JACC: CARDIOVASCULAR IMAGING VOL. ■, NO. ■, 2021

EDITORIAL COMMENT

Machine Learning for ECG Diagnosis of LV Dysfunction*

Rhodri H. Davies, PhD,[a,b] Aled R. Jones, MBBS[c,d]

Since the first report of the human electrocardiogram (ECG) by Augustus Waller in 1887 (1), its diagnostic use has continually expanded. Providing a window to the electric activity of the heart, the ECG allowed the documentation and classification of arrhythmias, soon followed by estimation of atrial and ventricular size (2). Some years later, dynamic changes associated with myocardial infarction were discovered and shown to correlate with clinical and histopathologic changes (2). Researchers have continued to extract a diversifying range of clinically useful data from this simple bedside test, demonstrating the wealth of information contained within it and making the ECG ubiquitous in clinical medicine.

Efforts to glean yet more information from the surface electrocardiogram have recently led to the application of machine learning methods to ECG analysis. Although the first attempt to automate ECG analysis was 50 years ago (3), the recent resurgence of machine learning has accelerated efforts to automatically recognize, measure, and characterize patterns not visible to the human eye. Examples include identifying patients with paroxysmal atrial fibrillation from a sinus rhythm ECG, as well as accurately predicting age and sex from a baseline trace (4). Furthermore, several algorithms for ECG diagnosis of left ventricular (LV) dysfunction (LVD) have recently been reported (5,6), with the U.S. Food and Drug Administration granting emergency authorization for such use (7).

In this issue of *iJACC*, Potter et al (8) develop a machine learning algorithm to identify LVD from 12-lead ECGs and evaluate its use as a screening tool for identifying at-risk asymptomatic individuals. The authors use machine learning of preprocessed ECG data to identify LVD, as defined by echocardiography. Proprietary software calculates a large set of parameters from the processed ECG signals, and these are used to teach ("train") a machine learning model to distinguish patients with LVD from those without. Both the signal processing (continuous wavelet transform) and machine learning (random forest) methods used in the work are well established and have been validated in many applications.

The definition of LVD used by Potter et al (8) is based on a combination of diastolic echocardiogram parameters and global longitudinal strain and deviates from standard criteria. This is not universal across similar studies. Attia et al (6), eg, use a single criterion of LV ejection fraction of <35%, which has a stronger evidence base for initializing treatment but will fail to identify milder cases or patients in the early phases of LV dysfunction. Concentrating on severe LV systolic dysfunction will identify patients with a more pronounced phenotype, likely translating to more marked ECG changes and simplifying the task of a classification algorithm; the definition of disease must be borne in mind when comparing the reported performance of these studies.

Central to any machine learning method is data: the training data on which the model is taught ("trained") and the independent test data on which it is evaluated. The training data are a key determinant of how well a model generalizes to new data and, ultimately, how well it performs. The goal is to learn underlying patterns and avoid overfitting to noise and spurious trends present in the training data. The training data should contain enough information to represent all intrinsic variation in the population to which the model will ultimately be applied. It is

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

From the [a]University College London, London, United Kingdom; [b]University Hospital of Wales, Cardiff, United Kingdom; [c]Queen Mary University London, London, United Kingdom; and [d]St. Bartholomew's Hospital, London, United Kingdom.

Alan Fraser, BSc, MB ChB, served as the Guest Editor for this paper.

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.

ISSN 1936-878X/$36.00 https://doi.org/10.1016/j.jcmg.2021.05.015

therefore likely that the model described by Potter et al (8) would benefit from a significant expansion of the training data set to allow extrapolation of the model to the multimillion patient screening program proposed. Collecting and cleaning such data can be a time-consuming and tedious process, and although many health record archives are in formats that lend themselves to convenient data scraping, there are many ethical and regulatory issues that must be considered. Expansion of the training data would permit the use of more data-hungry machine learning methods, particularly deep learning using convolutional neural networks, which has been the catalyst behind the recent step change in machine learning performance (and interest) (9).

Robust validation of a model on an independent data set is an essential step before clinical deployment. A noticeable trend in the data described by Potter et al (8) is the significant difference in clinical characteristics between training and test sets (see Table 2 of the paper), with the latter containing significantly fewer patients with the disease of interest (LVD). This is the likely cause for better model performance on the test set compared to the training data, which is the opposite of what is usually seen. This highlights the need to interpret model performance with respect to the population to which it is applied.

A frequent criticism of machine learning techniques applied to clinical data is the lack of transparency into how they make decisions. This is important in clinical research, where new scientific insights are sought from data, but also vital to ensure buy-in from clinicians and patients. In terms of clinical use, however, sound validation is arguably of greater importance. If the performance parameters of an algorithm (however it functions) are well defined, its role in clinical decision making can be made clear. Potter et al (8) are therefore right to emphasize the use of their algorithm in the context of screening, where its true benefit likely lies, highlighting asymptomatic individuals who are at an increased chance of benefiting from echocardiography.

A large-scale screening program for LVD based on current clinical and biochemical tests would result in an unacceptable number of false positives and false negatives. Potter et al (8) have shown the potential for delivering acceptable performance by integrating advanced analysis of the standard 12-lead ECG—a cheap, portable, and reproducible investigation. Although more work is required to translate this into a standard clinical tool, the evidence that the ECG contains useful information on LV function continues to grow.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

The authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Rhodri H. Davies, University College London, Institute of Cardiovascular Sciences, Gower Street, London WC1E 6BT, United Kingdom. E-mail: rhodri.davies@ucl.ac.uk.

REFERENCES

1. Waller AD. A demonstration on man of electromotive changes accompanying the heart's beat. J Physiol 1887;8(5):229–34.

2. Fye WB. A history of the origin, evolution, and impact of electrocardiography. Am J Cardiol 1994;73(13):937–49.

3. Pipberger HV, Arms RJ, Stallmann FW. Automatic screening of normal and abnormal electrocardiograms by means of digital electronic computer. Proc Soc Exp Biol Med 1961;106(1):130–2.

4. Somani S, Russak AJ, Richter F, et al. Deep learning and the electrocardiogram: review of the current state-of-the-art. *Europace* Feb 10, 2021 [E-pub ahead of print]

5. Attia ZI, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. Nat Med 2019;25(1):70–4.

6. Kwon J, Kim K-H, Jeon K-H, et al. Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification. Korean Circ J 2019;49(7):629–39.

7. US Food and Drug Administration. Coronavirus (COVID-19) update: daily roundup May 12, 2020. FDA. Published May 12, 2020, https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-daily-roundup-may-12-2020. Accessed April 28, 2021.

8. Potter EL, Rodrigues CHM, Ascher DB, et al. Machine learning of ECG waveforms to improve selection for testing for asymptomatic left ventricular dysfunction. J Am Coll Cardiol Img 2021;14:XXX-XXX.

9. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.

KEY WORDS ECG, left ventricular dysfunction, machine learning