Predicting heart failure with preserved ejection fraction: revisiting an old friend with new knowledge

Robert M. A. van der Boon

Department of Cardiology, Erasmus Medical Center Rotterdam, 's-Gravendijkwal 230, 3015 CE, Rotterdam, the Netherlands

This editorial refers to 'Artificial intelligence assessment for early detection of heart failure with preserved ejection fraction based on electrocardiographic features' by J. Kwon et al., on page 106.

Introduction

Heart failure (HF) is a growing health problem affecting millions of people worldwide. Heart failure with preserved ejection fraction (HFpEF) accounts for at least 50% of HF diagnoses and its prevalence is expected to increase due to an aging population and a growing burden of obesity, hypertension, and diabetes mellitus.[1] Although HFpEF and HF with reduced ejection fraction (HFrEF) share a common phenotypic spectrum of symptoms, clinical signs, and poor health status, they are distinct entities.[2] The heterogeneity and poorly understood pathophysiology of HFpEF makes it a complex syndrome with to date no therapies demonstrating clear benefits in large trials.[3] Moreover, adverse events related to HFpEF have been steadily increasing which accounts for substantial morbidity, mortality, and increasing health care cost.[4,5] Hence, the need for new techniques to define and diagnose this population with the possibility of precision treatment and improving prognosis. Machine learning (ML) is everywhere around us and has made its way into clinical practice with emerging applications to interpret echocardiography, radiologic images, and pathology slides. Yet, the earliest form of ML within cardiology is the machine read an electrocardiogram (ECG) over 40 years ago. Where at the time accuracy was low, current neural networks achieve a high diagnostic performance outperforming cardiology residents.[6] Therefore, it is not surprising that we are revisiting this old friend with new knowledge.

Discussion

In this issue of *European Heart Journal – Digital Health*, Kwon *et al.* propose an ensemble neural network-based deep learning model (DLM) combining electrocardiographic data and four clinical variables to predict HFpEF.[7] A total of 22 148 patients from a Korean hospital were retrospectively included of whom data of 20 169 patients were used to derive the DLM. Internal validation was performed in the remaining 1979 patients and external validation in 11 955 patients from a second institution. All patients underwent an ECG and echocardiogram, and left ventricular diastolic dysfunction (LVDD) was defined according to current guidelines.[3] Cases with LVDD in the presence of a (near) normal ejection fraction with symptoms or signs of heart failure were deemed as having HFpEF. The algorithm resulted in an area under the curve (AUC) of 0.87 [95% confidence interval (CI) 0.86–0.88], sensitivity of 0.77 (95% CI 0.75–0.79), and specificity of 0.82 (95% CI 0.81–0.83) in the external validation cohort, results that are similar to a recently published clinical risk score.[8] Additionally, in patients initially classified by the DLM as having HFpEF without evidence of LVDD by echocardiographic assessment (false positives), a significant proportion of 'high-risk' patients eventually developed HFpEF during follow-up (33.6% vs. 8.4%, $P < 0.001$).

Overall, the study by Kwon *et al.* confirms the feasibility of a DLM to discriminate between patients with and without HFpEF and highlights the possibility of screening and early diagnosis by means of the ECG. In contrast to available risk scores, the strength of this approach lies within the utilization of an inexpensive and almost universal available tool in combination with four patient characteristics to predict a prevalent disease. Their analysis regarding the 'false positives' suggests that subtle ECG abnormalities may identify early stages of HFpEF when overt disease is not yet present, a phenomenon also described by Attia *et al.*[9] in their work on

HFrEF. Furthermore, by applying guided gradient backpropagation and a sensitivity map analysis the authors visualize the electrocardiographic regions(R-wave and T-wave axis) that significantly affected the DLM. Not only does this help to unravel the 'black box' myth surrounding ML, it also expands our knowledge on subtle disease-specific electrocardiographic features improving ECG interpretation in daily practice.[10,11]

Although showing great promise, we should be careful before applying the presented algorithm in clinical practice. This DLM was developed upon a retrospective analysis which increases the possibility of incidental findings and correlations. The authors already noted this limitation that will be addressed in a prospective study designed to reproduce the current findings. More importantly, defining HFpEF is challenging and only combining the presence of LVDD with signs or symptoms of HF might be an oversimplification of a complex disease. This raises the question if the DLM truly predicts HFpEF or LVDD, a problem that should be addressed by using the HFA-PEFF diagnostic algorithm in future studies.[12] By using the Youden J-statistic sensitivity and specificity were optimized leading to a very good diagnostic accuracy (AUC) of the DLM. However, in the case of a screening test one might be more interested in maximizing either the sensitivity or the specificity in order to effectively rule-out or rule-in disease, respectively. Moreover, the trade-off between these variables introduces false positives or false negatives leading to unnecessary testing or false reassurance of patients. These problems are not apparent when focusing on the diagnostic accuracy of a model but become important when regarded from a clinical and economical perspective. Last, DLMs may demonstrate a lower accuracy when used beyond their training and validation data. Consequently, extensive external validation of the model is necessary prior to clinical implementation. This process should not only be repeated in a different institution, as was the case here, but also in a different geographic location (beyond Korea) to provide evidence of its applicability. Multicentre collaboration or open data sharing are necessary to facilitate this on a large scale.

Despite these challenges, the findings by Kwon *et al.* offer a glimpse into a future in which we are able to identify those at high risk of developing HFpEF in a simple but efficient way. Perhaps, this may even be applicable in regions with limited cardiovascular care with the use of wearables or lifestyle devices. Early recognition of these patients can initiate lifestyle intervention and treatment of known risk factors. The latter may lead to improved quality of life, significant reductions in HF associated adverse events and reduced costs benefiting patients, physicians, and healthcare systems. Also, subclassification of patients with HFpEF using deep phenotyping may help to approach this heterogeneous disease and selectively direct individual patients towards clinical trials that match their underlying pathophysiologic features. A strategy eventually leading to effective patient management.[13,14] Additionally, follow-up of untreated patients might help delineate the complex natural course of this multisystem disease. Machine learning is becoming increasingly important in clinical medicine and medical research, yet we must remember to have a healthy amount of scepticism when interpreting its results. Hopefully, coupled with new imaging technology and molecular analysis, machine learning will assist in a more accurate diagnosis,

targeted and efficacious treatment, and a true personalized approach for HFpEF.

References

1. van Riet EE, Hoes AW, Wagenaar KP, Limburg A, Landman MA and Rutten FH. Epidemiology of heart failure: the prevalence of heart failure and ventricular dysfunction in older adults over time. A systematic review. *Eur J Heart Fail* 2016;**18**:242–52.

2. Borlaug BA and Redfield MM. Diastolic and systolic heart failure are distinct phenotypes within the heart failure spectrum. *Circulation* 2011;**123**:2006–13; discussion 2014.

3. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, Falk V, González-Juanatey JR, Harjola V-P, Jankowska EA, Jessup M, Linde C, Nihoyannopoulos P, Parissis JT, Pieske B, Riley JP, Rosano GMC, Ruilope LM, Ruschitzka F, Rutten FH, van der Meer P; ESC Scientific Document Group. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J* 2016;**37**:2129–2200.

4. Dunlay SM, Shah ND, Shi Q, Morlan B, VanHouten H, Long KH and Roger VL. Lifetime costs of medical care after heart failure diagnosis. *Circ Cardiovasc Qual Outcomes* 2011;**4**:68–75.

5. Bursi F, Weston SA, Redfield MM, Jacobsen SJ, Pakhomov S, Nkomo VT, Meverden RA and Roger VL. Systolic and diastolic heart failure in the community. *JAMA* 2006;**296**:2209–16.

6. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP and Ng AY. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019;**25**:65–69.

7. Kwon J, Kim K-H, Eisen HJ, Cho Y, Jeon K-H, Lee SY, Park J and Oh B-H. Artificial intelligence assessment for early detection of heart failure with preserved ejection fraction based on electrocardiographic features. *Eur Heart J - Digital Health* 2021;**2**:106–116.

8. Gohar A, Kievit RF, Valstar GB, Hoes AW, Van Riet EE, van Mourik Y, Bertens LC, Boonman-Winter LJ, Bots ML, Den Ruijter HM and Rutten FH. Opportunistic screening models for high-risk men and women to detect diastolic dysfunction and heart failure with preserved ejection fraction in the community. *Eur J Prev Cardiol* 2019;**26**:613–623.

9. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM, Asirvatham SJ, Scott CG, Carter RE and Friedman PA. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med* 2019;**25**:70–74.

10. van de Leur R, Taha K, Bos MN, van der Heijden JF, Gupta D, Cramer MJ, Hassink RJ, van der Harst P, Doevendans PA, Asselbergs FW and van Es R. Discovering and visualizing disease-specific electrocardiogram features using deep learning: proof-of-concept in phospholamban gene mutation carriers. *Circ Arrhythm Electrophysiol* 2021. Doi: 10.1161/CIRCEP.120.009056.

11. Bleijendaal H, Ramos LA, Lopes RR, Verstraelen TE, Baalman SWE, Oudkerk Pool MD, Tjong FVY, Melgarejo-Meseguer FM, Gimeno-Blanes FJ, Gimeno-Blanes JR, Amin AS, Winter MM, Marquering HA, Kok WEM, Zwinderman AH, Wilde AAM and Pinto YM. Computer versus cardiologist: is a machine learning algorithm able to outperform an expert in diagnosing a phospholamban p.Arg14del mutation on the electrocardiogram? *Heart Rhythm* 2021;**18**:79–87.

12. Pieske B, Tschöpe C, de Boer RA, Fraser AG, Anker SD, Donal E, Edelmann F, Fu M, Guazzi M, Lam CSP, Lancellotti P, Melenovsky V, Morris DA, Nagel E, Pieske-Kraigher E, Ponikowski P, Solomon SD, Vasan RS, Rutten FH, Voors AA, Ruschitzka F, Paulus WJ, Seferovic P and Filippatos G. How to diagnose heart failure with preserved ejection fraction: the HFA-PEFF diagnostic algorithm: a consensus recommendation from the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). *Eur Heart J* 2019;**40**:3297–3317.

13. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, Bonow RO, Huang CC and Deo RC. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation* 2015;**131**:269–79.

14. Shah SJ, Borlaug BA, Kitzman DW, McCulloch AD, Blaxall BC, Agarwal R, Chirinos JA, Collins S, Deo RC, Gladwin MT, Granzier H, Hummel SL, Kass DA, Redfield MM, Sam F, Wang TJ, Desvigne-Nickens P and Adhikari BB. Research priorities for heart failure with preserved ejection fraction: national heart, lung, and blood institute working group summary. *Circulation* 2020;**141**:1001–1026.