Deep learning detects heart failure with preserved ejection fraction using a baseline electrocardiogram

Article *in* European Heart Journal - Digital Health · September 2021
DOI: 10.1093/ehjdh/ztab081

European Heart Journal Digital Health
doi:10.1093/ehjdh/ztab081

SHORT REPORT

Deep learning detects heart failure with preserved ejection fraction using a baseline electrocardiogram

Matthias Unterhuber [1]*, Karl-Philipp Rommel [1], Karl-Patrik Kresoja [1], Julia Lurz[2], Jelena Kornej[3,4], Gerhard Hindricks[2,5], Markus Scholz [4,6], Holger Thiele [1,5], and Philipp Lurz[1,5]

[1]Department of Cardiology, Heart Center Leipzig at University Leipzig, Strümpellstraße 39, 04289 Leipzig, Germany; [2]Department of Electrophysiology, Heart Center Leipzig at University Leipzig, Strümpellstraße 39, 04289 Leipzig, Germany; [3]School of Medicine, Cardiovascular Medicine, Boston University School of Medicine, 72 E Concord St, Boston, MA 02118; [4]LIFE Research Center of Civilization Diseases, Philipp-Rosenthal-Straße, 27 04103 Leipzig, Germany; [5]Leipzig Heart Institute, Russenstraße 69A, 04289 Leipzig, Germany; and [6]Institute for Medical Informatics, Statistics and Epidemiology, Medical Faculty, Universität Leipzig Härtelstraße, 16-18 04107 Leipzig, Germany

Received 28 May 2021; revised 23 August 2021; editorial decision 9 September 2021accepted 14 September 2021; online publish-ahead-of-print 17 September 2021

Aims

Heart failure with preserved ejection fraction (HFpEF) is a rapidly growing global health problem. To date, diagnosis of HFpEF is based on clinical, invasive, and laboratory examinations. Electrocardiographic findings may vary, and there are no known typical ECG features for HFpEF.

Methods and results

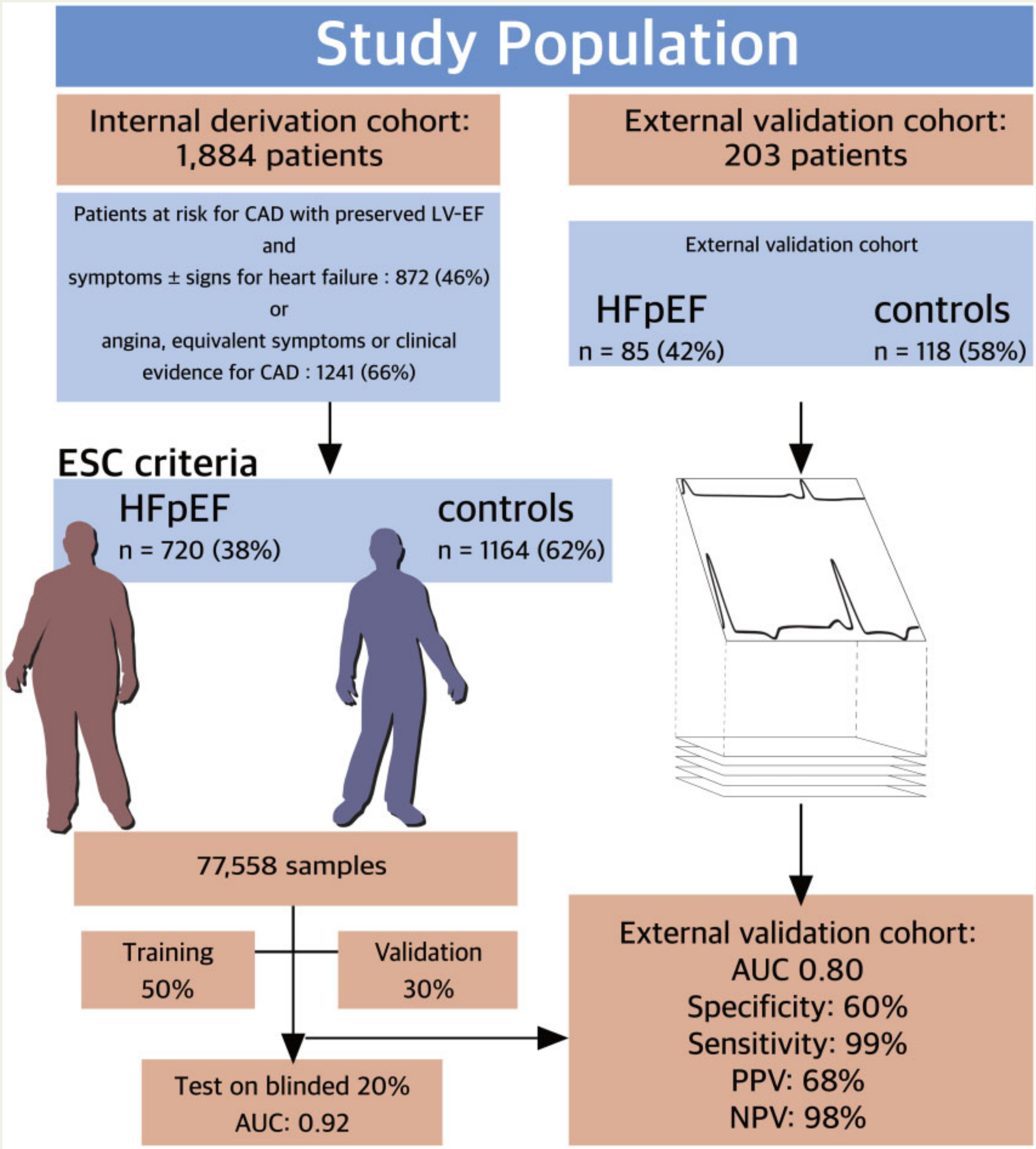
This study included two patient cohorts. In the derivation cohort, we included $n = 1884$ patients who presented with exertional dyspnoea or equivalent and preserved ejection fraction (≥50%) and clinical suspicion for coronary artery disease. The ECGs were divided in segments, yielding a total of 77 558 samples. We trained a convolutional neural network (CNN) to classify HFpEF and control patients according to European Society of Cardiology (ESC) criteria. An external group of 203 volunteers in a prospective heart failure screening programme served as a validation cohort of the CNN. The external validation of the CNN yielded an area under the curve of 0.80 [95% confidence interval (CI) 0.74–0.86] for detection of HFpEF according to ESC criteria, with a sensitivity of 0.99 (95% CI 0.98–0.99) and a specificity of 0.60 (95% CI 0.56–0.64), with a positive predictive value of 0.68 (95%CI 0.64–0.72) and a negative predictive value of 0.98 (95% CI 0.95–0.99).

Conclusion

In this study, we report the first deep learning-enabled CNN for identifying patients with HFpEF according to ESC criteria including NT-proBNP measurements in the diagnostic algorithm among patients at risk. The suitability of the CNN was validated on an external validation cohort of patients at risk for developing heart failure, showing a convincing screening performance.

* Corresponding author. Tel: +49-341-865-1428, Fax: +49-341-865-1461, Email: Matthias.Unterhuber@medizin.uni-leipzig.de

Graphical Abstract



Overview of the study design, model construction, and results. AUC, area under the curve; CAD, coronary artery disease; HFpEF, heart failure with preserved ejection fraction; LV-EF, left ventricular ejection fraction; NPV, negative predictive value; PPV, positive predictive value.

Keywords Artificial intelligence • Electrocardiogram • Heart failure with preserved ejection fraction

Introduction

Heart failure with preserved ejection fraction (HFpEF) is one of the most frequent cardiac causes of exertional dyspnoea. The reference standard for the diagnosis of HFpEF is an invasive workup with right-heart catheterization.[1] Many approaches have been developed for guiding non-invasive diagnostic pathways in HFpEF.[2,3] However, the current guideline definition of signs/symptoms of HFpEF combined with natriuretic peptides as well as structural and functional alterations on echocardiography is still valid.[1] In HFpEF, electrocardiographic findings may vary from a normal ECG to overt atrial and/or ventricular conduction delays which are recognized in various diagnostic algorithms, nonetheless, there are no unambiguous features that allow an accurate ECG diagnosis of HFpEF.[1,2]

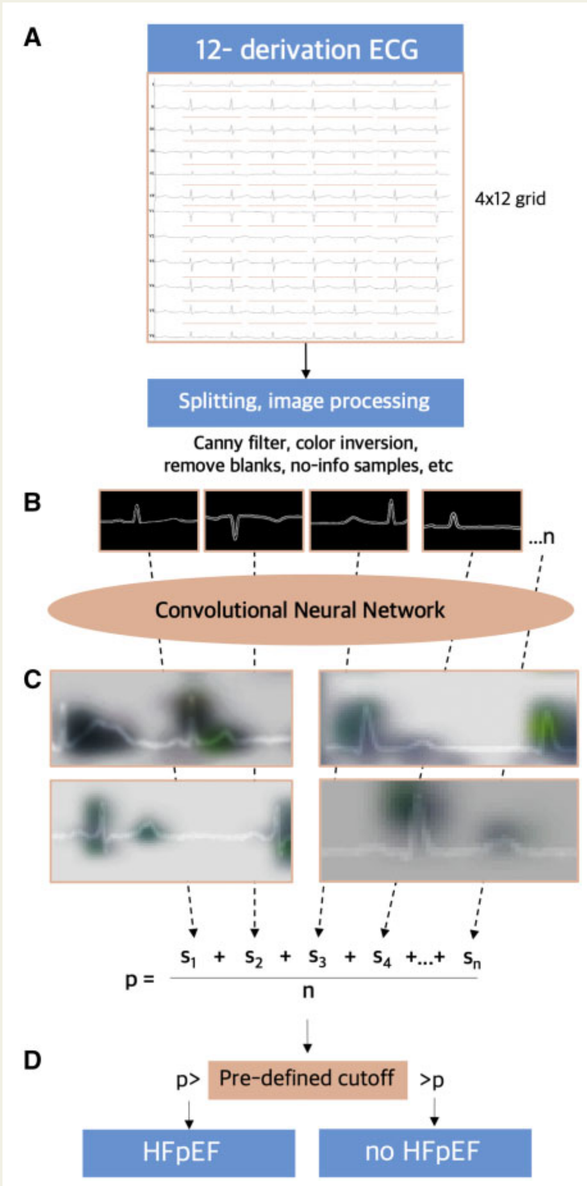


Figure 1 Image processing and classification steps of the algorithm. (*A*) After splitting into a 4 × 12 grid with 2-s segments for each derivation. (*B*) Filters are applied to invert the colours, sharpen edges of the tracings, and recognize edges with canny filters. (*C*) After pre-processing, each segment is passed into the convolutional neural network, yielding a probability of heart failure with preserved ejection fraction (s). The heatmaps are extracted from the network showing the main activation regions used for image classification, mainly QRS and ST-segments. The higher colour saturation, the more the region activates the networks' classifying confidence. (*D*) An arithmetic mean is computed across the calculated probabilities and checked against a pre-defined cut-off value from the derivation dataset, ultimately classifying an ECG as heart failure with preserved ejection fraction or not heart failure with preserved ejection fraction. HFpEF, heart failure with preserved ejection fraction; p, arithmetic mean of HFpEF probability; S, probability of HFpEF class according to sigmoid activation function.

Table 1 Baseline demographic, clinical, and echocardiographic characteristics of patients classified by the convolutional neural network

	CNN training cohort according to ESC criteria		
	Classified as HFpEF *N* = 720	Classified as no HFpEF *N* = 1164	*P*-value
Age, year	66 ± 10	59 ± 10	<0.001
Female gender, *n* (%)	330 (46)	418 (36)	<0.001
BMI, kg/m^2	31 ± 5	30 ± 5	<0.001
E/E′ over 12, *n* (%)	233 (32)	115 (10)	<0.001
LAEDVI, mL/m^2	29 ± 10	25 ± 8	0.005
NT-proBNP, ng/L	282 (178–545)	56 (34–86)	<0.001
LV mass index, g/m^2	128 (108–157)	116 (95–139)	<0.001
	External validation cohort		
	Classified as HFpEF *N* = 94	Classified as no HFpEF *N* = 109	*P*-value
Age, year	74 ± 8	71 ± 12	0.065
Female gender, *n* (%)	28 (29)	40 (36)	0.373
BMI, kg/m^2	28 (25–31)	29 (26–31)	0.368
E/E′ over 12, *n* (%)	24 (25)	12 (11)	0.010
LAEDVI, mL/m^2	34 (28–39)	27 (24–31)	<0.001
NT-proBNP, ng/L	186 (140–342)	94 (71–136)	<0.001
LV mass index, g/m^2	130 (114–152)	122 (108–138)	0.026

BMI, body mass index; CNN, convolutional neural network; ESC, European Society of Cardiology; HFpEF, heart failure with preserved ejection fraction; LAEDVI, left atrial end-diastolic volume index; LV, left ventricle.

Artificial intelligence gained attention in the last decade as ECG-enabled deep learning algorithms (DLA) and convolutional neural networks (CNNs) are able to detect a manifold of conditions.[4–7] However, these studies assessed echocardiographic features for diastolic dysfunction without assessing or reporting NT-proBNP, despite being a surrogate of increased wall stress, a hallmark of HFpEF diagnosis.[1] This study sought to evaluate whether a DLA can detect the diagnosis of HFpEF according to the current European Society of Cardiology (ESC) guidelines, including echocardiographic alterations, as well as increased natriuretic peptides, from baseline 12-lead ECGs.

Methods

We included 1884 patients who presented with exertional dyspnoea or equivalent and preserved ejection fraction (≥50%) with clinical suspicion for coronary artery disease (CAD) in the derivation cohort to train the model. All baseline ECGs were digitally recorded at the index visit. All patients underwent echocardiography and coronary angiography as well as invasive pressure measurements in a subset of patients (*n* = 1689, 90%). The ECGs were recorded for 10 s and divided in 2-s segments for each of the 12 leads. A ImageMagick "canny filter" was applied, blank and non-informative segments removed, yielding a total of 77 558 samples. The model was trained using Keras with TensorFlow-GPU (Google, Mountain View, CA, USA) in Python 3.6 and statistics were performed with R-4.0.3 (R Foundation, Vienna, Austria) on a custom-built workstation. The CNN was composed of four convolutional layers, each of which was followed by a dedicated 'ReLu' activation function and a 'max-pooling' layer. The network architecture was chosen through a cross-validation approach and hyper-parameter tuning through random grid selection with multiple layers, kernel size, and filter values. Data were condensed into an output layer with a 'Sigmoid' activation function, as a non-exclusive classifier was needed following the hypothesis that a single ECG segment could contain both HFpEF and non-HFpEF-specific characteristics. The optimization was done by a root-mean-square propagation algorithm. Data sets were divided in 50% training (942 patients), 30% internal validation (565 patients), and 20% test sets (377 patients). The test dataset was withheld and blinded from the network in the beginning to test the accuracy on a 'never-seen' dataset. Following training, the model was tested using the withheld data set. The arithmetic mean of the probabilities of the patient's segments was computed to perform classification. The output threshold was set to 0.4 using the Youden index and the approximation method to maximize sensitivity in the derivation cohort and was used for further predictions. An overview of the pre-processing steps and model algorithm is depicted in *Figure 1*.

Results

According to the ESC criteria, 720 patients (38%) were identified as HFpEF patients and 1164 (62%) as controls. The baseline characteristics are summarized in *Table 1*. Heart failure with preserved ejection fraction patients were older, more frequently females and had a higher body mass index. Heart failure with preserved ejection fraction patients had significantly higher E/E′ values, left atrial volume indices, left ventricular end-diastolic pressures ($P < 0.001$) and higher prevalence of left anterior fascicular ($n = 72$ vs. $n = 66$, $P = 0.027$) but no difference regarding right or left bundle branch block ($n = 22$ vs. $n = 18$, respectively, $P = 0.19$). Overall, 115 patients (6%) presented with atrial fibrillation, ($n = 6$ control vs. $n = 109$ HFpEF, $P < 0.001$). Coronary angiography revealed CAD without the need for intervention in 608 patients (52%) of the control group and in 460 patients (64%) of the HFpEF group ($P < 0.001$). The area under the curve (AUC) of the CNN on the blinded test set was 0.92 [95% confidence interval (CI) 0.91–0.94], allowing for a discrimination between HFpEF and controls with a sensitivity of 0.98 (95% CI 0.97–0.99) and a specificity of 0.63 (95% CI 0.59–0.67). The model was validated by using an external cohort of $n = 203$ volunteers that were in a prospective screening programme for having cardiovascular risk factors and a preserved ejection fraction. These patients underwent ECG recording, laboratory analysis, and echocardiography. The model predictions were tested on the ECGs of the validation cohort, which achieved an AUC of 0.80 (95% CI 0.74–0.86) for detection of HFpEF according to ESC criteria, maintaining the high sensitivity of 0.99 (95% CI 0.98–0.99) and a specificity of 0.60 (95% CI 0.56–0.64), with a positive predictive value of 0.68 (95% CI 0.64–0.72) and a negative predictive value of 0.98 (95% CI 0.95–0.99). The study outline, model building workflow, and the main results are depicted in the *Graphical abstract*. The baseline characteristics of the patients identified as HFpEF by the CNN are displayed in *Table 1* and show a clear distinction, with patients classified as HFpEF having significant higher LA pressure estimates and higher NT-proBNP.

Conclusions

In this study, we report the first deep learning-enabled CNN for the identification of patients with HFpEF according to ESC criteria including NT-proBNP measurements in the diagnostic algorithm among patients at risk for HFpEF. By analysing 12-lead ECGs, the model showed that HFpEF may have specific electrocardiographic characteristics that can be recognized by artificial intelligence algorithms. Importantly, the reliable screening suitability of the CNN was tested on an external validation cohort of patients at risk for developing heart failure, showing a convincing performance in excluding the diagnosis of HFpEF.

Beyond traditional ECG interpretation, machine learning-enabled CNN algorithms could become a valuable and easily applicable screening tool to rule out the diagnosis of HFpEF using a 12-lead ECG with the chance of identifying patients even in an early stage of HFpEF. Further research is needed to validate these findings in larger cohorts.

Funding

Initial funding of the Leipzig Heart Study was supported by the Roland Ernst Foundation. The continuation of the Leipzig (LIFE) Heart Study is supported by LIFE—Leipzig Research Center for Civilization Diseases, Leipzig University. LIFE is funded by means of the Free State of Saxony within the framework of its excellence initiative.

Conflict of interest: P.L.: Consultant to Abbott, Medtronic, and Edwards. All other authors declare no conflict of interest.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

1. 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. 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.
2. 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, 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.
3. Reddy YNV, Carter RE, Obokata M, Redfield MM, Borlaug BA. A simple, evidence-based approach to help guide diagnosis of heart failure with preserved ejection fraction. *Circulation* 2018;**138**:861–870.
4. 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, Friedman PA. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nat Med* 2019;**25**: 70–74.
5. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, Carter RE, Yao X, Rabinstein AA, Erickson BJ, Kapa S, Friedman PA. 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.
6. Kagiyama N, Piccirilli M, Yanamala N, Shrestha S, Fario PD, Casaclang-Verzosa G, Tarhuni WM, Nezarat N, Budoff MJ, Narula J, Sengupta PP. Machine learning assessment of left ventricular diastolic function based on electrocardiographic features. *J Am Coll Cardiol* 2020;**76**:930–941.
7. Kwon J, Kim K-H, Eisen HJ, Cho Y, Jeon K-H, Lee SY, Park J, Oh B-H. Artificial intelligence assessment for early detection of heart failure with preserved ejection fraction based on electrocardiographic features. *Eur Heart J - Digit Health*. 2021;**2**: 106–116.