHR and outcome. Compared with the fully connected networks used in DeepSurv, CNNs have the characteristic of weight sharing, which can effectively reduce the number of parameters. In fully connected networks, each neuron in 1 layer is connected with all the neurons in the preceding or successive layer; therefore, the connection parameters between the 2 layers would fold increase as the number of neurons increases. For example, when there are i input variables, the input layer has i neurons. If we suppose the next layer has j neurons, the number of connection weights would be $i \times j$. For high-dimensional input, the number of parameters in DeepSurv would be dramatically high and thus more training samples are needed. The implementation of a CNN in the analysis of EHR data is unconventional, but some studies have found that a CNN is a reliable method to represent and analyze EHR data.³⁴⁻³⁶ In this study, the proposed CNN model shows a superior performance compared with DeepSurv and a traditional statistical model.

In CNNSurv, we only use 4 predictors: logNT-proBNP,

PAF, LAHV and LAV, which are commonly available in most hospital data. LAV gives the most important contribution to models, and it indicates that LAV has the strongest relationship with AF recurrence. This may due to its association with structural remodeling in AF. A similar conclusion was also found by Mesquita et al²¹ and Wang et al.³⁷ Type of AF was a commonly used predictor for AF recurrence prediction in some studies,^{7,21,30-33,38} and these studies indicated that non-PAF patients had a higher risk of recurrence. This conclusion is also compatible with our results, as shown in **Supplementary Table 3**, where the proportions of patients with non-PAF in the low-, medium- and high-risk groups are 0%, 21.4% and 88.5%, respectively. NT-proBNP is confirmed as the strongest predictor of incident AF,³⁹ and Zhang et al suggested that the NT-proBNP level would be a predictor of AF recurrence after catheter ablation because a higher NT-proBNP level was associated with a higher risk of recurrence.⁴⁰ In the 3 risk groups classified by the proposed model, patients in the high-risk group have the highest logNT-proBNP (2.8 ± 0.4). Furthermore, the LAHV of patients in the high-risk group are apparently larger than those of patients in the low-risk group: 13.0 ± 7.2 vs 8.4 ± 3.0 mL. This result is also mentioned by Teixeira et al, whereby a larger LAHV was associated with AF recurrence.⁴¹

In the 3 risk groups, increased LA and RA structural variables are associated with higher risks of recurrence, as described in **Supplementary Table 3**. A significant functional MR or TR is defined as at least a moderate degree of MR or TR, and the high-risk group has higher proportions of significant functional MR and TR than those of the low- and medium-risk groups (MR: high vs. medium vs. low: 14.0% vs. 8.5% vs. 2.5%; TR: 25.6% vs. 7.8% vs. 2.6%). Patients in the high-risk group show a higher HR and a higher proportion of patients have HT. Variables of gender, height, weight, body surface area and body mass index have no significant difference among the 3 risk groups. Some studies suggested that age is a predictor for AF recurrence,^{7,21,32,33} but in this study, age does not show a statistically significant difference among the risk groups. The reason might be that 70% of patients in this study are aged >60 years, and all patients are aged >20 years and only 1.9% of them are aged <40 years.

Limitations of this study are as follows:

- (1) The recurrence event in patients may be underestimated due to a lack of continuous ECG monitoring; however, the purpose of catheter ablation is to improve patients' symptoms; the follow-up protocol used in this study can capture the essential prognostic condition after treatment.²¹
- (2) Some variables have a high proportion of missing values such as variables related to sleep, or are not recorded in this dataset; for example, previous ablations, thus hindering their inclusion in model development.
- (3) The proposed model is developed based on a small cohort. As a result, although the proposed model shows a considerable discrimination and calibration performance in our dataset, the patients in this study are biased to being elderly and lacking an external dataset. Therefore, the generalization of the model should be further validated on samples from multiple hospitals and populations.
- (4) We cannot provide an explicit survival function or equation, and we cannot suggest specific cut-off values of predictors because of the "black-box" characteristic of the model.

Conclusions

In this study, we proposed a DL model with 4 predictors: LAV, PAF, logNT-proBNP and LAAV. The model demonstrates better discrimination and calibration ability than conventional statistical analysis, and may provide valuable guidance to determine patients who will benefit most from the ablation treatment and avoid unnecessary treatment and costs; for hospitals, it may guide the rational use of medical resources.

Acknowledgments

This research is supported by a JSPS Kakenhi Basic Research Fund, C 18K11532 and 21K10287, and a Competitive Research Fund from The University of Aizu, 2021-P-5.

Disclosures

The authors declare no conflicts of interest.

IRB Information

The study protocol was approved by the Institutional Review Board of Toho University Ohashi Medical Center (Approval No.: H21049),

and informed consent was obtained from patients before participating in the study and releasing of the study data.

Data Availability

The deidentified participant data will not be shared.

References

1. Fuster V, Rydén LE, Asinger RW, Cannom DS, Crijns HJ, Frye RL, et al. ACC/AHA/ESC Guidelines for the management of patients with atrial fibrillation: Executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines and Policy Conferences (Committee to Develop Guidelines for the Management of Patients with Atrial Fibrillation) developed in collaboration with the North American Society of Pacing and Electrophysiology. *Circulation* 2001; **104**: 2118–2150.
2. Packer DL, Mark DB, Robb RA, Monahan KH, Bahnson TD, Poole JE, et al. Effect of catheter ablation vs antiarrhythmic drug therapy on mortality, stroke, bleeding, and cardiac arrest among patients with atrial fibrillation: The CABANA randomized clinical trial. *JAMA* 2019; **321**: 1261–1274.
3. Pallisgaard JL, Gislason GH, Hansen J, Johannessen A, Torp-Pedersen C, Rasmussen PV, et al. Temporal trends in atrial fibrillation recurrence rates after ablation between 2005 and 2014: A nationwide Danish cohort study. *Eur Heart J* 2018; **39**: 442–449.
4. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, 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). *Eur Heart J* 2020; **42**: 373–498.
5. Iwasaki YK, Nishida K, Kato T, Nattel S. Atrial fibrillation pathophysiology: Implications for management. *Circulation* 2011; **124**: 2264–2274.
6. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol* 1996; **49**: 907–916.
7. Kornej J, Hindricks G, Shoemaker MB, Husser D, Arya A, Sommer P, et al. The APPLE score: A novel and simple score for the prediction of rhythm outcomes after catheter ablation of atrial fibrillation. *Clin Res Cardiol* 2015; **104**: 871–876.
8. Potpara TS, Mujovic N, Sivasambu B, Shantsila A, Marinkovic M, Calkins H, et al. Validation of the MB-LATER score for prediction of late recurrence after catheter-ablation of atrial fibrillation. *Int J Cardiol* 2019; **276**: 130–135.
9. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 2018; **1**: 1–10.
10. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol* 2018; **18**: 1–12.
11. Kim DW, Lee S, Kwon S, Nam W, Cha IH, Kim HJ. Deep learning-based survival prediction of oral cancer patients. *Sci Rep* 2019; **9**: 1–10.
12. She Y, Jin Z, Wu J, Deng J, Zhang L, Su H, et al. Development and validation of a deep learning model for non-small cell lung cancer survival. *JAMA Netw Open* 2020; **3**: e205842.
13. Liang W, Yao J, Chen A, Lv Q, Zanin M, Liu J, et al. Early triage of critically ill COVID-19 patients using deep learning. *Nat Commun* 2020; **11**: 1–7.
14. Cox DR. Regression models and life-tables. *J R Stat Soc Ser B Methodol* 1972; **34**: 187–220.
15. JCS Joint Working Group. Guidelines for pharmacotherapy of atrial fibrillation (JCS 2013): Digest version. *Circ J* 2014; **78**: 1997–2021.
16. Takagi T, Nakamura K, Hashimoto H, Asami M, Ishii R, Enomoto Y, et al. The impact of sleep apnea on right atrial structural remodeling with atrial fibrillation. *J Cardiol* 2020; **75**: 665–672.
17. Abe Y, Akamatsu K, Ito K, Matsumura Y, Shimeno K, Naruko T, et al. Prevalence and prognostic significance of functional mitral and tricuspid regurgitation despite preserved left ventricular ejection fraction in atrial fibrillation patients. *Circ J* 2018; **82**: 1451–1458.
18. Fuchs A, Mejdahl MR, Kühl JT, Stisen ZR, Nilsson EJP, Køber LV, et al. Normal values of left ventricular mass and cardiac chamber volumes assessed by 320-detector computed tomography angiography in the Copenhagen General Population Study.

- Eur Heart J Cardiovasc Imaging* 2016; **17**: 1009–1017.
19. Masuda M, Fujita M, Iida O, Okamoto S, Ishihara T, Nanto K, et al. Influence of underlying substrate on atrial tachyarrhythmias after pulmonary vein isolation. *Heart Rhythm* 2016; **13**: 870–878.
  20. Yamasaki H, Tada H, Sekiguchi Y, Igarashi M, Arimoto T, Machino T, et al. Prevalence and characteristics of asymptomatic excessive transmural injury after radiofrequency catheter ablation of atrial fibrillation. *Heart Rhythm* 2011; **8**: 826–832.
  21. Mesquita J, Ferreira AM, Cavaco D, Moscoso Costa F, Carmo P, Marques H, et al. Development and validation of a risk score for predicting atrial fibrillation recurrence after a first catheter ablation procedure: ATLAS score. *Europace* 2018; **20**: f428–f435.
  22. Shalabi LA, Shaaban Z, Kasasbeh B. Data mining: A preprocessing engine. *J Comput Sci* 2006; **2**: 735–739.
  23. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning, PMLR* 2015; **37**: 448–456.
  24. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014; **15**: 1929–1958.
  25. Huang Z, Johnson TS, Han Z, Helm B, Cao S, Zhang C, et al. Deep learning-based cancer survival prognosis from RNA-seq data: Approaches and evaluations. *BMC Medical Genom* 2020; **13**: 1–12.
  26. Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Ann Intern Med* 2019; **170**: W1–W33.
  27. Frank E, Harrel JR, Robert MC, David BP, Kerry LL, Robert AR. Evaluating the yield of medical tests. *JAMA* 1982; **247**: 2543–2546.
  28. Lambert J, Chevret S. Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent ROC curves. *Stat Methods Med Res* 2016; **25**: 2088–2102.
  29. Miyagawa S, Pak K, Hikoso S, Ohtani T, Amiya E, Sakata Y, et al. Japan heart failure model: Derivation and accuracy of survival prediction in Japanese heart failure patients. *Circ Rep* 2019; **1**: 29–34.
  30. Berkowitsch A, Kuniss M, Greiss H, Wojcik M, Zaltsberg S, Lehinant S, et al. Impact of impaired renal function and metabolic syndrome on the recurrence of atrial fibrillation after catheter ablation: A long term follow-up. *Pacing Clin Electrophysiol* 2012; **35**: 532–543.
  31. Canpolat U, Aytemir K, Yorgun H, Şahiner L, Kaya EB, Oto A. A proposal for a new scoring system in the prediction of catheter ablation outcomes: Promising results from the Turkish Cryoablation Registry. *Int J Cardiol* 2013; **169**: 201–206.
  32. Winkle RA, Jarman JW, Mead RH, Engel G, Kong MH, Fleming W, et al. Predicting atrial fibrillation ablation outcome: The CAAP-AF score. *Heart Rhythm* 2016; **13**: 2119–2125.
  33. Tang RB, Dong JZ, Long DY, Yu RH, Ning M, Jiang CX, et al. Efficacy of catheter ablation of atrial fibrillation beyond HATCH score. *Chin Med J* 2012; **125**: 3425–3429.
  34. Gupta V, Sachdeva S, Bhalla S. A novel deep similarity learning approach to electronic health records data. *IEEE Access* 2020; **8**: 209278–209295.
  35. Zhao J, Feng Q, Wu P, Lupu RA, Wilke RA, Wells QS, et al. Learning from longitudinal data in electronic health record and genetic data to improve cardiovascular event prediction. *Sci Rep* 2019; **9**: 717.
  36. Cheng Y, Wang F, Zhang P, Hu J. Risk prediction with electronic health records: A deep learning approach. *Proceedings of the 2016 SIAM International Conference on Data Mining (SDM)*, doi:10.1137/1.9781611974348.49 (accessed June 27, 2021).
  37. Wang Q, Zhuo C, Shang Y, Zhao J, Chen N, Lv N, et al. U-shaped relationship between left atrium size on echocardiography and 1-year recurrence of atrial fibrillation after radiofrequency catheter ablation: Prognostic Value Study. *Circ J* 2019; **83**: 1463–1471.
  38. Watanabe R, Nagashima K, Wakamatsu Y, Otsuka N, Yokoyama K, Matsumoto N, et al. Different determinants of the recurrence of atrial fibrillation and adverse clinical events in the mid-term period after atrial fibrillation ablation. *Circ J*, doi:10.1253/circj.CJ-21-0326.
  39. Svennberg E, Lindahl B, Berglund L, Eggers KM, Venge P, Zethelius B, et al. NT-proBNP is a powerful predictor for incident atrial fibrillation: Validation of a multimarker approach. *Int J Cardiol* 2016; **223**: 74–81.
  40. Zhang Y, Chen A, Song L, Li M, Chen Y, He B. Association between baseline natriuretic peptides and atrial fibrillation recurrence after catheter ablation: A meta-analysis. *Int Heart J* 2016; **57**: 183–189.
  41. Teixeira PP, Oliveira MM, Ramos R, Rio P, Cunha PS, Delgado AS, et al. Left atrial appendage volume as a new predictor of atrial fibrillation recurrence after catheter ablation. *J Interv Card Electrophysiol* 2017; **49**: 165–171.

Supplementary Files

Please find supplementary file(s);
<http://dx.doi.org/10.1253/circj.CJ-21-0622>