Journal of the American Heart Association

EDITORIAL

Going Deep With ECG and Aortic Stenosis: Touchdown or Incomplete Pass?

Patrick A. Gladding, MBChB, PhD; Will Hewitt; Todd T. Schlegel, MD

A little learning is a dangerous thing; Drink deep, or taste not the Pierian spring.

Alexander Pope

Aortic stenosis (AS) is a growing problem in aging Western populations.[1] Many patients with AS are unaware of it and present late, often after irreversible left ventricular hypertrophy/fibrosis and/or diastolic or systolic dysfunction.[2] Whereas surgical guidelines currently dictate that either open heart surgery or transaortic valve replacement are reserved for those with clear symptoms, emerging evidence suggests that intervention may be preferred in some populations with severe asymptomatic AS.[3] With expanding indications and thresholds for transaortic valve replacement, screening and access to evidence-based therapies are increasingly important. Furthermore, with limited access to echocardiography, the main imaging modality for AS, there is a need to identify alternative diagnostic modalities that are portable and community focused.

See Article by Kwon et al.

In this issue of the *Journal of the American Heart Association* (*JAHA*), Kwon et al. took a first step in evaluating a deep learning (DL) algorithm with a view toward detecting moderate-to-severe AS via ECG.[4] Their optimized DL approach employed both a multilayer perceptron and a convolutional neural network as an ensemble to interrogate patterns within 12-lead and single-lead ECGs from 43 051 patients, of whom 1413 (3.3%) had > moderate AS. Patients in the derivation data set were all-comers to a cardiovascular teaching hospital. Those with AS were more often female, with notably low mean body mass index (24 kg/m^2) in a Korean population, and often with other structural and functional heart abnormalities such as altered systolic and diastolic function, left ventricular (LV) mass, and pulmonary pressure. While high-quality (500 Hz) ECG data were used, just 7 basic measures of strictly conventional ECG were employed, along with demographic factors, to build and test the multilayer perceptron as well as statistical pattern recognition models including logistic regression. At the same time, state-of-the-art signal processing was used to generate the convolutional neural network. Standard statistics (area under the receiver operating characteristic curve) were then used to evaluate and internally validate the trained models in the same hospital and in 10 865 patients from a second, community-based hospital. Results were impressive with an AUC of 0.86 (sensitivity 80% and specificity 78%) in the external validation group. In fact, even a single ECG lead had an AUC of 0.82.

It is worth noting that despite much recent excitement regarding artificial intelligence applications to ECG, there are still very few studies of this kind, especially performed at this scale and across 2 sites with true external validation. Difficulties in securing high-quality, well-annotated data, as well as sufficiently

Key Words: Editorials ■ ECG ■ ECG criteria ■ ECG screening

Correspondence to: Todd T. Schlegel, MD, Department of Clinical Physiology, Karolinska Institutet, Stockholm, N/A SE-171 76 Sweden. E-mail: ttschlegel@gmail.com

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.

For Disclosures, see page 3.

JAHA is available at: www.ahajournals.org/journal/jaha

sophisticated data curation and handling skills to ensure proper predictions from large sets of often biased data, likely contribute to the paucity of such studies. Nonetheless, Kwon et al. have shown that with a robust DL pipeline, it seems possible to use digital ECG to reasonably accurately predict AS in ways that few if any individual clinicians could.

Utilizing a variety of techniques, several researchers have also recently shown that other cardiac pathologies such as LV systolic dysfunction,[5–7] prior atrial fibrillation,[8] LV hypertrophy[9] (ideally better redefined electrically as "LV electrical remodeling"),[10] and hypertrophic cardiomyopathy[11,12] can also be predicted from 12-lead ECGs. So with some high-quality studies of this kind showing promising results, 1 question to be asked is "why have these techniques, especially the DL techniques that have elicited the most excitement, not yet entered clinical practice"? One potential reason is that there are even fewer prospective, randomized studies of DL-derived tools that demonstrate improved healthcare resource utilization, cost, or patient outcomes.[13] Furthermore, while Kwon et al. and others have made attempts to demonstrate specific predictive features of DL algorithms (eg, via sensitivity [saliency] mapping or "attention heatmaps"), such secondary procedures do not rid the DL techniques of their obscure, "black box"-type features.[14] The multiple abstractions and convolutions within the types of DL applied by Kwon et al. and others often defy full transparency or explainability, 2 elements considered essential by some for the ethical use of artificial intelligence.[15,16] So without these elements, the potential for bias, including ethnic bias or the entrenched bias present within any data set used for training, is quite real. For example, it has been shown that ECG features vary by both race and sex.[17] Also, although DL models for some conditions such as LV systolic dysfunction have thus far translated fairly well to other ethnicities, the model used by Kwon et al. for AS must be validated not only in other ethnic groups, but also in other clinical groups who will undoubtedly have different types and/or degrees of cardiac comorbidities, as well as different types and levels of ECG background noise, ECG sampling rates, etc.

Moreover, although DL has increasingly "come of age," it must also be understood that more trusted and fully transparent tools such as logistic regression demonstrate AUC results that are not inferior to, and sometimes superior to, those of the DL techniques,[18] something that we suspect might have also occurred in the study of Kwon et al. had more powerful advanced ECG parameters been derived and carefully "feature selected" into their logistic regression procedures.[7] Moreover, other statistical pattern recognition techniques such as advanced ECG-related discriminant analysis can also often add critical discriminatory power for separating 1 cardiac pathology from another, something that until proven otherwise, DL techniques such as that of Kwon et al. might have comparative difficulty accomplishing. With rigorous derivation and careful study of multiple discrete advanced and conventional ECG measures together, it is usually possible to readily identify the key features that serve as the main predictors within one's logistic regression, discriminant analysis, or other forms of ECG-based statistical pattern recognition. These other techniques can therefore allow one to "discard the black box," or, at a minimum, to open up one's black box for appropriate scrutiny and transparency. Also, the careful use, even alone, of powerful advanced ECG features now easily derived from any 12-lead ECG (eg, the spatial QRST angle),[19] is also more practical and potentially universally applicable for the arguably most important task of predicting clinical outcomes. Moreover, discrete features of advanced ECG have also recently demonstrated genetic associations in genomewide association studies, with potential pharmacogenomic implications.[20]

A final issue worth noting before clinical implementation of DL-type techniques is what the implications of the delivery of distributable systems would be when magnified to scale. Here basic epidemiological statistics and modeling might help. For a condition with prevalence of 3.3%, and a diagnostic tool with sensitivity 80% and specificity 78%, the likelihood of *any* test being positive is 23%. For any positive result, the likelihood that it is a true positive is only 11%, with the vast majority of positive calls being false. Thus, blindly following such an artificial intelligence result with a referral to echocardiography would rapidly swamp any system's ability to cope. However, careful use within a population with a high prevalence of the disease would be a different story. For example, use in a murmur clinic, or in conjunction with point of care ultrasound or phonocardiography or biomarkers, could also provide a means of further improving diagnostic performance.

In conclusion, the study by Kwon et al. provides a noteworthy advance within the ECG field. At the same time, however, excitement about the potential for clinical use of ECG-based DL algorithms must tempered by the recognition of their limitations, obscuration, and need for further validation, as well as the existence of equally accurate techniques that are possibly more trustworthy, understandable, and ethical.

ARTICLE INFORMATION

Affiliations

From the Cardiology Department, Waitemata District Health Board, Auckland, New Zealand (P.A.G.); Auckland Bioengineering Institute,

Auckland, New Zealand (P.A.G., W.H.); Department of Clinical Physiology, Karolinska Institutet, Stockholm, Sweden (T.T.S.); Nicollier-Schlegel Sàrl, Trélex, Switzerland (T.T.S.).

Disclosures

Dr Schlegel is a principal of Nicollier-Schlegel SARL, Trélex, Switzerland, a company that performs clinical and research-related advanced ECG consultancy. The remaining authors have no disclosures to report.

REFERENCES

1. Thoenes M, Bramlage P, Zamorano P, Messika-Zeitoun D, Wendt D, Kasel M, Kurucova J, Steeds RP. Patient screening for early detection of aortic stenosis (AS)-review of current practice and future perspectives. *J Thorac Dis*. 2018;10:5584–5594.
2. Everett RJ, Clavel M-A, Pibarot P, Dweck MR. Timing of intervention in aortic stenosis: a review of current and future strategies. *Heart*. 2018;104:2067.
3. Kang D-H, Park S-J, Lee S-A, Lee S, Kim D-H, Kim H-K, Yun S-C, Hong G-R, Song J-M, Chung C-H, et al. Early surgery or conservative care for asymptomatic aortic stenosis. *N Engl J Med*. 2019;382:111–119.
4. Kwon J-M, Lee SY, Jeon K-H, Lee Y, Kim K-H, Park J, Oh B-H, Lee M-M. Deep learning based algorithm for detecting aortic stenosis using electrocardiography. *J Am Heart Assoc*. 2020;9:e014717. DOI: 10.1161/JAHA.119.014717.
5. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM, et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nat Med*. 2019;25:70–74.
6. Kwon JM, Kim KH, Jeon KH, Kim HM, Kim MJ, Lim SM, Song PS, Park J, Choi RK, Oh BH. Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification. *Korean Circ J*. 2019;49:629–639.
7. Schlegel TT, Kulecz WB, Feiveson AH, Greco EC, DePalma JL, Starc V, Vrtovec B, Rahman MA, Bungo MW, Hayat MJ, et al. Accuracy of advanced versus strictly conventional 12-lead ECG for detection and screening of coronary artery disease, left ventricular hypertrophy and left ventricular systolic dysfunction. *BMC Cardiovasc Disord*. 2010;10:28.
8. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, Carter RE, Yao X, Rabinstein AA, Erickson BJ, 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.
9. Kwon JM, Jeon KH, Kim HM, Kim MJ, Lim SM, Kim KH, Song PS, Park J, Choi RK, Oh BH. Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography. *Europace*. 2020;22:412–419.
10. Bacharova L, Estes HE, Schocken DD, Ugander M, Soliman EZ, Hill JA, Bang LE, Schlegel TT. The 4th report of the working group on ECG diagnosis of left ventricular hypertrophy. *J Electrocardiol*. 2017;50:11–15.
11. Ko WY, Siontis KC, Attia ZI, Carter RE, Kapa S, Ommen SR, Demuth SJ, Ackerman MJ, Gersh BJ, Arruda-Olson AM, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J Am Coll Cardiol*. 2020;75:722–733.
12. Potter SLP, Holmqvist F, Platonov PG, Steding K, Arheden H, Pahlm O, Starc V, McKenna WJ, Schlegel TT. Detection of hypertrophic cardiomyopathy is improved when using advanced rather than strictly conventional 12-lead electrocardiogram. *J Electrocardiol*. 2010;43:713–718.
13. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44–56.
14. The Lancet Respiratory M. Opening the black box of machine learning. *Lancet Respir Med*. 2018;6:801.
15. Coeckelbergh M. Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Sci Eng Ethics*. 2019. Available at: https://link.springer.com/article/10.1007%2Fs11948-019-00146-8.
16. London AJ. Artificial intelligence and black-box medical decisions: accuracy versus explainability. *Hastings Cent Rep*. 2019;49:15–21.
17. Noseworthy PA, Attia ZI, Brewer LC, Hayes SN, Yao X, Kapa S, Friedman PA, Lopez-Jimenez F. Assessing and mitigating bias in medical artificial intelligence: the effects of race and ethnicity on a deep learning model for ECG analysis. *Circ Arrhythm Electrophysiol*. 2020. Available at: https://www.ncbi.nlm.nih.gov/pubmed/32064914. DOI: 10.1161/CIRCEP.119.007988. [Epub ahead of print].
18. 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.
19. Oehler A, Feldman T, Henrikson CA, Tereshchenko LG. QRS-T angle: a review. *Ann Noninvasive Electrocardiol*. 2014;19:534–542.
20. Tereshchenko LG, Sotoodehnia N, Sitlani CM, Ashar FN, Kabir M, Biggs ML, Morley MP, Waks JW, Soliman EZ, Buxton AE, et al. Genome-wide associations of global electrical heterogeneity ECG phenotype: the ARIC (Atherosclerosis Risk in Communities) Study and CHS (Cardiovascular Health Study). *J Am Heart Assoc*. 2018;7:e008160. DOI: 10.1161/JAHA.117.008160.