JACC: ADVANCES © 2025 THE AUTHORS. PUBLISHED BY ELSEVIER ON BEHALF OF THE AMERICAN COLLEGE OF CARDIOLOGY FOUNDATION. THIS IS AN OPEN ACCESS ARTICLE UNDER THE CC BY LICENSE (http://creativecommons.org/licenses/by/4.0/).

**ORIGINAL RESEARCH** 

# Automated Detection of Reduced Ejection Fraction Using an ECG-Enabled Digital Stethoscope

| 1 | h |  |
|---|---|--|
|   | 2 |  |

# A Large Cohort Validation

Ling Guo, PhD,<sup>a</sup> Gregg S. Pressman, MD,<sup>b</sup> Spencer N. Kieu, BS,<sup>a</sup> Scott B. Marrus, MD, PhD,<sup>c</sup> George Mathew, PhD,<sup>a</sup> John Prince, PhD,<sup>a</sup> Emileigh Lastowski, MS,<sup>a</sup> Rosalie V. McDonough, MD, MSc,<sup>a</sup> Caroline Currie, BA,<sup>a</sup> John N. Maidens, PhD,<sup>a</sup> Hussein Al-Sudani, MD,<sup>c</sup> Evan Friend, BA,<sup>b</sup> Deepak Padmanabhan, MD,<sup>d</sup> Preetham Kumar, MD,<sup>e</sup> Edward Kersh, MD,<sup>f</sup> Subramaniam Venkatraman, PhD,<sup>a</sup> Salima Qamruddin, MD<sup>g</sup>

## ABSTRACT

**BACKGROUND** Asymptomatic left ventricular systolic dysfunction (ALVSD) affects 7 million globally, leading to delayed diagnosis and treatment, high mortality, and substantial downstream health care costs. Current detection methods for ALVSD are inadequate, necessitating the development of improved diagnostic tools. Recently, electrocardiogram-based algorithms have shown promise in detecting ALVSD.

**OBJECTIVES** The authors developed and validated a convolutional neural network (CNN) model using single-lead electrocardiogram and phonocardiogram inputs captured by a digital stethoscope to assess its utility in detecting individuals with actionably low ejection fractions (EF) in a large cohort of patients.

**METHODS** 2,960 adults undergoing echocardiography from 4 U.S. health care networks were enrolled in this multicenter observational study. Patient data were captured using a digital stethoscope, and echocardiograms were performed within 1 week of data collection. The algorithm's performance was compared against echocardiographic EF (EF measurements, categorizing EF as normal and mildly reduced [>40%] or moderate and severely reduced [≤40%]).

**RESULTS** The CNN model demonstrated an area under the receiver operating characteristic curve of 0.85, with a sensitivity of 77.5%, specificity of 78.3%, positive predictive value of 20.3%, and negative predictive value of 98.0%. Among those with an abnormal artificial intelligence screen but EF >40% (false positives), 25% had an EF between 41%-49% and 63% had conduction/rhythm abnormalities. Subgroup analyses indicated consistent performance across various demographics and comorbidities.

**CONCLUSIONS** The CNN model, utilizing a digital stethoscope, offers a noninvasive and scalable method for early detection of individuals with EF  $\leq$ 40%. This technology has the potential to facilitate early diagnosis and treatment of heart failure, thereby improving patient outcomes. (JACC Adv. 2025;4:101619) © 2025 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

From the <sup>a</sup>Eko Health, Inc, Emeryville, California, USA; <sup>b</sup>Jefferson Einstein Hospital, Thomas Jefferson University, Philadelphia, Pennsylvania, USA; <sup>c</sup>Prairie Education & Research Cooperative, Springfield, Illinois, USA; <sup>d</sup>Sri Jayadeva Institute of Cardiovascular Sciences and Research, Bengaluru, India; <sup>e</sup>Heart and Vascular Institute, MedStar Washington Hospital Center, Washington, District of Columbia, USA; <sup>f</sup>Sutter Care at Home, Fairfield, California, USA; and the <sup>g</sup>Ochnser Heart and Vascular Institute, Ochsner Medical Center, New Orleans, Louisiana, USA.

#### ABBREVIATIONS AND ACRONYMS

2

#### AI = artificial intelligence

ALVSD = asymptomatic left ventricular systolic dysfunction

AUROC = area under the receiver operating characteristic curve

CNN = convolutional neural network

ECG = electrocardiogram

EF = ejection fraction

ML = machine learning

**NPV** = negative predictive value

NTproBNP = N-terminal pro B-type natriuretic peptide

PCG = phonocardiogram PPCM = peripartum

cardiomyopathy

**PPV** = positive predictive value

eart failure remains a significant global health concern. Congestive heart failure affects 5 million people and consumes more than 30 billion in health care expenditure.<sup>1,2</sup> Asymptomatic left ventricular systolic dysfunction(LVSD) has an estimated prevalence of 3% to 6%, affecting approximately 7 million people, and is at least as common in the community as systolic heart failure.<sup>3</sup> Because it often occurs in the absence of known cardiovascular disease, this condition may go unrecognized and untreated.<sup>3</sup> Asymptomatic LVSD, classified as stage B heart failure, is defined as structural heart disease without current or prior symptoms. Early intervention is known to be effective, yet late diagnosis persists, leading to heart failure being identified in acute care settings despite earlier documentation of potential symptoms in primary care.<sup>4</sup> Delayed diagnosis contributes to the

unacceptably high mortality rates associated with heart failure.<sup>5</sup> Additionally, the underutilization of echocardiography and late diagnosis of heart failure incur substantial costs for health care systems.<sup>6</sup> Underlying this is the lack of simple and scalable screening tools available for early detection and intervention.<sup>7</sup>

Since its invention over 200 years ago, the stethoscope has remained an indispensable tool in the clinical examination of patients.<sup>8</sup> With advancements in technology, notably the advent of artificial intelligence (AI) and digital stethoscopes capable of capturing both heart sound (phonocardiogram [PCG]) and electrocardiogram (ECG) signals, diagnostic capabilities of the stethoscope have substantially improved for a wide range of users. While echocardiography remains the gold standard for diagnosing low ejection fraction (EF), its widespread use is limited by resource constraints, requirement for extensive training, operator-dependent variability, and cost. PCG and ECG signals, captured by a digital stethoscope and processed through AI-enabled algorithms, provide a noninvasive method to detect a low EF that can be used equally effectively by individuals with all levels of training and experience.<sup>1,9</sup>

Digital stethoscopes have already demonstrated their ability to improve diagnostic accuracy in several

clinical contexts, including the detection of heart murmurs and lung sounds. By incorporating AI into these devices, we aim to further enhance their utility in the detection of low EF, offering a noninvasive, accessible tool for early heart failure diagnosis.<sup>10,11</sup>

To our knowledge, few studies have explored AI-enhanced digital stethoscopes for detecting low EF. These either did not incorporate PCG data into their algorithm or involved different patient populations (eg, peripartum individuals). The usage of PCG for the detection of low EF is new in this study. Models based on PCG signal alone are less predictive for low EF compared to models based on ECG alone. In a previous study,<sup>12</sup> it was observed that a model using PCG alone had an area under the curve of 77.8%, while a model using single-lead ECG alone had an area under the curve of 80.4% for detecting low EF. In the current study, we show that a model that simultaneously analyzes synchronously captured PCG and ECG performs better than a model that uses single-lead ECG alone. Our study incorporates both single-lead ECG and PCG signals into the algorithm to detect LVEF  $\leq$ 40% in a diverse, real-world cohort. Our aim was to validate the predictive accuracy of this algorithm in a real-world setting in identifying patients with an EF of 40% or less, thus facilitating early detection and management.

#### **METHODS**

**STUDY POPULATION AND SETTING.** This multicenter observational study was conducted using data from 4 geographically distinct U.S. health care networks: Jefferson Einstein Philadelphia Hospital (Pennsylvania), Prairie Cardiovascular Consultants (Illinois), MedStar Health Research Institute (Maryland), and Ochsner Heart and Vascular Institute (Louisiana). The study population consisted of adults aged 18 years and older undergoing a clinically indicated echocardiogram for any reason within 7 days of study procedures and who were willing and capable of providing informed consent. Nonprobability, exhaustive sampling method was used to enroll individuals in the study. Exclusion criteria included individuals who were unwilling or unable to provide informed consent, as well as hospitalized patients. Prior diagnosis of heart failure or reduced EF did not preclude patients from participating in the study. Patient demographic

Manuscript received August 19, 2024; revised manuscript received January 15, 2025, accepted January 18, 2025.

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.

data, including age, race/ethnicity, sex, and medical history, were collected. The responsible Institutional Review Boards and ethics committees of the participating centers approved this study.

ALGORITHM DEVELOPMENT. The development of the low EF convolutional neural network (CNN) model built upon previous work conducted by Attia et al.<sup>1,9</sup> Briefly, the model uses an ensemble of 3 CNNs. The primary component of the ensemble (referred to as the ECG model) was trained on 44,959 12-lead ECG and echocardiogram pairings from the Mayo Clinic database.<sup>1</sup> The inputs to the ECG model were single leads from the 12-lead ECGs in the Mayo Clinic data set and the model was trained to predict whether LVEF was normal (>40%) or low ( $\leq$ 40%). Information on which ECG lead was being used was not provided to the ECG model. The model was trained in this manner so that the resulting model is robust and not sensitive to the location/orientation at which the ECG lead is captured. To further refine the algorithm, 18,999 paired PCG and ECG recordings from 1,852 subjects sourced from a proprietary data set, referred to as the Eko Training Dataset, were utilized. The training data set was acquired from 8 clinical study sites (both within and outside the United States) independent of the study sites used for testing. Subjects in the training data set underwent transthoracic echocardiography within 30 days of the ECG-PCG recordings to ensure accurate ground truth labeling. These data enabled the fine-tuning of the ECG model and the creation of 2 additional models in the ensemble that utilize PCG as well, enabling the low EF CNN model to accept both single-lead ECG and PCG inputs.

The final model is an ensemble of 3 models (ECG model, PCG model, and PCG-ECG model). The input to the PCG model is the one-dimensional time-series or PCG waveform at 2,000 Hz, and the input to the PCG-ECG model is synchronously captured PCG and ECG waveforms. All 3 models have a ResNet<sup>13</sup> style architecture and they all have 2 nodes in the output layer, one representing the probability of "low ejection fraction," and one representing the probability of "normal ejection fraction" (which is 1 - probability of "low ejection fraction"). The probability outputs from the 3 models were averaged to get the final probability output of the ensemble model. A simple threshold was then applied on the final averaged probability output. If the model output for low EF is equal or greater than the threshold, the output of the Ejection Classification Algorithm is "low ejection fraction," otherwise the output is "normal ejection fraction." The probability threshold for the final



The Eko DUO digital stethoscope is placed on the chest in a modified lead ii orientation, with the device positioned along the midline of the sternum, extending diagonally toward the left hip. This placement is designed to optimize the quality of both the ECG and phonocardiogram (PCG) signals for accurate ejection fraction analysis. ECG = electrocardiogram.

ensemble model was determined using a separate internal validation set (which was not used for training the model or the final evaluation). The threshold was chosen to maximize the sum of sensitivity and specificity on the internal validation set.

The Eko Test Dataset includes paired ECG and PCG recordings and echocardiograms from 2,960 unique subjects. The data set was collected exclusively from the 4 study sites using the predefined inclusion and exclusion criteria and represents a wholly independent test set with no overlap with the training data. This test set was powered to estimate sensitivity and specificity with the lower bound of the 95% CI to be above 70%, assuming a sensitivity of 80% and a low EF prevalence of ~15%. Recruitment continued until 210 subjects with low EF were included in the data set to ensure sufficient precision for sensitivity estimation.

**STUDY PROCEDURES.** For data capture, a single 15second ECG and PCG recording was obtained using the Eko DUO digital stethoscope (Eko Health, Inc). The stethoscope was placed in a modified lead II orientation (**Figure 1**) on patients in a seated position. Previous validation work determined that this position and orientation provided the best algorithm performance.<sup>14</sup> All participants underwent echocardiography within a maximum of 7 days of the corresponding ECG and PCG recordings. 3



**EF DETERMINATION.** The EF of each person was measured by the echocardiogram machine's integrated cardiac quantification software at the time of the echocardiogram or a manual biplane Simpson's measurement. For all included patients, EF measurements were then over-read by board-certified cardiologists at each site.<sup>15</sup> The final EF determination was based on the EF provided in the clinical report and, for this study, was dichotomized as  $\leq 40\%$ or >40%. Interobserver variability of EF measurements was performed in a subset of 194 patients from 2 independent readers and was 0.706 (Cohen's kappa).

**CNN MODEL ANALYSIS**. The signal quality of the ECG and PCG recordings was assessed using a deep learning classifier previously validated against expert/cardiologist consensus, serving as the ground truth. Only recordings deemed to have good signal quality underwent EF classification. Specifically, if the ECG signal quality was assessed as poor, no EF classification result was provided, irrespective of the PCG signal quality. In cases where the ECG signal quality was deemed good, but the PCG signal quality was poor, only the ECG model was utilized. Conversely, when both the ECG and PCG signal qualities were satisfactory, an ensemble model was employed (Figure 2). The algorithm identified

instances of "low ejection fraction detected" (defined as an EF of  $\leq$ 40%) when the output of the EF classification algorithm surpassed a predetermined threshold. The potential AI outputs for low EF were "low ejection fraction detected," "normal ejection fraction detected," or "poor ECG signal."

**OUTCOMES.** The primary aim of this study was to determine the performance of an ECG-enabled digital stethoscope in identifying patients with reduced EF of  $\leq$ 40% using an algorithm derived from single-lead ECG and PCG inputs. The primary endpoints included the performance metrics sensitivity and specificity. Secondary endpoints included positive predictive value (PPV), negative predictive value (NPV), and diagnostic yield (**Central Illustration**).

**STATISTICAL ANALYSIS.** Patient demographic and clinical data for the entire cohort are presented using descriptive statistics. A comprehensive panel of algorithm performance metrics, including sensitivity, specificity, PPV, NPV, and diagnostic yield, were calculated. Diagnostic yield is defined as the percentage of recordings that the algorithm recognized as having good signal quality and which were suitable for a low EF interpretation. Exploratory subgroup analyses were conducted to assess the effects of participant characteristics such as sex, race, and medical history on algorithm performance.



Workflow and outcomes of an AI-enabled digital stethoscope for early detection of asymptomatic left ventricular systolic dysfunction. The device captures single-lead ecg and heart sound data, which is rapidly analyzed via a cloud-based algorithm to identify patients with ejection fractions (EF)  $\leq$ 40%. The convolutional neural network model showed an AUROC of 0.85, with 77.5% sensitivity and 78.3% specificity, demonstrating its potential as a scalable, noninvasive tool for early heart failure diagnosis. Abbreviations as in Figures 1 to 3.

| TABLE 1 Demographic and Medical Information of Enrolled Patients |                    |                      |                        |                      |  |  |
|--|--------------------|----------------------|------------------------|----------------------|--|--|
|  | All<br>(N = 2,960) | EF ≤40%<br>(n = 210) | EF >40%<br>(n = 2,750) | P Value <sup>a</sup> |  |  |
| Age, y   | 66 (56-75)         | 69 (59-77)           | 66 (55-74)             | 0.006                |  |  |
| Female   | 1,508 (51)         | 70 (33.3)            | 1,439 (52.3)           | < 0.001              |  |  |
| BMI, kg/m <sup>2</sup>   | 30.3 (26.0-35.2)   | 29 (25.5-33.5)       | 30.4 (26.0-35.2)       | 0.02                 |  |  |
| Ejection fraction, %   | 60 (53-86)         | 34 (30-37)           | 60.5 (55-65)           | N/A                  |  |  |
| Race   |                    |                      |                        | 0.05                 |  |  |
| White  | 2,011 (67.9)       | 127 (60.5)           | 1,884 (68.5)           |                      |  |  |
| Black or African American  | 748 (25.3)         | 66 (31.4)            | 682 (24.8)             |                      |  |  |
| American Indian or Alaska Native                                 | 28 (0.95)          | 0 (0)                | 28 (1.0)               |                      |  |  |
| Asian  | 21 (0.71)          | 1 (0.5)              | 20 (0.7)               |                      |  |  |
| Native Hawaiian or other Pacific Islander                        | 5 (0.17)           | 1 (0.5)              | 4 (0.1)                |                      |  |  |
| Other  | 145 (5.0)          | 15 (7.1)             | 132 (4.8)              |                      |  |  |
| Prevalence of comorbidities <sup>b</sup>                         |                    |                      |                        |                      |  |  |
| Hypertension   | 2,296 (77.6)       | 177 (84.3)           | 2,119 (77.1)           | 0.02                 |  |  |
| Coronary artery disease  | 1,053 (35.6)       | 114 (54.3)           | 939 (34.1)             | <0.001               |  |  |
| Diabetes mellitus  | 800 (27.0)         | 76 (36.2)            | 724 (26.3)             | 0.002                |  |  |
| Cardiomyopathy <sup>c</sup>                                      | 489 (16.5)         | 116 (55.2)           | 373 (13.6)             | <0.001               |  |  |
| Obesity (BMI>30 kg/m <sup>2</sup> )                              | 1,525 (51.5)       | 90 (42.9)            | 1,435 (52.2)           | 0.01                 |  |  |
| Permanent atrial fibrillation                                    | 74 (2.5)           | 12 (5.7)             | 62 (2.3)               | 0.004                |  |  |
| Chronic obstructive pulmonary disease                            | 289 (9.8)          | 29 (13.8)            | 260 (9.5)              | 0.05                 |  |  |
| Renal failure  | 258 (8.7)          | 29 (13.8)            | 260 (9.5)              | 0.12                 |  |  |
| Obstructive sleep apnea  | 429 (14.5)         | 26 (12.4)            | 403 (14.7)             | 0.42                 |  |  |
| Left bundle branch block   | 155 (5.2)          | 38 (18.1)            | 117 (4.3)              | <0.001               |  |  |
| Receiving cardiotoxic drugs                                      | 26 (0.88)          | 0 (0)                | 26 (0.95)              | 0.30                 |  |  |
| Aortic stenosis  | 157 (5.3)          | 10 (4.8)             | 147 (5.3)              | 0.84                 |  |  |
| Myocardial infarction  | 379 (12.8)         | 56 (26.7)            | 323 (11.7)             | <0.001               |  |  |

Values are median (IQR) or n (%). <sup>a</sup>Chi-squared for continuous variables, Wilcoxon rank-sum for categorical variables. <sup>b</sup>The total exceeds 100% because some subjects had more than one risk factor. <sup>-</sup>Cardiomyopathy refers to diagnosis of any disease of the heart muscle, including heart failure.

 $\mathsf{EF}=\mathsf{ejection}\ \mathsf{fraction}.$ 

Finally, as an exploratory analysis, we tested the performance of both an "augmented" model that included low EF AI output and pertinent clinical variables known to be associated with heart failure with reduced EF (ie, age, biological sex, body mass index, hypertension, diabetes, coronary artery disease, and presence of a pacemaker<sup>16-21</sup>) and a "clinical" model, using the same clinical parameters, but without low EF AI output. Backward selection was employed to identify the most relevant variables for model inclusion. Model performance, as indicated by area under the receiver operating characteristic curve (AUROC), was compared to that of the "baseline" model (ie, univariable prediction of reduced EF  $\leq$ 40% based on low EF AI output alone) using DeLong's test. A P value of < 0.05 was considered statistically significant. Statistical analyses were performed using RStudio (2023.6.1.524).

# RESULTS

**DEMOGRAPHIC DATA**. From October 2020 to May 2022, 2,988 patients (median age 66 years [IQR: 56-75], 51% female) were enrolled. Due to missing

echocardiographic EF calculation, 28 subjects from the 4 sites were excluded from the study, resulting in a final cohort of 2,960. A total of 210 patients (7.1%) had an EF of  $\leq$ 40%, as identified by their study echocardiogram. Detailed demographic information can be seen in Table 1. The breakdown by participating site can be seen in Supplemental Table 1. The majority of participants were referred through primary care (55.2%), followed by general cardiology (18.2%), emergency care (11.7%), other (13.6%), and urgent care (0.24%); for the remaining 1.0%, information was not available. The most commonly identified comorbidities were hypertension (2,296, 77.6%), obesity (1,525, 51.5%), coronary artery disease (1,053, 35.6%), and diabetes (800, 27.0%) (Table 1). The inclusion criteria were patients scheduled for an echocardiogram for any reason. Among this cohort, the most common symptomatic indication for the echocardiogram was dyspnea (1,248, 42%).

**PERFORMANCE OF LOW EF AI.** The overall performance of the EF classification algorithm resulted in an AUROC of 0.852 (95% CI: 0.826-0.878), sensitivity of 77.5% (95% CI: 70.7-83.4), specificity of 78.3%



(95% CI: 76.7-79.9), PPV of 20.3% (95% CI: 17.3-23.5), NPV of 98.0% (95% CI: 97.3-98.6), and a diagnostic yield of 90.7% (95% CI: 89.6-91.7). An augmented logistic regression model including age, sex, presence of a pacemaker, and body mass index in addition to low EF AI output increased the AUROC slightly yet significantly from 0.85 to 0.87 (P < 0.001). The presence of clinical variables alone (ie, without the AI output included in the model) resulted in a significant reduction in AUROC to 0.710 (P < 0.001) (Figure 3, Table 2).

**EXPLORATORY SUBGROUP ANALYSES. By patient characteristic.** A subanalysis demonstrating the performance of the low EF AI for subjects with specific cardiovascular conduction disorders is provided in **Table 3**. These patients showed a decreased specificity compared to other patients. No significant differences were observed, although the subgroups were small with wide confidence intervals. Further subgroup analyses were conducted to assess performance as stratified by age, biological sex, and race (**Table 4**). Specificity varied slightly across biological

| TABLE 2 Performance Metrics of Low EF AI Classification Against Ground Truth of Echocardiography  |                  |                  |                  |                  |                  |  |  |
|---|------------------|------------------|------------------|------------------|------------------|--|--|
|   | Sensitivity      | Specificity      | PPV              | NPV              | Diagnostic Yield |  |  |
| Low EF AI   | 77.5 (70.7-83.4) | 78.3 (76.7-79.9) | 20.3 (17.3-23.5) | 98.0 (97.3-98.6) | 90.7 (89.6-91.7) |  |  |
| Values are % (95% Cl).<br>AI = artificial intelligence; NPV = negative predictive value; PPV = positive predictive value; other abbreviation as in Table 1. |                  |                  |                  |                  |                  |  |  |

sex (male: 73.6% [95% CI: 71.3-75.8] vs female: 81.8% [95% CI: 79.7-83.9]) and was lower for patients aged >70 years (70.7% [95% CI: 67.7-73.6]) vs the entire cohort (78.3% [95% CI: 76.7-79.9]). No other significant differences were observed.

False positives. An in-depth analysis of the false positives revealed that approximately 25% (131/543) had an EF between 41% and 49% on echocardiography, and hence closer to the designated threshold. Furthermore, 342 (63%) had a history or presence of at least one of the following: (permanent) atrial fibrillation, left bundle branch block, a pacemaker, and/or wide QRS complex (>120 ms). The rates of conduction/rhythm abnormalities for both false positives and true negatives are shown in Table 5.

# DISCUSSION

With this study, we aimed to prospectively validate the real-world potential of a novel digital stethoscope coupled with real-time AI analytics to identify patients with reduced EF (≤40%). Using single-lead ECG and PCG inputs collected from the stethoscope during a single 15-second recording, we demonstrated an AUROC of 0.85. This external, real-world validation of this technology supports its potential value as a noninvasive, scalable, and relatively low-cost method for identifying those at increased risk of reduced EF. Earlier diagnosis of a previously unrecognized low EF would allow for the earlier initiation of therapies proven to improve quality of life, decrease hospitalizations, and prolong survival.<sup>22</sup>

In our study, 210 (7.1%) of patients were found to have an EF  $\leq$ 40%, consistent with prevalence rates from previous studies involving higher risk populations.<sup>23-25</sup> The real-world implementation of the technology in this population demonstrated discriminatory capabilities comparable to other commonly utilized screening tests, such as mammography for breast cancer and cervical cytology for cervical cancer.<sup>1</sup> An "augmented" logistic regression model including age, sex, presence of a pacemaker, and body mass index, in addition to low EF AI output, increased the AUROC from 0.85 to 0.87.

Arguably, there are several potential EF cutoff thresholds that could be of value when screening a population, with other solutions selecting EF of 35% and still others EF of 50%.<sup>1,26</sup> We selected EF  $\leq$ 40% after input from clinical experts and regulatory agencies. Also, this matches the American Heart Association/American College of Cardiology definition of heart failure with reduced EF, making it immediately actionable with strong evidence for the benefit of multiple guideline-directed medical therapies.<sup>27</sup> The operating point of the US Food and Drug Administration-cleared low EF prediction model studied here was selected to optimize the sensitivity and specificity. When implemented in the setting of the current study, sensitivity and specificity were satisfactory, at ~78% for both. With the low prevalence of individuals having an EF  $\leq$ 40%, this resulted in a PPV of only  $\sim 20\%$ . Of note, one in 4 patients with a false positive low EF alert was found to have mildly reduced EFs between 40% and 50%. The NPV of the model was excellent at 98.0%. As more knowledge is gained through widespread implementation, consideration of the risk characteristics of the population being screened and the goals of screening will play a critical role in optimizing the technology further.

|                               | Recordings | Sensitivity      | Specificity      | PPV              | NPV              | Diagnostic Yield |
|-------------------------------|------------|------------------|------------------|------------------|------------------|------------------|
| Permanent atrial fibrillation | 74         | 100.0 (71.5-NA)  | 50.0 (35.8-64.2) | 29.7 (15.9-47.0) | 100.0 (86.8-NA)  | 85.1 (75.0-92.3) |
| Left bundle branch block      | 155        | 93.9 (79.8-99.3) | 28.2 (19.7-37.9) | 29.5 (21.0-39.2) | 93.5 (78.6-99.2) | 87.7 (81.5-92.5) |
| Pacemaker                     | 278        | 87.3 (76.5-94.4) | 41.1 (33.8-48.8) | 34.8 (27.4-42.8) | 90.0 (81.2-95.6) | 85.6 (80.9-89.5) |

Abbreviations as in Table 2.

| TABLE 4 Performance Metrics Across Exploratory Subgroup Analyses    |            |                  |                  |                  |                  |                  |
|---|------------|------------------|------------------|------------------|------------------|------------------|
|   | Recordings | Sensitivity      | Specificity      | PPV              | NPV              | Diagnostic Yield |
| Biological sex  |            |                  |                  |                  |                  |                  |
| Male  | 1,451      | 79.5 (71.0-86.4) | 74.4 (71.8-76.8) | 23.5 (19.4-28.0) | 97.3 (96.1-98.3) | 89.5 (87.8-91.1) |
| Female  | 1,509      | 73.8 (60.9-84.2) | 81.9 (79.7-83.9) | 15.8 (11.8-20.6) | 98.5 (97.6-99.2) | 91.8 (90.3-93.1) |
| Age (y)   |            |                  |                  |                  |                  |                  |
| 18-30   | 94         | 100 (15.8- nan)  | 87.1 (78.0-93.4) | 15.4 (1.9-45.4)  | 100 (95.1-nan)   | 92.6 (85.3-97.0) |
| 31-50   | 400        | 68.8 (41.3-89.0) | 87.0 (83.1-90.4) | 19.3 (10.0-31.9) | 98.4 (96.3-99.5) | 92.7 (89.8-95.1) |
| 51-70   | 1,274      | 73.6 (61.9-83.3) | 81.6 (79.2-83.9) | 20.9 (16.0-26.4) | 97.9 (96.8-98.7) | 91.4 (89.8-92.9) |
| >70   | 1,192      | 81.8 (72.2-89.2) | 70.7 (67.7-73.6) | 20.2 (16.1-24.7) | 97.7 (96.3-98.7) | 89.0 (87.1-90.7) |
| Race  |            |                  |                  |                  |                  |                  |
| Black/African American  | 748        | 72.2 (58.4-83.5) | 78.6 (75.2-81.8) | 22.7 (16.6-29.7) | 97.0 (95.1-98.3) | 90.4 (88.0-92.4) |
| White   | 2,011      | 81.7 (73.1-88.4) | 77.4 (75.4-79.4) | 18.6 (15.2-22.4) | 98.5 (97.7-99.1) | 91.0 (89.7-92.3) |
| Other   | 147        | 64.3 (35.1-87.2) | 85.3 (77.6-91.2) | 34.6 (17.2-55.7) | 95.2 (89.1-98.4) | 88.4 (82.1-93.1) |
| Asian   | 21         | 100.0 (2.5-NA)   | 88.2 (63.6-98.5) | 33.3 (0.8-90.6)  | 100.0 (78.2-NA)  | 85.7 (63.7-97.0) |
| American Indian or<br>Alaska Native                                 | 28         | -                | 92.3 (74.9-99.1) | 0.0 (NA-84.2)    | 100.0 (85.8-NA)  | 92.9 (76.5-99.1) |
| Values are n or % (95% CI).<br>Abbreviations as in <b>Table 2</b> . |            |                  |                  |                  |                  |                  |

There is still a great deal to learn from the situation where the model predicts a low EF but echocardiography finds an EF of >40%. Beyond the fact that many have an EF between 40 and 50%, longitudinal follow-up in some studies of false positive patients has found a 4-fold increased risk of developing a low EF over the ensuing years relative to those with a true negative finding.<sup>3</sup> In addition, it will be important to more thoroughly explore other clinical features that influence the performance of the model, such as underlying atrial fibrillation, presence of a pacemaker, or left bundle branch block, as identified in this study, in order to better optimize its performance.

Additionally, the usage of PCG for the detection of low EF in this study is new. The PCG model that is part of the final ensemble used for this study, by itself, is less predictive for low EF compared to the ECG model alone. However, PCG models based on new neural network architectures have been shown to have performance closer to ECG models.<sup>12</sup> Understanding what features of the PCG and/or ECG are triggering predictions for low EF is the subject of future work. To make this possible, new machine learning (ML) approaches and analysis methods are required.

Currently, approximately 50 cardiovascular AI/ML medical devices have received US Food and Drug Administration clearance.<sup>28</sup> Despite the increased focus of the scientific research community on health care-related AI research, there is a paucity of data available from prospective validation studies for US Food and Drug Administration-cleared AI/ML

cardiovascular devices.<sup>29</sup> Previous algorithms have shown promising performance in detecting LVSD, with area under the curve values ranging from 0.83 to 0.97. The majority of these studies were based on 12lead ECGs obtained as part of routine clinical care.<sup>1,30</sup> More recent studies have explored reduced-lead ECG (ie, single-lead) approaches. For example, Khunte et al<sup>31</sup> reported a novel strategy adapting AI for noisy single-lead ECGs obtained from wearable and portable devices, achieving an AUROC of 0.87 for the detection of EF <40% on noise-augmented ECGs. Another study demonstrated the feasibility of detecting EF ≤40% using single-lead ECGs from smartwatches, achieving an AUROC of 0.89.32 Additionally, Bachtiger et al14 validated the use of AI applied to single-lead ECGs recorded with an ECG-

| TABLE 5Rates of Comorbidities in the True Negative Vs FalsePositive Cohorts  |                                  |                                 |                      |  |  |  |
|--|----------------------------------|---------------------------------|----------------------|--|--|--|
|  | True<br>Negatives<br>(n = 1,963) | False<br>Positives<br>(n = 543) | P Value <sup>a</sup> |  |  |  |
| EF: 41%-49%  | 133 (6.8%)                       | 131 (25%)                       | < 0.001              |  |  |  |
| Permanent AF   | 26 (1.3%)                        | 26 (4.8%)                       | < 0.001              |  |  |  |
| AF   | 126 (6.4%)                       | 108 (19.9%)                     | < 0.001              |  |  |  |
| LBBB   | 29 (1.5%)                        | 74 (13.6%)                      | < 0.001              |  |  |  |
| Pacemaker  | 72 (3.7%)                        | 103 (19.0%)                     | < 0.001              |  |  |  |
| Wide QRS complex <sup>b</sup><br>(>120 ms)   | 247 (12.6%)                      | 266 (49.0%)                     | <0.001               |  |  |  |
| Values are n (%). <sup>a</sup> Chi-squared test. <sup>b</sup> Either in addition to LBBB and pacemaker or isolated.<br>AF = atrial fibrillation; LBBB = left bundle branch block; other abbreviation as in Table 1 |                                  |                                 |                      |  |  |  |

enabled digital stethoscope, an earlier version of the device used in this study, achieving an AUROC of 0.85 in a multicenter real-world clinical setting. There is further promise for the application of these technologies in unique populations, such as those with peripartum cardiomyopathy (PPCM). A study conducted in Nigeria, which has the highest reported incidence of PPCM worldwide, demonstrated that AI-guided screening with a digital stethoscope significantly improved the diagnosis of pregnancy-related cardiomyopathy. In that study, the AI paired with a digital stethoscope had an AUROC of 0.98 for detecting EF <50% and 0.99 for detecting EF <40%, doubling the number of diagnosed cases compared to usual care.<sup>33,34</sup> Interestingly, a significant increase in PPCM diagnosis was only observed for the AI-enabled single-lead ECG digital stethoscope model and not the 12-lead-based algorithm that was simultaneously studied.<sup>34</sup> This was perhaps because the selected cutoff for determination of the primary outcome, where LVSD was defined as LVEF <50%, was not in line with the categorization used during the 12-lead model derivation (trained to detect an LVEF  $\leq 35\%$ ).<sup>1</sup> The potential for AI-guided screening using digital stethoscopes in lower-income countries, where access to advanced imaging modalities is limited, highlights the broader applicability and impact of this technology for early detection and management of cardiac conditions.35 Recent work has suggested that combining this tool with AI-guided echocardiographic imaging can lead to screening solutions at scale.35

**STUDY LIMITATIONS.** Despite the promising results of this study, there are several limitations that need to be acknowledged. Firstly, while the discriminatory power, sensitivity, and specificity of the algorithm were good, there is substantial room for improvement. We believe that the addition of multiple ECG leads and/or the addition of clinical characteristics or other biomarkers such as BNP and NTproBNP could further optimize the identification of individuals with an undiagnosed low EF by potentially both improving sensitivity and PPV. Another limitation, shared by almost all ML models, is that the exact features of the ECG and PCG leading to the prediction of a low versus normal EF are not readily identifiable. In addition, the accuracy of the ground truth EF data from echocardiography and potential biases therein can influence the performance and reliability of the algorithm, although the influence of any variability in EF determination would likely only impact those with EFs close to (eg,  $\pm 5\%$ ) 40%. Furthermore, as the study population was comprised of patients already referred for an echocardiogram, the generalizability of our findings to the broader population may be limited. Finally, longitudinal studies examining patient outcomes are required to assess the clinical utility and cost-effectiveness of the diagnostic tool. Such studies would provide valuable insights into the long-term effectiveness and impact on patient management strategies.

# CONCLUSIONS

In conclusion, our study demonstrates the potential of a CNN model using single-lead ECG and PCG inputs collected with a digital stethoscope as a noninvasive tool for the detection of EF  $\leq$ 40%. The algorithm showed good performance, with sensitivity and specificity comparable to other common medical screening tests. While our findings highlight the promise of this technology for improving detection rates, further research is needed to evaluate its impact on clinical outcomes and cost-effectiveness. Future work will focus on adapting the model for 3-lead input and optimizing algorithm performance using transformer models.

# FUNDING SUPPORT AND AUTHOR DISCLOSURES

Authors Guo, Kieu, Mathew, Prince, Lastowski, McDonough, Currie, Maidens, and Venkatraman are affiliated with Eko Health, Inc that develops digital stethoscopes, software, and algorithms to detect cardiovascular and pulmonary disease; hold company equity; and were responsible for algorithm development, collated data analysis, and codrafting of the manuscript. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Salima Qamruddin, Edward D. Frohlich Endowed Mentorrship position, Technical Director Echocardiography Laboratory, Ochsner Medical Center, Director Women's Cardiovascular Wellness Clinic, Ochsner Heart and Vascular Institute, Assistant Professor, Ochsner Clinical School, University of Queensland, Australia, 1514 Jefferson Highway, New Orleans, Louisiana 70121, USA. E-mail: salimaq5@gmail.com OR salima. qamruddin@ochsner.org.

### PERSPECTIVES

#### **COMPETENCY IN PATIENT CARE AND**

**PROCEDURAL SKILLS:** The CNN model using a digital stethoscope enables earlier detection of patients with reduced EF (EF  $\leq$  40%), facilitating timely intervention and management of left ventricular systolic dysfunction. This technology provides a noninvasive and accessible method for screening, which can be easily integrated into various clinical settings, including primary care and remote healthcare environments. The ability to identify at-risk patients early has the potential to reduce downstream healthcare costs associated with late-stage heart failure treatment.

**TRANSLATIONAL OUTLOOK:** For successful clinical translation, several factors need to be addressed. Further research is necessary to streamline the integration of this

CNN model into existing clinical workflows, ensuring it complements current diagnostic procedures without causing disruptions. Additionally, validating the algorithm's performance across diverse populations and different healthcare settings is crucial to ensure its generalizability and effectiveness. Addressing the clinical significance of false positives, particularly those with EF between 41% and 49%, will help refine the algorithm and enhance its predictive accuracy. Providing adequate training for healthcare providers to use this technology effectively will be essential for widespread adoption. Lastly, conducting longitudinal studies to assess the impact of early detection using this CNN model on longterm patient outcomes, including morbidity and mortality rates, will be vital in demonstrating its overall clinical utility.

#### REFERENCES

**1.** 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):70-74. https://doi.org/ 10.1038/s41591-018-0240-2

2. Ammar KA, Jacobsen SJ, Mahoney DW, et al. Prevalence and prognostic significance of heart failure stages: application of the American College of Cardiology/American Heart Association heart failure staging criteria in the community. *Circulation*. 2007;115(12):1563-1570. https://doi.org/10. 1161/circulationaha.106.666818

**3.** Wang TJ, Levy D, Benjamin EJ, Vasan RS. The epidemiology of "asymptomatic" left ventricular systolic dysfunction: implications for screening. *Ann Intern Med.* 2003;138(11):907–916. https://doi.org/10.7326/0003-4819-138-11-200306030-00012

4. Sara JD, Toya T, Taher R, Lerman A, Gersh B, Anavekar NS. Asymptomatic left ventricle systolic dysfunction. *Eur Cardiol*. 2020;15:e13. https://doi. org/10.15420/ecr.2019.14

5. Verhestraeten C, Heggermont WA, Maris M. Clinical inertia in the treatment of heart failure: a major issue to tackle. *Heart Fail Rev.* 2021;26(6): 1359-1370. https://doi.org/10.1007/s10741-020-09979-z

**6.** Heidenreich PA, Fonarow GC, Opsha Y, Sandhu AT, Sweitzer NK, Warraich HJ. Economic issues in heart failure in the United States. *J Card Fail.* 2022;28(3):453-466. https://doi.org/10. 1016/j.cardfail.2021.12.017

7. Docherty KF, Lam CSP, Rakisheva A, et al. Heart failure diagnosis in the general community - who, how and when? A clinical consensus statement of the Heart Failure Association (HFA) of the European Society of Cardiology (ESC). *Eur J Heart Fail*. 2023;25(8):1185-1198. https://doi.org/10.1002/ ejhf.2946

8. Rice T. 'The hallmark of a doctor': the stethoscope and the making of medical identity. J Mater Cult. 2010;15(3):287-301. https://doi.org/10.1177/ 1359183510373985

 Attia ZI, Dugan J, Rideout A, et al. Automated detection of low ejection fraction from a one-lead electrocardiogram: application of an AI algorithm to an electrocardiogram-enabled Digital Stethoscope. *Eur Heart J Digit Health.* 2022;3(3):373-379. https://doi.org/10.1093/ehjdh/ztac030

**10.** Prince J, Maidens J, Kieu S, et al. Deep learning algorithms to detect murmurs associated with structural heart disease. *J Am Heart Assoc.* 2023;12(20):e030377. https://doi.org/10.1161/jaha.123.030377

**11.** Arjoune Y, Nguyen TN, Salvador T, et al. StethAid: a digital auscultation platform for pediatrics. *Sensors (Basel)*. 2023;23(12):5750. https://doi.org/10.3390/s23125750

**12.** Mathew G, Barbosa D, Prince J, Venkatraman S. Foundation models for cardiovascular disease detection via biosignals from digital stethoscopes. *NPJ Cardiovasc Health.* 2024;1(1):25. https://doi.org/10.1038/s44325-024-00027-5

13. Computer Vision Foundation. URL: https://www. cv-foundation.org/openaccess/content\_cvpr\_2016/ papers/He\_Deep\_Residual\_Learning\_CVPR\_2016\_ paper.pdf

**14.** Bachtiger P, Petri CF, Scott FE, et al. Point-ofcare screening for heart failure with reduced ejection fraction using artificial intelligence during ECG-enabled stethoscope examination in London, UK: a prospective, observational, multicentre

#### study. Lancet Digit Health. 2022;4(2):e117-e125. https://doi.org/10.1016/s2589-7500(21)00256-9

**15.** Lang RM, Bierig M, Devereux RB, et al. Recommendations for chamber quantification: a report from the American Society of Echocardiography's guidelines and standards committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. J Am Soc Echocardiogr. 2005;18(12):1440-1463. https://doi.org/10.1016/j.echo.2005.10.005

16. Ho KK, Pinsky JL, Kannel WB, Levy D. The epidemiology of heart failure: the Framingham Study. J Am Coll Cardiol. 1993;22(4 Suppl A):6a-13a. https://doi.org/10.1016/0735-1097(93)90455-a

**17.** Gardin JM, Siscovick D, Anton-Culver H, et al. Sex, age, and disease affect echocardiographic left ventricular mass and systolic function in the freeliving elderly. The Cardiovascular Health Study. *Circulation*. 1995;91(6):1739–1748. https://doi. org/10.1161/01.cir.91.6.1739

**18.** Redfield MM, Jacobsen SJ, Burnett JC Jr, Mahoney DW, Bailey KR, Rodeheffer RJ. Burden of systolic and diastolic ventricular dysfunction in the community: appreciating the scope of the heart failure epidemic. *JAMA*. 2003;289(2):194–202. https://doi.org/10.1001/jama.289.2.194

 Kannel WB, Ho K, Thom T. Changing epidemiological features of cardiac failure. Br Heart J. 1994;72(2 Suppl):S3-S9. https://doi.org/10.1136/ hrt.72.2\_suppl.S3

**20.** He J, Ogden LG, Bazzano LA, Vupputuri S, Loria C, Whelton PK. Risk factors for congestive heart failure in US men and women: NHANES I epidemiologic follow-up study. *Arch Intern Med*. 2001;161(7):996-1002. https://doi.org/10.1001/ archinte.161.7.996

**21.** Massie BM, Shah NB. Evolving trends in the epidemiologic factors of heart failure: rationale for preventive strategies and comprehensive disease management. *Am Heart J.* 1997;133(6):703-712. https://doi.org/10.1016/s0002-8703(97)70173-x

**22.** Abdin A, Anker SD, Butler J, et al. 'Time is prognosis' in heart failure: time-to-treatment initiation as a modifiable risk factor. *ESC Heart Fail*. 2021;8(6):4444–4453. https://doi.org/10. 1002/ehf2.13646

23. Nielsen OW, McDonagh TA, Robb SD, Dargie HJ. Retrospective analysis of thecosteffectiveness of using plasmabrain natriuretic peptide inscreening for left ventricularsystolic dysfunction in the general population. J Am Coll Cardiol. 2003;41(1):113–120. https://doi.org/10. 1016/S0735-1097(02)02625-6

24. Boonman-de Winter LJM, Rutten FH, Cramer MJM, et al. High prevalence of previously unknown heart failure and left ventricular dysfunction in patients with type 2 diabetes. *Diabetologia*. 2012;55(8):2154–2162. https://doi.org/ 10.1007/s00125-012-2579-0

**25.** Roger VL. Epidemiology of heart failure. *Circ Res.* 2021;128(10):1421-1434. https://doi.org/10. 1161/CIRCRESAHA.121.318172

**26.** Yao X, Rushlow DR, Inselman JW, et al. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nat* 

# *Med.* 2021;27(5):815-819. https://doi.org/10. 1038/s41591-021-01335-4

**27.** Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Joint Committee on clinical practice guidelines. *Circulation*. 2022;145(18):e895-e1032. https://doi.org/10.1161/cir.000000000001063

28. Armoundas AA, Narayan SM, Arnett DK, et al. Use of artificial intelligence in improving outcomes in heart disease: a scientific statement from the American Heart Association. *Circulation*. 2024;149(14):e1028-e1050. https://doi.org/10. 1161/CIR.000000000001201

**29.** CBInsights. State of AI 2021 report. Accessed July 26, 2024. https://www.cbinsights.com/ research/report/ai-trends-2021/

**30.** Ulloa-Cerna AE, Jing L, Pfeifer JM, et al. rECHOmmend: an ECG-based machine learning approach for identifying patients at increased risk of undiagnosed structural heart disease detectable by echocardiography. *Circulation.* 2022;146(1):36-47. https://doi.org/10.1161/circulationaha.121.057869

**31.** Khunte A, Sangha V, Oikonomou EK, et al. Detection of left ventricular systolic dysfunction from single-lead electrocardiography adapted for portable and wearable devices. *NPJ Digital Med.* 2023;6(1):124. https://doi.org/10.1038/s41746-023-00869-w

**32.** Attia ZI, Harmon DM, Dugan J, et al. Prospective evaluation of smartwatch-enabled detection of left ventricular dysfunction. *Nat Med.* 2022;28(12):2497-2503. https://doi.org/10. 1038/s41591-022-02053-1

**33.** Adedinsewo DA, Morales-Lara AC, Dugan J, et al. Screening for peripartum cardiomyopathies using artificial intelligence in Nigeria (SPEC-Al Nigeria): clinical trial rationale and design. *Am Heart J.* 2023;261:64-74. https://doi.org/10.1016/j.ahi.2023.03.008

**34.** Adedinsewo DA, Morales-Lara AC, Afolabi BB, et al. Artificial intelligence guided screening for cardiomyopathies in an obstetric population: a pragmatic randomized clinical trial. *Nat Med.* 2024;30(10):2897-2906. https://doi.org/10.1038/ s41591-024-03243-9

**35.** Peck D, Rwebembera J, Nakagaayi D, et al. The use of artificial intelligence guidance for rheumatic heart disease screening by novices. *J Am Soc Echocardiogr.* 2023;36(7):724-732. https://doi.org/10.1016/j.echo.2023.03.001

**KEY WORDS** artificial intelligence, asymptomatic left ventricular systolic dysfunction, digital health, heart failure detection, reduced ejection fraction

**APPENDIX** For supplemental tables and figures, please see the online version of this paper.