EDITORIAL COMMENT

Artificial Intelligence to Complement, Not Replace, Clinical Knowledge

Reading Between the Lines*

Joshua P. Barrios, PhD,[a] Madhav R. Seshadri, MD,[b] Geoffrey H. Tison, MD, MPH[a,c,d]

A cornerstone of cardiovascular diagnosis, the electrocardiogram provides an information-rich record of cardiac electric activity. Although standard clinical electrocardiographic (ECG) interpretation relies on recognizing well-understood ECG markers to diagnose abnormalities, in the past 10 years, machine learning or artificial intelligence (AI) methods applied to ECG analysis have demonstrated that ECG waveforms contain more information than was previously appreciated. Modern AI models are capable of learning novel patterns within raw ECG voltage data, which have allowed them, in some cases, to predict diagnoses or outcomes that were previously not possible by electrocardiography alone.

In a recent compelling example, Attia et al[1] demonstrated that a neural network AI model could analyze the raw waveforms from a sinus rhythm electrocardiogram and estimate the risk for future atrial fibrillation (AF) over the subsequent 31 days. Patients who are currently in AF, among the most common rhythm abnormalities, exhibit the well-recognized pattern of irregularly irregular QRS activations and lack P waves. And although some sinus rhythm ECG predictors of future risk for AF have been previously recognized,[2] there are no commonly applied criteria to readily predict the likelihood of future AF from sinus rhythm electrocardiograms, making this a novel AI-enabled task with implications for earlier AF intervention and prevention.

Before such AI models are brought into real-world clinical practice, several important considerations should be made. The performance of a broadly trained AI model should be evaluated in a target clinical population, particularly if the target population differs substantially from the population in which the model was derived. Also, one must determine how to integrate an AI model's predictions with existing clinical knowledge. In addition, the AI model's predictions should ideally be integrated into the clinical workflow to ensure that they complement rather than complicate patient care.

In this issue of *JACC: CardioOncology*, Christopoulos et al[3] present a study that begins to address at least the first 2 of these considerations. The investigators used the same AI model previously described by Attia et al[1] to predict AF risk from baseline sinus rhythm electrocardiograms in patients with chronic lymphocytic leukemia (CLL). Quantifying the risk for AF in patients with CLL is especially important because of the association of Bruton's tyrosine kinase (BTK) inhibitors, a standard treatment for CLL in the frontline and relapsed or refractory settings, with AF.[4] The study demonstrated that the previously trained AI model successfully stratified increased AF risk among patients with CLL, with high AI ECG prediction associated with an HR of 3.9 and 2 to 5 times higher rates of 2-, 5-, and 10-year risk for AF compared with those with low predicted AI ECG risk. This study provides an important step toward the clinical implementation of this model in the CLL population.

Perhaps more important, this study illuminates the point that AI can, in the right settings, capture information that is complementary to existing clinical

From the [a]Division of Cardiology, Department of Medicine, University of California-San Francisco, San Francisco, California, USA; [b]Division of Hematology and Oncology, University of California-San Francisco, San Francisco, California, USA; [c]Cardiovascular Research Institute, University of California-San Francisco, San Francisco, California, USA; and the [d]Bakar Computational Health Sciences Institute, University of California-San Francisco, San Francisco, California, USA.

knowledge and practice. In this case, the investigators show that AI captured information from electrocardiograms to augment the ability to predict AF risk compared with using clinical variables in the Mayo CLL AF risk score alone. In multivariable analysis, both the AI model score and the highest risk categories of the Mayo CLL AF score remained independently and significantly associated with higher AF risk in the CLL population. The discriminative ability of either the ECG AI model or the Mayo CLL AF score alone had a C statistic of 0.62 to 0.66. The combination of both increased the C statistic to 0.71, though the 95% CIs were wide and overlapped, making it unclear if the combination provided a statistically significant increase. These results underscore several important points. First, neither AI analysis of the electrocardiogram nor clinical variables alone tell the entire story regarding future AF risk. This is underscored by the investigators' Figure 2B, which shows that even among those with high AI scores, clinical variables still provided powerful ability to discriminate future AF risk. Second, when used properly, AI can serve to augment (not replace) existing clinical knowledge and standard practice, as this analysis shows. Third, in the case of future AF risk for patients with CLL, there are risk factors that remain to be understood, beyond the clinical variables and the ECG, as the combined C statistic was 0.71. This study provides a valuable example of how AI analysis can expand the diagnostic utility of tests such as electrocardiography by providing complementary information from a data source (a sinus rhythm electrocardiogram) that clinicians would not easily be able to take into consideration otherwise.

A valuable clinical question to consider for any risk prediction model is whether something can be done differently on the basis of the information it provides. In this regard, CLL provides an example of a population for which prediction of increased AF risk may be clinically actionable. The covalent BTK inhibitor ibrutinib is highly effective in CLL, but it is associated with cardiovascular adverse events, including AF in up to 16% of patients.[5] Second-generation covalent BTK inhibitors such as acalabrutinib and zanubrutinib, as well as noncovalent BTK inhibitors such as pirtobrutinib, are associated with lower rates of AF,[6-8] but risk for AF and other cardiovascular adverse events appears to be a class effect. Although prior AF or risk for AF is not necessarily a contraindication to the use of BTK inhibitors, alternative therapies such as venetoclax-based regimens might

be preferred in such patients. Understanding the risk for AF and related complications is therefore important to adequately counsel patients prior to starting therapy for CLL. In this analysis, a higher baseline AI ECG score was associated with incidence of AF in the cohort of patients with newly diagnosed CLL, but this association was not seen on multivariable analysis among patients treated with BTK inhibitors. The investigators discuss several potential reasons for this, including the possibility that an algorithm trained using AF occurrence in the general population may not predict AF induced by BTK inhibitors.

This study begins to address several important considerations about applying the AF risk prediction AI model in real-world clinical settings and provides a valuable demonstration of how the AI model performs in the CLL population. However, several questions remain, including its utility in guiding therapy. Demonstrating the ability of AI ECG analysis to predict AF and other arrhythmias with use of BTK inhibitors would be critical for clinical application in this setting, which this study was not able to show. Another important consideration is model generalizability, which describes a model's ability to perform well on unseen data, such as patients from a different geographic region or institution, and is an important factor for robust model performance.[9] Because both the original AI model derivation[1] and this study[3] were performed at the same institution, it remains to be determined how this AI model would perform beyond the Mayo Clinic. The importance of external generalizability for any predictive model reemphasizes the urgent need for multicenter collaboration and data sharing to facilitate the development of robust and generalizable predictive models, AI or otherwise. It is also important to examine how this AI model can best be integrated into the clinical workflow to provide the AF risk information to clinicians and whether the chosen approach positively affects relevant clinical outcomes. Last, the clinical adoption of AI tools should be driven by clear and rigorous evidence, ideally in the form of randomized clinical trials, showing that the AI model and its method of clinical workflow integration leads to meaningfully improved outcomes for patients.

In summary, this study highlights an important example wherein AI can capture complementary information to augment clinical knowledge and practice and sets the stage for future research that examines how this complementary information can be integrated into clinical practice.

ADDRESS FOR CORRESPONDENCE: Dr Geoffrey H. Tison, Division of Cardiology, Department of Medicine, Bakar Computational Health Sciences Institute, University of California-San Francisco, San Francisco, California 94158, USA. E-mail: geoff.tison@ucsf.edu.

REFERENCES

1. Attia ZI, Noseworthy PA, Lopez-Jimenez F, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. *Lancet*. 2019;394:861–867.

2. Fukunami M, Yamada T, Ohmori M, et al. Detection of patients at risk for paroxysmal atrial fibrillation during sinus rhythm by P wave-triggered signal-averaged electrocardiogram. *Circulation*. 1991;83:162–169.

3. Christopoulos G, Attia ZI, Achenbach SJ, et al. Artificial intelligence electrocardiography to predict atrial fibrillation in patients with chronic lymphocytic leukemia. *J Am Coll Cardiol CardioOnc*. 2024;6(2):251–263.

4. Quartermaine C, Ghazi SM, Yasin A, et al. Cardiovascular toxicities of BTK inhibitors in chronic lymphocytic leukemia: *JACC: CardioOncology* state-of-the-art review. *J Am Coll Cardiol CardioOnc*. 2023;5:570–590.

5. Burger JA, Barr PM, Robak T, et al. Long-term efficacy and safety of first-line ibrutinib treatment for patients with CLL/SLL: 5 years of follow-up from the phase 3 RESONATE-2 study. *Leukemia*. 2020;34:787–798.

6. Byrd JC, Hillmen P, Ghia P, et al. Acalabrutinib versus ibrutinib in previously treated chronic lymphocytic leukemia: results of the first randomized phase III trial. *J Clin Oncol*. 2021;39:3441–3452.

7. Brown JR, Eichhorst B, Hillmen P, et al. Zanubrutinib or ibrutinib in relapsed or refractory chronic lymphocytic leukemia. *N Engl J Med*. 2023;388:319–332.

8. Mato AR, Woyach JA, Brown JR, et al. Pirtobrutinib after a covalent BTK inhibitor in chronic lymphocytic leukemia. *N Engl J Med*. 2023;389:33–44.

9. Koh PW, Sagawa S, Marklund H, et al. WILDS: a benchmark of in-the-wild distribution shifts. *Proc Mach Learn Res*. 2021;139:5637–5664.

KEY WORDS arrhythmia, hypertension, leukemia