







Automated detection of low ejection fraction from a one-lead electrocardiogram: application of an AI algorithm to an electrocardiogram-enabled Digital Stethoscope

Zachi I. Attia ¹, Jennifer Dugan ¹, Adam Rideout², John N. Maidens², Subramaniam Venkatraman², Ling Guo², Peter A. Noseworthy ¹, Patricia A. Pellikka ¹, Steve L. Pham², Suraj Kapa ¹, Paul A. Friedman¹, and Francisco Lopez-Jimenez ¹*

¹Department of Cardiovascular Medicine, Mayo Clinic College of Medicine, Rochester, MN, USA; and ²Eko Devices, Inc., Berkeley, CA, USA

Received 1 November 2021; revised 21 April 2022; online publish-ahead-of-print 23 May 2022

Aims

Electrocardiogram (ECG)-enabled stethoscope (ECG-Scope) acquires a single-lead ECGs during cardiac auscultation and may facilitate real-time screening for pathologies not routinely identified by cardiac auscultation alone. We previously demonstrated an artificial intelligence (AI) algorithm can identify left ventricular dysfunction (LVSD) [defined as ejection fraction (EF) \leq 40%] with an area under the curve (AUC) of 0.91 using a 12-lead ECG.

Methods and results

One hundred patients referred for clinically indicated echocardiography were prospectively recruited. ECG-Scope recordings with the patient supine and sitting were obtained in multiple electrode locations at the time of the echocardiogram. The AI algorithm for the detection of LVSD was retrained using single leads from ECG-12 and validated against ECG-Scope to determine accuracy for low EF detection (\leq 35%, $<$ 40%, or $<$ 50%). We evaluated the algorithm with respect to body position and lead location. Amongst 100 patients (aged 61.3 ± 13.8 ; 61% male, BMI: 30.0 ± 5.4), eight had $EF \leq 40\%$, and six had EF 40–50%. The best single recording position was V2 with the patient supine [AUC: 0.88 (CI: 0.80–0.97) for $EF \leq 35\%$, 0.85 (CI: 0.75–0.95) for $EF \leq 40\%$, and 0.81 (CI: 0.71–0.90) for $EF < 50\%$]. When using an AI model to select the recording automatically, AUC was 0.91 (CI: 0.84–0.97) for $EF \leq 35\%$, 0.89 (CI: 0.83–0.96) for $EF \leq 40\%$, and 0.84 (CI: 0.73–0.94) for $EF < 50\%$.

Conclusion

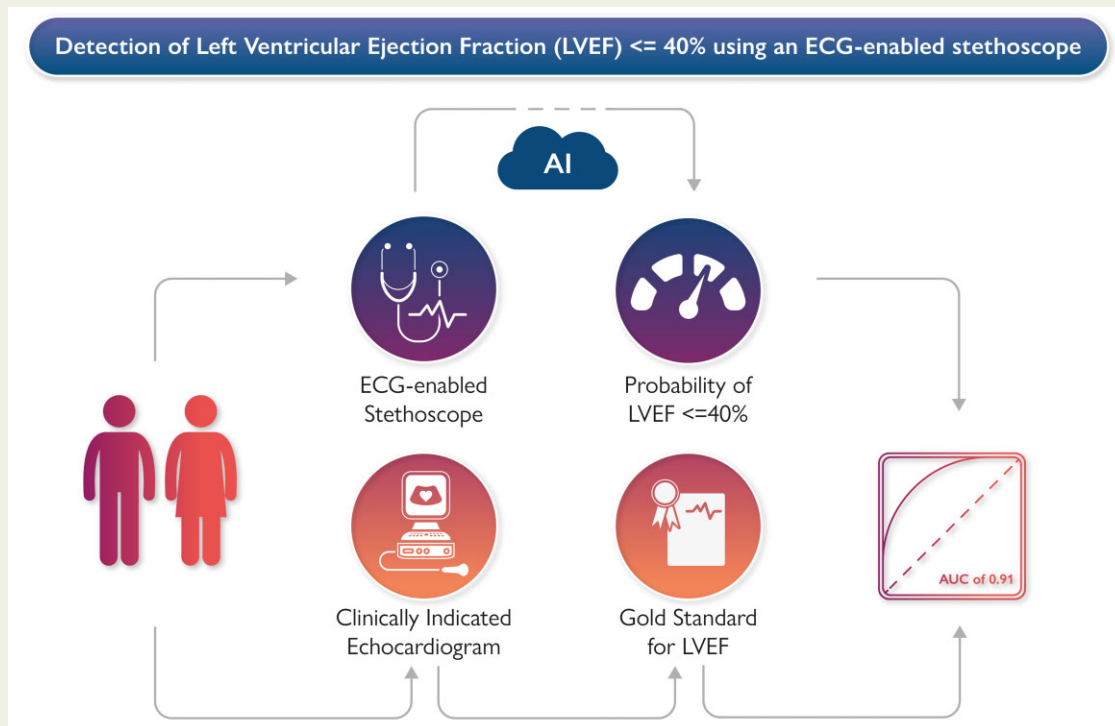
An AI algorithm applied to an ECG-enabled stethoscope recording in standard auscultation positions reliably detected the presence of a low EF in this prospective study of patients referred for echocardiography. The ability to screen patients with a possible low EF during routine physical examination may facilitate rapid detection of LVSD.

* Corresponding author. Tel: +1 (507) 284-8087. Email: lopez@mayo.edu

© The Author(s) 2022. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Graphical Abstract



Keywords

Artificial intelligence • Electrocardiogram • Stethoscope • Ejection fraction • Heart failure

Introduction

Since the invention of the stethoscope in 1819, it has become well recognized that the stethoscope is an essential component of any diagnostician's armamentarium—that the science of auscultation can serve the well-trained physician in not only identifying diseases early in their course but to also augment findings obtained via other, often more costly tests. However, while auscultation may offer one perspective on cardiac health, the electrocardiogram (ECG) may offer additional value.

Recent data has suggested that, by applying algorithms derived out of large, well-annotated data sets using various artificial intelligence (AI) techniques, it is possible to determine potassium values, patient age or sex, or the presence of a low ejection fraction (EF) from the ECG alone.^{1–3} In particular, in the case of a low EF, whereas there are well-recognized treatments to reduce associated morbidity and mortality, around 8% of the population may be otherwise asymptomatic and go undiagnosed.⁴ Thus, as a low EF may not always be obvious during routine clinical examination, the ability of the ECG to automatically suggest the presence of a low EF and to drive the clinician to refer for confirmatory testing (e.g. echocardiography) is clear. However, widespread acquisition of routine 12-lead ECGs in ostensibly healthy patients is neither cost-effective nor efficient. A recent technology has demonstrated the ability to acquire a one-lead, digital ECG during routine cardiac auscultation.⁵ Applying AI-enabled, ECG-derived

clinical diagnostic algorithms to the signals acquired by such a device may allow for early identification of clinical pathology during the routine physical examination. Thus, we sought to evaluate the predictive accuracy of using such an ECG-enabled digital stethoscope in automatically identifying patients with a low EF via a previously reported AI algorithm derived from 12-lead ECGs.

Methods

According to a Mayo Clinic Institutional Review Board (IRB) approved protocol, 100 consecutive patients were approached for inclusion in the study. All patients were referred for outpatient transthoracic echocardiography for any indication to the Mayo Echocardiography Laboratory. Informed consent was obtained immediately before their echocardiogram.

Electrocardiogram/heart sound acquisition

After obtaining patient informed consent, the ECG-enabled stethoscope (Eko DUO, Eko Devices, Inc; Oakland, CA) was applied to the patient's chest in a variety of locations (Figure 1) both supine and sitting by a single operator (J.D.). ECG was recorded in 500 Hz for 15 s in each position. The AI algorithm (described below) was applied to all single-lead ECGs obtained to determine a probability score (0–1) of the patients having a low EF (defined as $\leq 35\%$, $\leq 40\%$, or $< 50\%$).

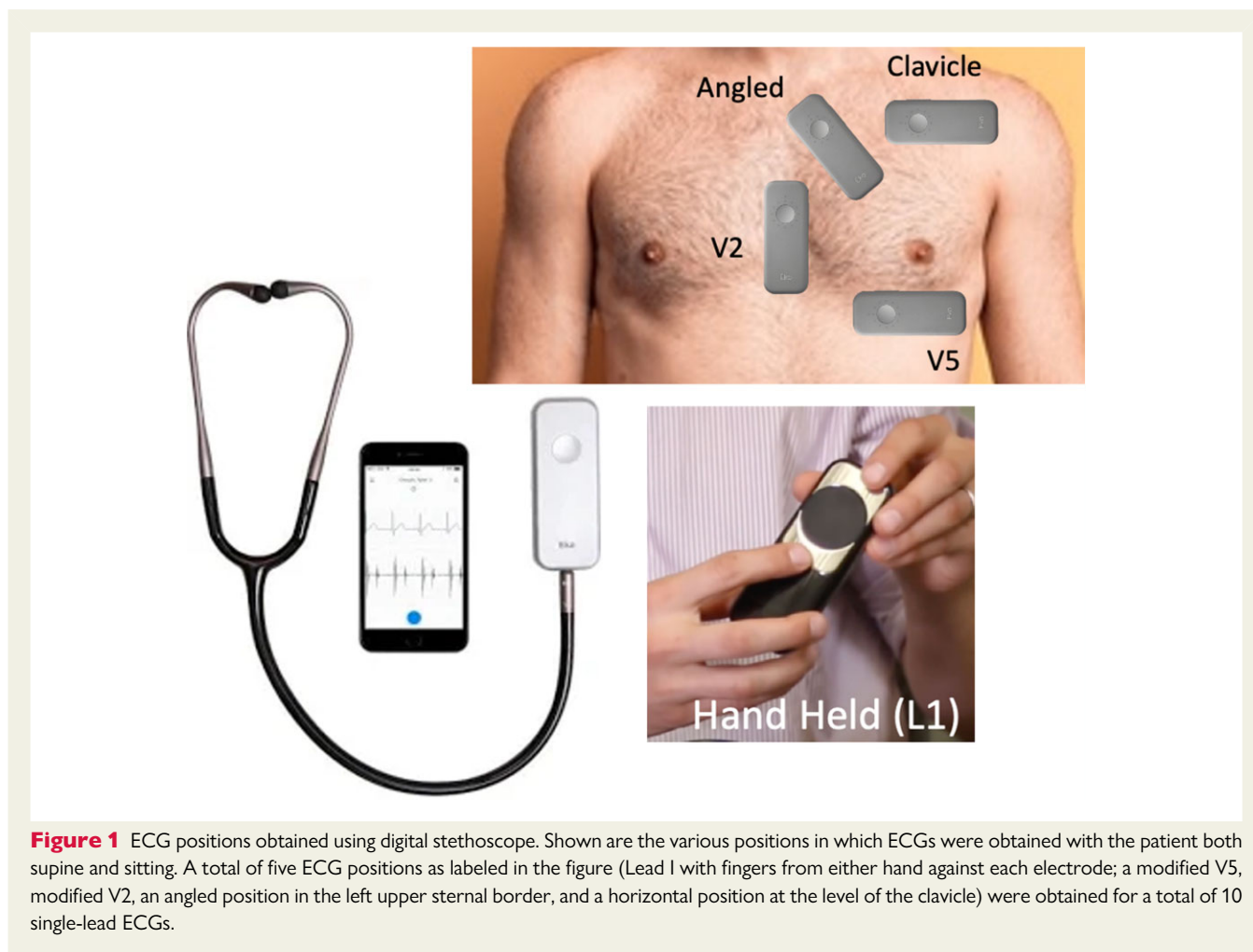


Figure 1 ECG positions obtained using digital stethoscope. Shown are the various positions in which ECGs were obtained with the patient both supine and sitting. A total of five ECG positions as labeled in the figure (Lead I with fingers from either hand against each electrode; a modified V5, modified V2, an angled position in the left upper sternal border, and a horizontal position at the level of the clavicle) were obtained for a total of 10 single-lead ECGs.

Construction of neural network for use on single-lead electrocardiogram

The development, validation, and network architecture for the 12-lead ECG EF prediction algorithm has been previously published.³ To retool the network for use on a single-lead ECG, a lead agnostic training method to allow the 12-lead AI-ECG model to function off of a single lead was performed. This consisted of retraining the model to use every individual lead as a unique, independent lead for the purposes of identifying a low EF similar to the methods described previously.³ This was carried out using the same cohort used to develop the original model, containing 35 970 patients (of those 3894 patients had an $EF \leq 40\%$) used for seeding the network weights (training set), and 8989 patients (of those 990 patients had an $EF \leq 40\%$) used for internal validation, by changing the model architecture described in³ from a convolutional neural networks with an input of (12×5000) to (1×5000) , as the original model was operating on each lead separately in all layers except for one that used to combined the features from the different leads (the 'spatial block' in the original manuscript), we removed the 'spatial block' and retrained the model, during training we fed each of the leads from the original 12 leads as an independent sample after normalizing it to have an maximum absolute amplitude of 1 (au), that is not affected from the ECG polarity. For the testing set from the original derivation cohort, the area under the curve (AUC) from averaging the scores for all 12 leads tested independently, generating a per-patient score, was 0.9 for detection of $EF \leq 35\%$.

Owing to the variability in ECGs recorded using a mobile form factor placed on the patient chest in different locations, we normalized all data sets used in this study to have a maximum amplitude of 1 unit. In addition, when training and testing the algorithm, we used one version in which the ECG was fed to the network as-is, and one version when the ECG amplitude was multiplied by '-1' to mimic a situation when the device electrodes are in opposite orientation. In the testing stage, the score of both versions was averaged to make the model invariant to electrode reversal.

Neural network to determine optimal electrocardiogram lead

When obtaining ECGs from multiple positions, certain leads may have better overall signal characteristics with minimal noise, baseline wander, and maximal contact to allow for adequate cardiac signal acquisition. An algorithm was developed to automatically identify the optimal ECG lead from all stethoscope acquired ECGs. Development of this algorithm consisted of training a convolutional neural network model to classify signal quality of a single-lead ECG strip into one of three classes: good, moderate and poor signal quality. This model was trained using a data set of 1408 annotated ECGs collected at clinical sites different from the study site used to collect the test set using the digital ECG stethoscope system. The model takes in a rolling window of two-second ECG segments in natural (multiplied by 1) and inverted (multiplied by -1) orientations. Model outputs were averaged across segments and orientations to determine a single output for each ECG strip. The model output then

was used to identify the optimal ECG signal for a given study subject by selecting the lead with maximum probability of good signal quality according to the model's averaged outputs. The single-lead AI-ECG model was run on each of the ECGs obtained and also on the optimal ECG lead based on this methodology.

Low ejection fraction outcome analysis

A total of 978 recordings were used for the analysis (summary of missing recordings per patients are presented in see [Supplementary material online, Table S1](#)). Each ECG recording was normalized and evaluated using both the EF-AI single-lead algorithm yielding a probability score (0–1) of the likelihood of having a low EF and the AI-Quality network. After deriving the scores for each lead location in the two body positions, an ensemble score (optimal score) was calculated by using the single recording with the highest AI-Quality score per patient. The network reports a score between 0 and 1 (0 = low likelihood of low EF; 1 = high likelihood) for each ECG that is evaluated. The area under the receiver operating characteristics curve (AUC) was the primary outcome measures and was determined for the test cohort.

Results

Demographics

A total of 100 patients (age 61.3 ± 13.8 years, 61% male) comprised the study group. Seven patients had an $EF \leq 35\%$, one had $35\% < EF \leq 40\%$, and six had $EF 40\text{--}50\%$. Indications for echocardiography included atrial fibrillation (N = 15); dyspnoea (N = 25); heart failure (N = 19); syncope (N = 6); bradycardia (N = 3); arrhythmia (N = 11); and other reasons (N = 21).

Accuracy of single-lead digital stethoscope-enabled electrocardiogram in identifying a low ejection fraction

[Figure 1](#) depicts the various positions in which ECGs were obtained via the ECG-enabled stethoscope in both supine and sitting positions.

[Table 1](#) summarizes the area under the receiver operator curve (AUC) for each of the positions in which the ECG was obtained for prediction of an $EF < 35\%$, $EF < 40\%$ and $< 50\%$. The single recording with the highest AUC was in the V2 position with the patient supine. However, the average score across all leads performed similarly robustly (AUC for identifying $EF < 35\%$ 0.86; $EF < 40\%$ 0.75; and $< 50\%$ 0.82) ([Figure 2](#)). When the system's algorithm was used to automatically select the optimal ECG, however, the AUC for prediction of low EF was highest [AUC for $< 35\%$ 0.91 CI: (0.84–0.97); for $< 40\%$ 0.89 CI: (0.83–0.96); for $< 50\%$ 0.84 CI: (0.72–0.96)].

Accuracy and model statistics for the invariant lead

Sensitivity, specificity, NPV, and PPV were calculated using the invariant lead model for the different EF thresholds. For $EF \leq 35\%$: sensitivity, 85.7%; specificity, 84.9%; PPV, 30.0%; and NPV, 98.8% using a threshold score of 56% ([Figure 3](#)). For $EF \leq 40\%$: sensitivity, 87.5%; specificity, 80.4%; PPV, 28.0%; NPV, 98.7% with a threshold score of 53.8% and for $EF < 50\%$ sensitivity, 85.7%; specificity, 80.2%; PPV, 41.4%, and NPV, 97.2% using a threshold score of 51.8%.

Discussion

The stethoscope has seen many changes over the past 200 years. However, the essential attribute being detected—namely heart sounds derived from the mechanical flow of blood—has barely evolved. Although other efforts to pair ECG data with heart sounds have existed over the years (e.g. phonocardiography), their use has been limited due to lack of convenience (stemming from the need of additional wires and connections to the patient's person) or of interpretability (due to the need of an experienced cardiologist to interpret the data).^{5,6}

Table 1 AUC by stethoscope/ECG position

| Body position | Lead location | AUC for $EF \leq 35\%$ | AUC for $EF \leq 40\%$ | AUC for $EF < 50\%$ | N | N | N | N | N |
|-----------------------|---------------|------------------------|------------------------|----------------------|-----|--------------|--------------|-----------|--------------|
| | | | | | | $EF \leq 35$ | $EF \leq 40$ | $EF < 50$ | $EF \geq 50$ |
| Sitting | AN | 0.83 CI: (0.64–1) | 0.77 CI: (0.58–0.97) | 0.76 CI: (0.63–0.89) | 98 | 7 | 8 | 14 | 69 |
| Sitting | CL | 0.83 CI: (0.69–0.97) | 0.8 CI: (0.67–0.94) | 0.77 CI: (0.64–0.9) | 98 | 7 | 8 | 14 | 69 |
| Sitting | L1 | 0.86 CI: (0.74–0.99) | 0.86 CI: (0.75–0.97) | 0.83 CI: (0.72–0.93) | 97 | 7 | 8 | 14 | 68 |
| Sitting | V2 | 0.79 CI: (0.65–0.94) | 0.8 CI: (0.67–0.93) | 0.73 CI: (0.56–0.9) | 96 | 7 | 8 | 13 | 68 |
| Sitting | V5 | 0.72 CI: (0.52–0.92) | 0.72 CI: (0.54–0.89) | 0.73 CI: (0.59–0.86) | 96 | 7 | 8 | 14 | 67 |
| Supine | AN | 0.78 CI: (0.54–1) | 0.77 CI: (0.56–0.97) | 0.73 CI: (0.56–0.9) | 100 | 7 | 8 | 14 | 71 |
| Supine | CL | 0.79 CI: (0.63–0.95) | 0.78 CI: (0.65–0.92) | 0.81 CI: (0.71–0.9) | 100 | 7 | 8 | 14 | 71 |
| Supine | L1 | 0.84 CI: (0.62–1) | 0.85 CI: (0.66–1) | 0.81 CI: (0.65–0.96) | 94 | 7 | 8 | 14 | 65 |
| Supine | V2 | 0.88 CI: (0.79–0.97) | 0.85 CI: (0.75–0.95) | 0.81 CI: (0.7–0.93) | 100 | 7 | 8 | 14 | 71 |
| Supine | V5 | 0.75 CI: (0.59–0.91) | 0.75 CI: (0.61–0.89) | 0.72 CI: (0.59–0.85) | 99 | 7 | 8 | 14 | 70 |
| Average | Invariant | 0.86 CI: (0.69–1) | 0.85 CI: (0.71–1) | 0.83 CI: (0.71–0.95) | 100 | 7 | 8 | 14 | 71 |
| Best lead per patient | Invariant | 0.91 CI: (0.84–0.97) | 0.89 CI: (0.83–0.96) | 0.84 CI: (0.72–0.96) | 100 | 7 | 8 | 14 | 71 |

The lead positions correlate with the positions demonstrated in [Figure 2](#). The optimal lead and average were defined as outlined in the methods, in some positions with missing ECGs, the analysis was carried out with available data and the number of patients in each group was reported.

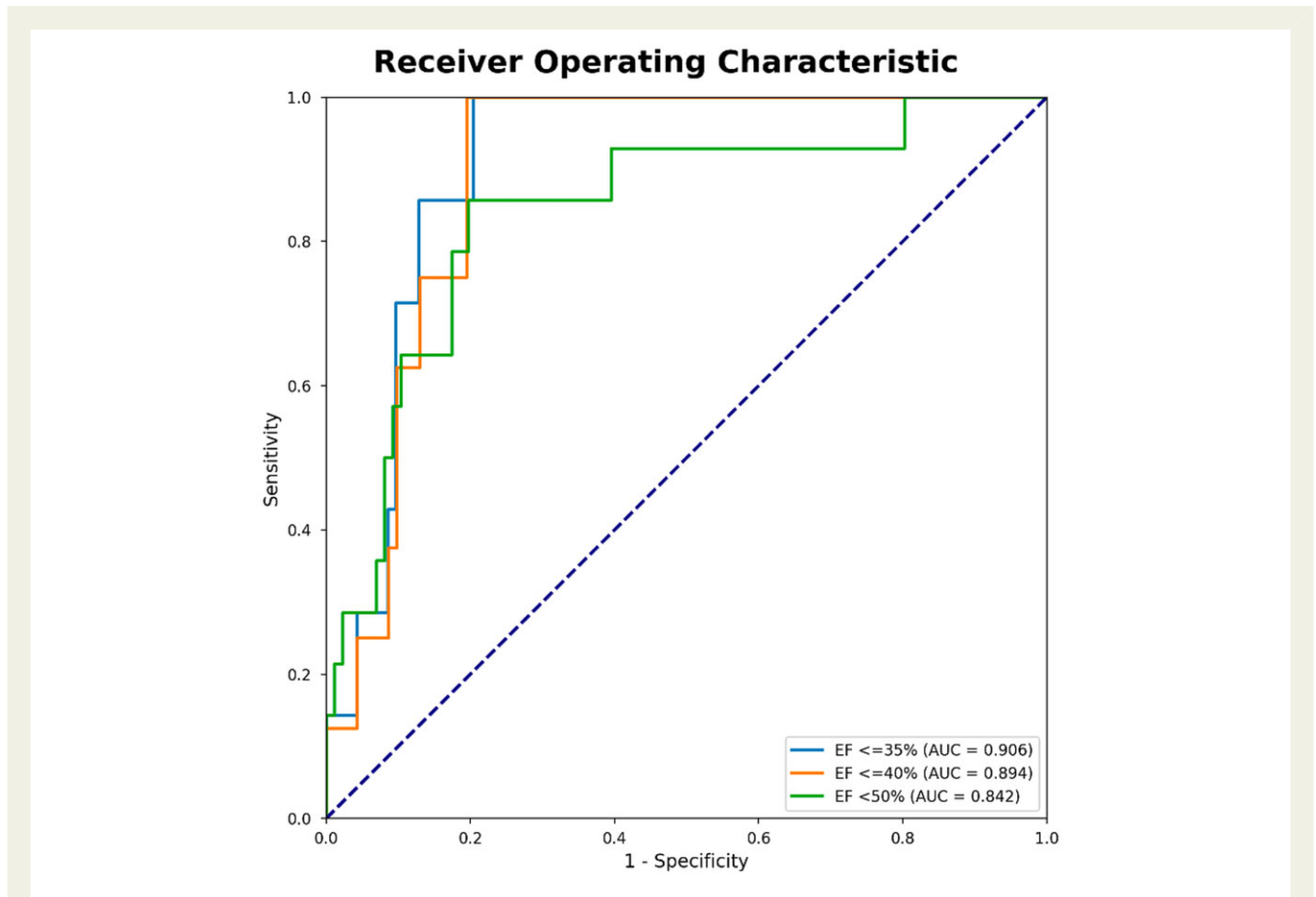


Figure 2 Receiver-operating curve for prediction of low EF. Shown is the ROC curve for prediction of a low EF using the single-lead, stethoscope-enabled ECG. The ROC curve shown is for the optimal lead signal from all ECGs obtained. The blue line indicates the ROC curve for prediction of an EF $\leq 35\%$ [AUC = 0.91 CI: (0.84–0.97)], the orange curve for prediction of an EF $\leq 40\%$ [AUC = 0.89 CI: (0.83–0.96)] and the orange curve for prediction of an EF $\leq 50\%$ [AUC = 0.84 CI: (0.72–0.96)].

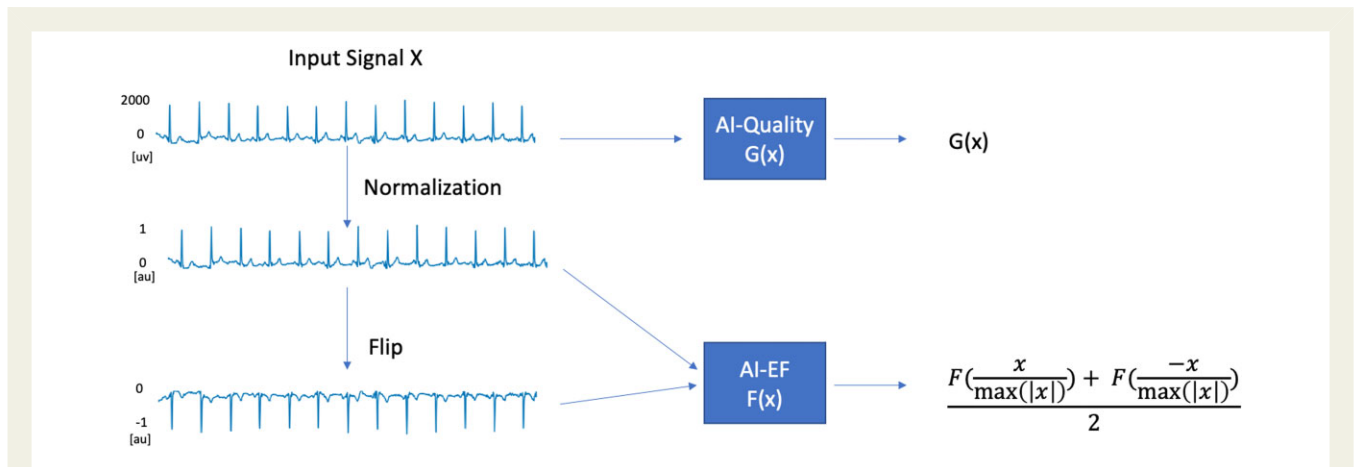
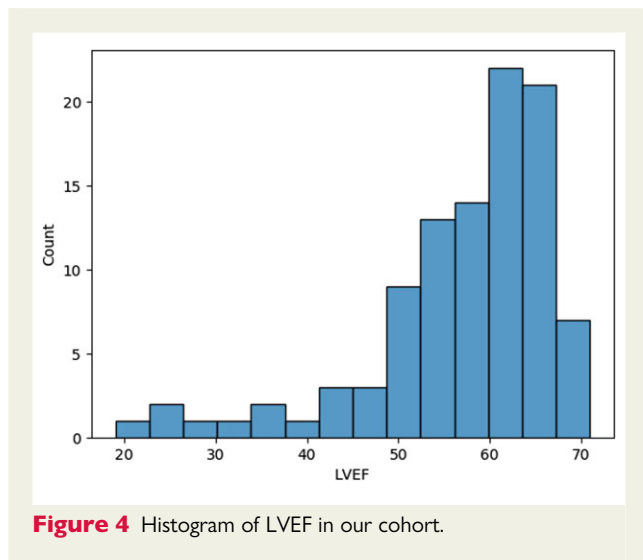


Figure 3 Pre-processing. To make the model results invariant to amplitude and direction, each ECG is normalized and flipped and the EF score is the average score of both directions.

Thus, although ECG data could certainly be acquired during normal bedside examination, it has not been done routinely (Figure 4).

The capacity of a stethoscope used for routine cardiac auscultation to simultaneously record an ECG at each position in which the device is placed may offer additional value in part due to



(i) avoidance of cumbersome patches, wires, or other externally attached machinery; and (ii) ability to augment the data via automated diagnostics, which may supplement the physician's interpretation of physical findings with minimal user interaction or prior expertise; and (iii) the fact that the use of a stethoscope is integrated into clinical workflows, facilitating ready adoption. The ability to augment what is otherwise a routine portion of any physical examination may have profound implications on early identification of disease, such as the presence of a low EF.

Our data suggests that it is feasible to routinely obtain ECG signals using an ECG-enabled digital stethoscope, process such signals through an AI-enabled algorithm trained from 12-lead ECGs; and automatically detect presence of a low EF with clinical useful predictive power from the stethoscope-enabled ECG when compared with 12-lead ECGs in the same patients in this prospective study.

The overall accuracy of the ECG-enabled stethoscope was good and while the V2 position supine had the best results followed by the average probability attained based on ECGs obtained from all positions together, an algorithm that selected the lead with the optimal signal overall performed the best (AUC = 0.91 for $EF \leq 35\%$; 0.84 for $EF < 50\%$). The reason for this might be due to recordings from chest wall locations with poorer signal quality resulting in an inaccurate prediction, and that the EF algorithm is dependent on having an optimal signal at the time of acquisition. However, during a normal physical examination, it is expected that the stethoscope is applied to a minimum of four locations for the purposes of cardiac auscultation. Furthermore, it is also considered routine to listen with the patient both supine and sitting.⁷ Thus, this variability in accuracy based on single site, single-position ECG acquisition is unlikely to be relevant when the stethoscope is used in routine clinical practice due to the fact that multiple ECGs would be obtained anyway.

Future research will be needed to determine the diagnostic accuracy for low EF in an ostensibly healthy population, the reproducibility of obtaining diagnostic quality ECG signals during routine physical examination by providers, and the clinical impact of such early detection algorithms for a low EF. In addition, while not specifically studied here, it is possible that heart sounds may contain similar data to drive prediction of a low EF equivalent to or better than an ECG alone, or

that the combination of both may improve diagnostic accuracy. Finally, impact of cost of such new digital technologies, cost of increased referrals for more advanced testing (in this case echocardiography), and acceptability of the augmented interpretations by clinicians will have to be weighed against impact on patient outcomes at a larger population level.

Limitations

There are several limitations to our study. First, the number of patients included was small. As all were referred for an echocardiogram, pre-test probability of some cardiac disease was high. Thus, the risk of false positives could be higher in an otherwise healthy population who would not have otherwise had an indication for an echocardiogram. Second, all stethoscope-enabled ECG acquisition was performed by a single individual. Thus, reproducibility of signal acquisition between different providers requires further study. Third, as is already a recognized current limitation of AI and, specifically, neural network techniques, the exact features of the ECG leading to the automated interpretation of a low versus normal EF are not easily identifiable. Finally, as we have reported previously, amongst patients with 'false positives' (i.e. presence of a normal EF when the algorithm predicted a low EF), there is a nearly five-fold increased risk of developing a low EF over follow-up when compared with patients where the ECG predicted the EF was normal and it actually was.⁽⁴⁾ Thus, it is possible that some of our 'false positives' may nevertheless reflect a high-risk cohort for development of a low EF in the future, though the follow-up period was too short to prove this.

Conclusion

In this prospective study of patients referred for echocardiography, it was feasible to automatically identify patients with depressed ventricular function using a stethoscope with embedded electrodes for automated ECG acquisition. The ability to augment bedside physical diagnostics through advancing digital technologies may improve early diagnosis of ventricular dysfunction and, potentially, other physical ailments. Future work evaluating impact on patient outcomes, costs of care, and reproducibility amongst different clinical providers will be needed to validate clinical impact.

Lead author biography



Zachi I. Attia, PhD, works in artificial intelligence, machine learning and signal processing. In his research, he uses machine learning to develop tools that enable the detection and prediction of diseases using cardiac biosignals. Heart disease is the leading cause of death in the United States. But many people are diagnosed and treated only after symptoms appear, after developing severe morbidities or when they present with sudden cardiac death. People who have no symptoms manifest silent markers. These markers precede overt disease and are recorded during routine tests. But they are missed by

clinicians, as these biomarkers are invisible to the human eye. Dr. Attia uses multimodal cardiac data including electrocardiograms, echocardiograms and angiograms to develop artificial intelligence models that can detect treatable but silent diseases with high accuracy. Using pragmatic clinical trials, he collaborates closely with cardiologists, cardiac surgeons and other medical experts to test these models and incorporate them into clinical practice. The goal is to empower clinicians to apply these methods in practice without requiring any AI expertise.

Supplementary material

Supplementary material is available at *European Heart Journal – Digital Health*.

Funding

This study was funded using institutional funds at Mayo Clinic for data collection and statistical analyses.

Conflict of interest: Mayo clinic have licensed this technology to Eko Devices and to Anumana, Mayo Clinic and the inventors of the technology (PAF,FLJ, ZIA and SK) might benefit from its commercialization, but will not benefit financially from the use of the technology on Mayo patients. A.R., J.N.M., S.V., L.G., and S.L.P. are employees of Eko.

Data availability

All requests for raw and analyzed data will be reviewed by the Mayo Clinic legal department and Mayo Clinic Ventures to verify

whether the request is subject to any intellectual property or confidentiality obligations. Requests for patient-related data not included in the paper will not be considered. Any data and materials that can be shared will be released via a Material Transfer Agreement.

References

1. Galloway CD, Valys AV, Shreibati JB, Treiman DL, Petterson FL, Gundotra VP, Albert DE, Attia ZI, Carter RE, Asirvatham SJ, Ackerman MJ, Noseworthy PA, Dillon JJ, Friedman PA. Development and validation of a deep-learning model to screen for hyperkalemia from the electrocardiogram. *JAMA Cardiol* 2019;**4**: 428–436.
2. Attia ZI, Friedman PA, Noseworthy PA, Lopez-Jimenez F, Ladewig DJ, Satam G, Pellikka PA, Munger TM, Asirvatham SJ, Scott CG, Carter RE, Kapa S. Age and sex estimation using artificial intelligence from standard 12-lead ECGs. *Circ Arrhythm Electrophysiol* 2019;**12**:e007284.
3. 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.
4. Wang TJ, Levy D, Benjamin EJ, Vasan RS. The epidemiology of “asymptomatic” left ventricular systolic dysfunction: implications for screening. *Ann Intern Med* 2003; **138**:907–916.
5. Varma N, Marrouche NF, Aguinaga L, Albert CM, Arbelo E, Choi JJ, et al. HRS/EHRA/APHRS/LAHRs/ACC/AHA worldwide practice update for telehealth and arrhythmia monitoring during and after a pandemic. *Circ Arrhythm Electrophysiol* 2020;**13**: e009007.
6. Sprague HB, Ongley PA. The clinical value of phonocardiography. *Circulation* 1954;**9**: 127–34.
7. Orient JM. *Sapira's Art & Scient of Bedside Diagnosis*. Philadelphia: Lippincott Williams and Wilkins; 2009.
8. Laennec RTH. *A Treatise on the Diseases of the Chest*. London: Hafner Publishing; 1821.