

EDITORIAL COMMENT

A Paradigm Shift in Risk Prediction in Patients With Atrial Fibrillation*



Eue-Keun Choi, MD, PhD, Soonil Kwon, MD

Atrial fibrillation (AF) and heart failure (HF) are the new cardiovascular epidemics. The prevalence of HF is anticipated to increase by 46%, affecting approximately 8 million people older than 18 years.¹ By 2030, HF-related expenses are expected to increase by 127% to \$69.8 billion. HF imposes a considerable cost burden on society because it is the most prevalent discharge diagnosis. AF is the most prevalent arrhythmia in clinical practice, with an estimated global prevalence of 46.3 million. The prevalence of AF in Asian countries has been increasing because of an aging population and increased detection rates. AF and HF are closely associated, and the prevalence of AF in recent HF studies has been reported to range from 13% to 27%. Although the relationship between the 2 conditions has not been fully understood, their coexistence can be partially explained by common risk factors, such as increasing age, hypertension, diabetes, and structural heart disease. AF can accelerate the occurrence or progression of HF in several ways. Increased heart rate and exercise-induced hypertrophy decrease cardiac output by shortening diastolic filling time. The irregular ventricular response may have a further impact. When both AF and HF are present, the prognosis is worse, with an elevated risk of stroke and mortality. Rate and rhythm control of AF in patients with HF is hindered by various factors. In recent randomized clinical trials, AF catheter ablation in patients with HF with reduced ejection fraction showed better outcomes than medical therapy,

including survival, quality of life, ventricular function, and HF hospitalization.^{2,3} However, selecting patients with AF who are at a high risk for hospitalization is still challenging. Therefore, a better strategy to identify patients with AF is needed to reduce the risk of HF aggravation.

In this issue of *JACC: Asia*, Hamatani et al⁴ reported the predictive power of machine learning (ML) for incident HF in patients with AF. The Fushimi AF Registry is a community-based prospective survey of patients with AF in Fushimi-ku, Kyoto, Japan. The authors constructed an ML model using the derivation and validation cohorts and predicted HF hospitalization. Six ML models were tested, and the top 7 variables (age, history of HF, creatinine clearance, cardiothoracic ratio on x-ray, left ventricular [LV] ejection fraction, LV end-systolic diameter, and LV asynergy) were used in the random forest algorithm. All the ML algorithms showed a higher performance for HF hospitalization in patients with AF compared to the conventional risk prediction model, that is, the Framingham HF risk model (area under the receiver-operating characteristic curve: 0.75 vs 0.67; $P < 0.001$). Although a history of HF was the most important variable among the 6 ML models, the ML model could stratify the probability of HF hospitalization in patients without a history of HF. When stratifying the patients by tertiles in the random forest algorithm using the 7 variables, high-risk patients had a 12-fold higher risk of HF hospitalization than low-risk patients.

Recently, ML has been rapidly introduced into medical research. However, some considerations exist when conducting or interpreting medical research using ML. First, ML does not always outperform traditional statistical analyses. Based on a systematic review,⁵ ML may not always have superior performance to logistic regression analysis when analyzing a clinical prediction model, as in this study. The performance of C-statistics for predicting HF among patients with AF using a combination of various clinical parameters showed an area under the

*Editorials published in *JACC: Asia* reflect the views of the authors and do not necessarily represent the views of *JACC: Asia* or the American College of Cardiology.

From the Department of Internal Medicine, Seoul National University College of Medicine and Seoul National University Hospital, Seoul, South Korea.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the [Author Center](#).

TABLE 1 Considerations and Checklists in Performing Medical Research Using ML^{9,10,12,13}

Considerations	Checklists
Distinguishing results from previous statistical analysis	<ul style="list-style-type: none"> ● Is ML analysis coherent with previous results? ● Is ML analysis outperforming previous results? ● Is ML analysis using unstructured data? ● Is ML analysis producing novel findings?
Appropriate internal and external validation	<ul style="list-style-type: none"> ● Is there an appropriate process for internal validation? ● Has external validation been performed and coherent results obtained? ● Is the population for external validation independent from that used for the construction and testing of the ML model?
Proper control of bias	<ul style="list-style-type: none"> ● Is the population size appropriate to perform the presented ML analysis? ● Is the performance validation process reasonable and appropriate? ● Have the missing data been handled properly? ● Is there control and measurement of potential overfitting? ● Is the study performed according to a checklist designed for medical research using ML?

ML = machine learning.

receiver-operating characteristic curve of 0.717 (95% CI: 0.705-0.732).⁶ Therefore, at first glance, the ML results of this study may not appear to be strikingly superior to those of previous non-ML analyses. However, considering that the current study analyzed only 4,395 patients, much fewer than the previous study (n = 23,503), the variables used in the models were different. ML analysis is expected to outperform statistical analysis when a larger population and more variables are used. Second, when ML is used in medical research, external validation is required for reproducibility and generalizability.⁷ Also, the population used for external validation should be independent of those used to construct and validate the ML model. An appropriate external validation process can determine whether the ML results are applicable to other populations. If the external validation results are different, selection bias of the source population can be suspected. Unfortunately, the current study included only patients from the Fushimi region of Kyoto, Japan. Verifying whether similar results are reproducible in other areas or ethnicities is necessary. Therefore, the superiority of the HF prediction model using ML requires further validation. Finally, when conducting medical research using ML, it is essential to be aware of the bias related to ML and make an effort to reduce it. Many ML studies may have bias associated with a small study population, inadequate handling of missing values, failure to control overfitting, and so on.⁸ If such a bias is not controlled, biased results may be produced, ultimately failing external validation. Therefore, when ML is used, accurate data labeling, appropriate data set construction, preprocessing, and proper ML model evaluation are needed to reduce bias and derive correct results.

When using ML methods in medical research, it would be helpful to use a checklist to reduce bias and report scientifically. With the advent of ML methodology in medical research, a number of checklists have been suggested. The CLAIM (Checklist for Artificial Intelligence in Medical Imaging)⁹ and MINIMAR (Minimum Information for Medical AI Reporting)¹⁰ are examples of currently available tools. The TRIPOD (Transparent Reporting of a Multivariable Prediction Model of Individual Prognosis or Diagnosis) statement and the PROBAST (Prediction Model Risk of Bias Assessment Tool) were previously used to evaluate the adequacy of the prediction model. To apply these tools in ML analysis, TRIPOD-Artificial Intelligence (AI) and PROBAST-AI are underway and expected to be published soon.¹¹ These evaluation tools will help reduce bias in ML research. **Table 1** summarizes some critical considerations for performing ML analyses in medical research.

Although Hamatani et al⁴ performed well-designed ML analyses to predict HF among patients with AF, some cautions exist when interpreting the results. First, this study did not investigate the impact of concomitant cardiovascular medications, which are crucial factors for incident HF. Hamatani et al⁴ responded that it was difficult to interpret the cause-effect relationship between cardiovascular medication use and incident HF. Nevertheless, poorly controlled AF and underlying cardiovascular diseases negatively affect HF; therefore, medication usage should be considered in predicting HF among patients with AF. Second, there were significant differences in some baseline characteristics and annual incidence rates of HF hospitalization (16% vs 11%; $P < 0.001$) between the derivation and validation

cohorts. Significantly deviated characteristics between derivation and validation cohorts may limit the reproducibility of the ML model. Third, it is unfortunate that unstructured data, such as raw data of electrocardiograms or x-ray images, were not analyzed using ML. For structured data, conventional statistical analysis may have comparable predictive performance to ML analysis. Finally, a lack of external validation may limit the generalizability of the study results.

In summary, predicting the high-risk group for HF among patients with AF is crucial. The current study evaluated the ML model for predicting incident HF in Japanese patients with AF using various clinical variables. Although the current ML analysis could improve the predictive power of HF among patients with AF, some limitations remain when interpreting

the study results. By combining the results of traditional statistical analysis and ML analysis, we hope to effectively predict incident HF in patients with AF.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

Dr Choi has received research grants or speaker fees from Abbott, Bayer, BMS/Pfizer, Biosense Webster, Chong Kun Dang, Daewoong Pharmaceutical Co, Daiichi-Sankyo, DeepQure, Dreamtech Co, Ltd, Jeil Pharmaceutical Co Ltd, Medtronic, Samjinpharm, Seers Technology, and Skylabs. Dr Kwon has reported that he has no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Eue-Keun Choi, Department of Internal Medicine, Seoul National University College of Medicine and Seoul National University Hospital, 101, Daehak-ro, Jongno-gu, Seoul 03080, South Korea. E-mail: choiek17@snu.ac.kr.

REFERENCES

1. Virani SS, Alonso A, Aparicio HJ, et al. Heart disease and stroke statistics—2021 update: a report from the American Heart Association. *Circulation*. 2021;143:e254–e743. <https://doi.org/10.1161/CIR.0000000000000950>
2. Marrouche NF, Brachmann J, Andresen D, et al. Catheter ablation for atrial fibrillation with heart failure. *N Engl J Med*. 2018;378:417–427.
3. Prabhu S, Taylor AJ, Costello BT, et al. Catheter ablation versus medical rate control in atrial fibrillation and systolic dysfunction: the CAMERA-MRI study. *J Am Coll Cardiol*. 2017;70:1949–1961.
4. Hamatani Y, Nishi H, Iguchi M, et al. Machine learning risk prediction for incident heart failure in patients with atrial fibrillation. *JACC: Asia*. 2022;2(6):706–716.
5. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22.
6. Krisai P, Johnson LSB, Moschovitis G, et al. Incidence and predictors of heart failure in patients with atrial fibrillation. *CJC Open*. 2021;3:1482–1489.
7. Ho SY, Phua K, Wong L, Bin Goh WW. Extensions of the external validation for checking learned model interpretability and generalizability. *Patterns (N Y)*. 2020;1:100129. <https://doi.org/10.1016/j.patter.2020.100129>
8. Andaur Navarro CL, Damen JAA, Takada T, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *BMJ*. 2021;375:n2281. <https://doi.org/10.1136/bmj.n2281>
9. Mongan J, Moy L, Kahn CE Jr. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell*. 2020;2:e200029. <https://doi.org/10.1148/ryai.2020200029>
10. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): developing reporting standards for artificial intelligence in health care. *J Am Med Inform Assoc*. 2020;27:2011–2015.
11. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open*. 2021;11:e048008. <https://doi.org/10.1136/bmjopen-2020-048008>
12. Shelmerdine SC, Arthurs OJ, Denniston A, Sebire NJ. Review of study reporting guidelines for clinical studies using artificial intelligence in healthcare. *BMJ Health Care Inform*. 2021;28(1):e100385. <https://doi.org/10.1136/bmjhci-2021-100385>
13. Whalen S, Schreiber J, Noble WS, Pollard KS. Navigating the pitfalls of applying machine learning in genomics. *Nat Rev Genet*. 2022;23:169–181.

KEY WORDS atrial fibrillation, heart failure, machine learning, risk prediction