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

PUBLISHED BY ELSEVIER

NEW RESEARCH PAPER

Machine Learning of ECG Waveforms to Improve Selection for Testing for Asymptomatic Left Ventricular Dysfunction Prompt

Elizabeth L. Potter, MBBS, BSc,[a,b] Carlos H.M. Rodrigues, BSc,[a,c] David B. Ascher, PhD,[a,c] Walter P. Abhayaratna, MBBS, PhD,[d,e] Partho P. Sengupta, MD,[f] Thomas H. Marwick, MBBS, PhD, MPH[a,b]

ABSTRACT

OBJECTIVES To identify whether machine learning from processing of continuous wave transforms (CWTs) to provide an "energy waveform" electrocardiogram (ewECG) could be integrated with echocardiographic assessment of subclinical systolic and diastolic left ventricular dysfunction (LVD).

BACKGROUND Asymptomatic LVD has management implications, but routine echocardiography is not undertaken in subjects at risk of heart failure. Signal processing of the surface ECG with the use of CWT can identify abnormal myocardial relaxation.

METHODS EwECG and echocardiography were undertaken in 398 participants at risk of heart failure (HF). Reduced global longitudinal strain (GLS ≤16%)), diastolic abnormalities (E/e′ >15, left atrial enlargement with E/e′ >10 or impaired relaxation) or LV hypertrophy defined LVD. EwECG feature selection and supervised machine-learning by random forest (RF) classifier was undertaken with 643 CWT-derived features and the Atherosclerosis Risk in Communities (ARIC) heart failure risk score.

RESULTS The ARIC score and 18 CWT features were selected to build a RF predictive model for LVD in a training dataset (n = 287; 60% female, median age 71 [interquartile range: 68 to 74] years). Model performance was tested in an independent group (n = 111; 49% female, median age 61 years [59 to 66 years]), demonstrating 85% sensitivity and 72% specificity (area under the receiver-operating characteristic curve [AUC]: 0.83; 95% confidence interval [CI]: 0.74 to 0.92). With ARIC score removed, sensitivity was 88% and specificity, 70% (AUC: 0.78; 95% CI: 0.70 to 0.86). RF models for reduced GLS and diastolic abnormalities including similar features had sensitivities that were unsuitable for screening. Conventional candidates for LVD screening (ARIC score, N-terminal pro-B-type natriuretic peptide, and standard automated ECG analysis) had inferior discriminative ability. Integration of ewECG in screening of people at risk of HF would reduce need for echocardiography by 45% while missing 12% of LVD cases.

CONCLUSIONS Machine learning applied to ewECG is a sensitive screening test for LVD, and its integration into screening of patients at risk for HF would reduce the number of echocardiograms by almost one-half.
(J Am Coll Cardiol Img 2021;■:■–■) © 2021 by the American College of Cardiology Foundation.

From the [a]Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia; [b]School of Public Health and Preventive Medicine, Monash University, Melbourne, Victoria, Australia; [c]School of Biomedical Sciences, Melbourne University, Melbourne, Victoria, Australia; [d]Australian National University Medical School, Australian National University, Canberra, Australian Capital Territory, Australia; [e]Division of Medicine, Canberra Hospital, Canberra, Australian Capital Territory, Australia; and the [f]West Virginia University Heart and Vascular Institute, Morgantown, West Virginia, USA.
Alan Fraser, MD, served as Guest Editor for this paper.

ISSN 1936-878X/$36.00 https://doi.org/10.1016/j.jcmg.2021.04.020

ABBREVIATIONS AND ACRONYMS

ARIC = Atherosclerosis Risk in Communities

CAD = coronary artery disease

cNRI = continuous net reclassification improvement

CWT = continuous wavelet transform

DD = diastolic dysfunction

ewECG = energy waveform electrocardiography

GLS = global longitudinal strain

HF = heart failure

IDI = integrated discrimination improvement

LVD = left ventricular dysfunction

RF = random forest

The echocardiographic recognition of structural and functional cardiac abnormalities among patients with risk factors for heart failure (HF) identifies asymptomatic left ventricular dysfunction (LVD) and thereby guides therapy (1). Despite this, routine echocardiography is rarely performed in people with HF risk factors, other than in coronary artery disease (CAD) (2). No viable screening test has evolved to better direct selection for echocardiography.

Standard 12-lead electrocardiography (ECG) is a potential initial investigation in people with HF risk factors. A range of ECG abnormalities, e.g., arrhythmias, conduction disturbances, and voltage patterns, may relate to underlying LVD. The potential for a form of signal processing, known as continuous wavelet transform (CWT), to reveal abnormalities in a standard ECG signal has been recognized for more than 20 years (3,4). However, only recently have patterns in CWT-processed ECG signals, or "energy waveform" ECG (ewECG), been demonstrated to predict functional LV abnormalities, specifically abnormal relaxation (5,6). Recording an ewECG requires no additional time or expertise and simultaneously displays a standard 12-lead ECG trace, making it a feasible test for use in the community. However, although there is an association between repolarization measures and abnormal myocardial relaxation (7), whether prior findings in populations presenting for echocardiography (6) extend to the detection of asymptomatic systolic LVD in community populations at risk for HF is unknown. Indeed, the most appropriate use of this technology would be to guide definitive echocardiographic assessment as part of a screening process. We hypothesized that machine-learning algorithms applied to ewECG data could identify LVD in a community population at risk of HF and that depolarization and repolarization CWT features were associated with systolic and diastolic dysfunction (DD), respectively.

METHODS

STUDY PARTICIPANTS. Participants were recruited from the community as part of the ongoing Victorian Study of Echocardiographic Detection of Left Ventricular Dysfunction (Vic-ELF; ACTRN 12617000116325) in Melbourne, Australia. Participants were aged ≥65 years with at least one of the following HF risk factors: obesity (body mass index ≥30 kg/m^2), type 2 diabetes mellitus, or hypertension (systolic blood pressure ≥140 mm Hg or on medication). Exclusion criteria included LV ejection fraction ≤40%, known symptomatic HF (or diagnosed at baseline screening), known CAD (excluded because of the routine use of echocardiographic assessment in this group), moderate or greater valvular heart disease, renal impairment, and symptoms of HF.

The testing dataset comprised an analogous group of 111 prospectively recruited asymptomatic people without established cardiovascular disease, with the same inclusion and exclusion criteria, in Canberra, Australia (Australian National University Medical School; n = 79), and in the USA (Icahn School of Medicine, New York, New York, and University of West Virginia, Morgantown, West Virginia; n = 32). This geographic heterogeneity aimed to test model generalizability. The relevant institutional review boards approved the study, and each participant gave written informed consent.

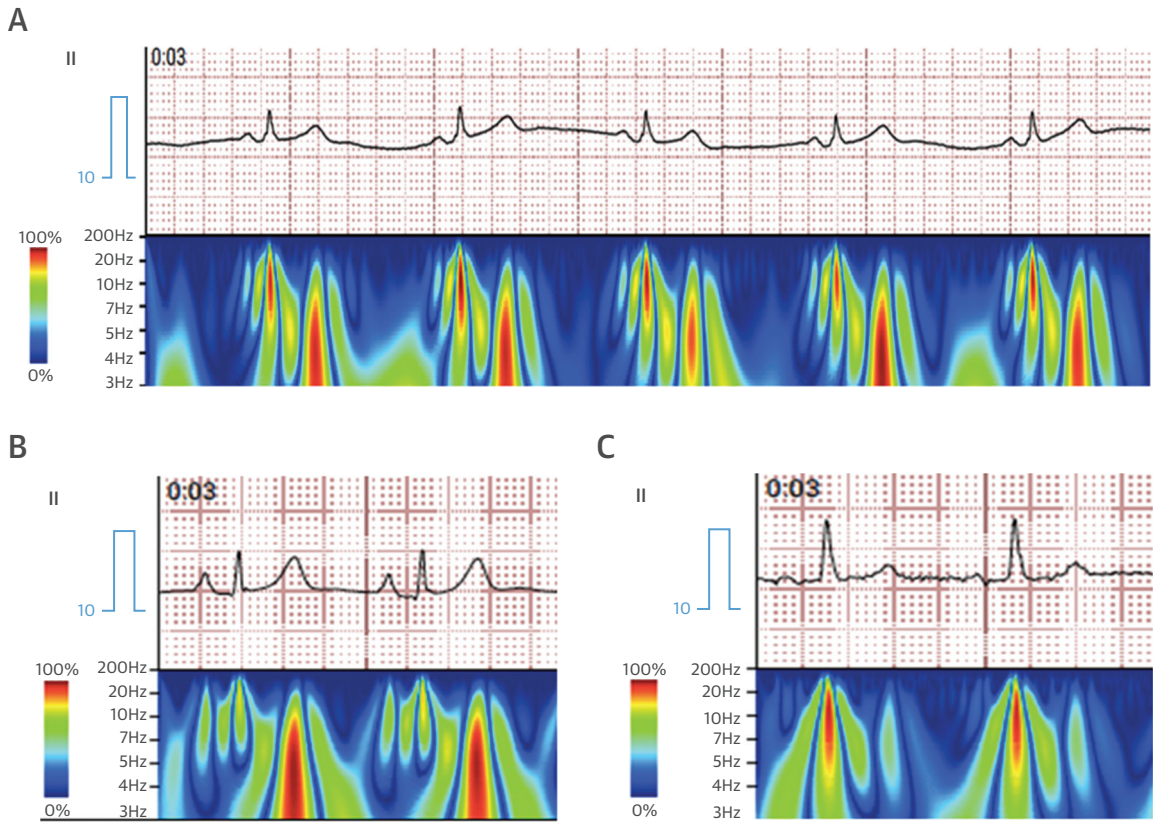
CLINICAL MEASURES. Baseline measures and procedures pertinent to this substudy included body mass index, resting averaged systolic blood pressure (SBP) and diastolic blood pressure (DBP), heart rate, documentation of cardiovascular risk factors, comorbidities, and medications. Clinical data were used to calculate the 4-year risk of incident symptomatic HF with the use of the Atherosclerosis Risk in Communities (ARIC) HF risk score, which has demonstrated utility in risk stratification in subclinical HF (8). Biochemical markers of renal function and N-terminal pro-B-type natriuretic peptide (NT-proBNP) were also measured.

STANDARD ECG AND ewECG. After standard ECG lead placement, subjects underwent ewECG evaluation with the use of a device that is CE marked but not approved for use in the USA (MyoVista version 2.0; HeartSciences, Southlake, Texas). The MyoVista ewECG interface displays a standard 12-lead ECG trace as well as an automated diagnostic interpretation based on the University of Glasgow 12-lead ECG interpretive analysis algorithm, which provides both quantitative parameters and qualitative interpretations (9,10). The ECG signal is

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.

Manuscript received December 9, 2020; revised manuscript received March 24, 2021, accepted April 8, 2021.

FIGURE 1 **Conventional ECG Traces and Corresponding ewECG Scalograms After Signal Processing Using CWT**



(A) Normal ewECG, as determined by MyoVista proprietary software and our machine-learning algorithm. Echocardiogram was normal. **(B)** Predicted abnormal by our machine-learning algorithm. Participant had abnormal systolic function (GLS 15%). Of potential significance is the lower energy associated with the QRS. **(C)** Predicted abnormal by our machine-learning algorithm. Participant had diastolic dysfunction. Note low energy associated with the T-wave. CWT = continuous wave transform; ECG = electrocardiography; ewECG = energy waveform ECG; GLS = global longitudinal strain.

deconstructed and presented graphically in an energy scalogram (red [high energy] to blue [low energy]) depicting an energy distribution by time (x-axis) and frequency (y-axis) (the "energy waveform") (Figure 1). As described previously (5), energy is expressed as coefficients reflecting agreement between wavelet and signal at varying scales, rather than a discrete energy measurement (11). A total of 643 CWT features (energies, frequencies and ratios) at defined points in the cardiac cycle are generated by proprietary software throughout the cardiac cycle. As our prior work demonstrated the existing automated interpretative algorithms to be insufficiently sensitive to detect LVD (12), we used the complete CWT output.

ECHOCARDIOGRAPHY. On the same day as ewECG, a transthoracic 2-dimensional and Doppler echocardiographic study was performed with the use of standard equipment (ACUSON SC2000; Siemens Healthcare USA, Mountain View, California) and transducer (4V1c, 1.25 to 4.5 MHz; 4Z1c, 1.5 to 3.5 MHz) in accordance with American Society of Echocardiography guidelines. LV systolic function was assessed by global longitudinal strain (GLS) computed with the use of speckle tracking (Syngo VVI; Siemens Healthcare USA). GLS was the average of regional strains in the apical 2-chamber, 4-chamber and long-axis views. Diastolic function was assessed by measuring mitral inflow peak early diastolic velocity (E), peak late diastolic velocity (A), E/A ratio, septal and lateral mitral annular early diastolic velocities (e′), and the E/e′ ratio. Left atrial volume index (LAVi) was calculated from maximal LA volume with the use of biplane images and indexed to body surface area; LA enlargement was defined as LAVi >34 ml/m^2. Left ventricular hypertrophy (LVH) was defined as LV mass index >95 g/m^2 in women and >115 g/m^2 in men.

TABLE 1 Baseline Characteristics by Subclinical Heart Failure Stage

	Stage A HF (n = 227)	LV Dysfunction (n = 171)	p Value
Clinical and biomarkers			
Age, yrs	68 (62-71)	71 (68-75)	<0.001
Sex	137 (60)	90 (40)	0.08
Hypertension	185 (82)	152 (90)	0.04
Type 2 diabetes mellitus	41 (18)	60 (35)	<0.001
Atrial fibrillation	9 (4)	13 (8)	0.12
Systolic BP, mm Hg	138 ± 15	142 ± 15	0.01
Diastolic BP, mm Hg	82 ± 9	83 ± 11	0.13
Heart rate, beats/min	63 ± 9	66 ± 10	0.002
BMI, g/m^2	31 ± 5	32 ± 6	0.09
ACE-I/ARB*	118 (75)	118 (73)	0.64
Beta-blockers*	17 (11)	25 (15)	0.22
NT-proBNP,† pg/ml	51 (30-94)	59 (33-101)	0.39
ARIC HF risk score	3.6 (1.22-6.60)	7.1 (3.80-12.90)	<0.001
Standard ECG abnormalities			
Atrial fibrillation	1 (0.4)	4 (2.3)	0.09
LBBB	0 (0.0)	3 (1.6)	0.05
LV hypertrophy	7 (3)	10 (6)	0.18
Abnormal ECG (per Glasgow analysis)	35 (15)	62 (36)	<0.001
Echocardiographic measures			
LV mass index, g/m^2	67 ± 16	71 ± 22	0.01
LV ejection fraction, %	64 ± 6	61 ± 7	<0.001
Global longitudinal strain, %	20 (18.9-21.0)	17 (15.4-18.6)	<0.001
E/A ratio	0.95 ± 0.28	0.80 ± 0.24	<0.001
Average e′, cm/s	8.1 ± 1.7	7.1 ± 1.9	<0.001
Average E/e′	8.3 (7.2-9.8)	9.3 (7.3-11.7)	0.003
LAVI, ml/m^2	30 (25-34)	37 (29-42)	<0.001

Values are median (interquartile range), n (%), or mean ± SD. *Not available in the Canberra group. †Available only in the training dataset.

ACE-I/ARB = angiotensin converting enzyme inhibitor/receptor blocker; ARIC = Atherosclerosis Risk In Communities; BMI = body mass index; BP = blood pressure; ECG = electrocardiography; HF = heart failure; LAVI = left atrial volume indexed to body surface area; LBBB = left bundle branch block; LV = left ventricular; NT-proBNP = N-terminal pro-B-type natriuretic peptide.

LVD was defined by either: 1) abnormal structure (LVH); 2) abnormal GLS ≤16% or borderline GLS (17%-18%) with impaired relaxation or left atrial enlargement (LAE); or iii) diastolic dysfunction (E/e′ >15 or E/e′ >10 with LAE or impaired relaxation with LAE).

We developed predictive models for: 1) LVD' 2) systolic dysfunction (GLS ≤16%); and 3) diastolic dysfunction.

MACHINE-LEARNING CLASSIFICATION MODEL. A supervised machine-learning approach was used to predict LVD status. We used the random forest (RF) classifier algorithm in the module Scikit-learn (Python Software Foundation, https://www.python.org) (13). All hyperparameters for the algorithm were entered according to the library's documentation and are summarized in Supplemental Table 5. Given the high dimensionality of the data, including the fact that ewECG features were more numerous than subjects, we undertook a process of feature selection to identify those features with the most predictive information relevant to LVD (Supplemental Figure 1). All CWT features plus the ARIC HF risk score were offered in feature selection. The ARIC score was included to evaluate the importance of this easily attainable clinical variable against ewECG.

The feature selection approach evaluated the performance of all individual features (*N*), using area under the receiver-operating characteristic curve (AUC), and selected the best performing in the first round. Then, each of the remaining (*N* − *1*) features was paired with the selected one to identify the pair that gave the best performance. This process was repeated until all features were selected and the combination of features that provided the best performance was ascertained. Other approaches were tested but were less successful (Supplemental Table 1). We also evaluated feature importance, extracted from the RF output (which uses Gini importance). This indicates the contribution made by each feature to the model's predictive performance. Practically, higher importance means that the decision-making error associated with a feature in the nodes across all decision trees in the forest is less than using other features in other nodes.

A 5-fold cross validation on the training dataset was used to internally validate model performance with subsequent external validation on the separate/test dataset. The output of the RF model is a continuous probability score with a threshold of 50% for dichotomizing predicted outcome, e.g., LVD versus no LVD. When evaluating the performance on the external dataset, modification of this probability threshold was investigated to see if performances could be optimized. Unless otherwise stated, the 50% threshold was found to be optimal.

STATISTICAL ANALYSIS. Continuous data are presented as mean ± SD or median and interquartile range (IQR) depending on distribution after visual assessment. Between group differences for categoric data were tested with the use of Pearson's chi-square, and for continuous variables the independent *t* test or Wilcoxon rank-sum test was used depending on normality of distribution. Because the most important characteristic of a screening test is high sensitivity, cutoff points were selected with the minimal number of false positives at a sensitivity closest to 90%. Discriminatory performance predictive models were assessed with the use of AUCs. To evaluate the incremental utility of the machine-learning models compared with the ARIC HF risk score as a base model, continuous net reclassification improvement (cNRI) and integrated discrimination improvement

(IDI) were calculated. cNRI measures improvements in probabilities within events (i.e., increased probability) and nonevents (i.e., decreased probabilities), with the addition of, in this case, the machine-learning models (14). IDI reflects the difference in discrimination slopes (probabilities for events minus nonevents) between 2 models and is reported herein as the absolute IDI (15). For all analyses, statistical significance was defined as a 2-tailed p value < 0.05. Analyses were conducted with the use of Stata 15.1 (StataCorp, College Station, Texas).

RESULTS

PARTICIPANTS. Overall we included 398 participants (57% female, median age 69 (IQR 66-73) years), and of these, 171 (43%) had LVD. Baseline characteristics by HF stage are presented in Table 1. Compared with people with only risk factors, LVD was associated with older age, a higher proportion of hypertension, type 2 diabetes, increased heart rate and SBP, and higher ARIC HF risk score. The proportions of abnormal Glasgow ECG analysis summaries were 15% and 36% for people with risk factors and LVD, respectively ($p < 0.001$). All echocardiographic measures differed significantly between HF stages.

PREDICTION OF LVD BY CONVENTIONAL METHODS. The ARIC HF risk score had an AUC of 0.72 (95% CI: 0.67 to 0.77) for LVD discrimination. An optimized cutoff point for sensitivity was identified as an ARIC HF risk score of 2.6, providing 90% sensitivity and 40% specificity. Similarly, the AUC for NT-proBNP was 0.53, with an optimized cutoff of 21 pg/ml providing a sensitivity of 88% and specificity of 14%. Finally, an abnormal ECG by Glasgow analysis had a sensitivity of 36% and a specificity of 85%. In those with available NT-proBNP, adding NT-proBNP to the ARIC HF risk score (AUC 0.65 [95% CI: 0.59 to 0.71]) did not significantly improve discriminatory ability versus ARIC alone (AUC 0.63 [95% CI: 0.56 to 0.69]; $p = 0.18$). Furthermore, the addition of both NT-proBNP and abnormal ECG by Glasgow analysis did not significantly improve discriminatory ability (AUC 0.67 [95% CI: 0.61 to 0.74]; $p = 0.06$) (Supplemental Figure 2).

PREDICTION OF LVD BY RF CLASSIFIER USING ewECG. Of the 398 subjects, 287 (72%) were used to train the RF prediction model and 111 (28%) were used to test model performance. Compared with the training dataset, subjects in the test dataset were significantly younger and there were lower proportions of women and participants with hypertension and diabetes. Furthermore, SBP, DBP, and ARIC HF risk were significantly lower. The prevalence of the LVD composite was 23% in the test dataset compared with 51% in the training dataset ($p < 0.001$), and a similar pattern was observed for abnormal GLS and diastolic abnormalities (Table 2).

TABLE 2 Baseline and Outcome Characteristics by Training Versus Test Dataset

	Training (n = 287)	Test (n = 111)	p Value
Clinical			
Age, yrs	71 (68-74)	61 (59-66)	<0.001
Sex	171 (60)	54 (49)	0.05
Hypertension	252 (88)	85 (77)	0.005
Type 2 diabetes mellitus	92 (32)	9 (8)	<0.001
Systolic BP, mm Hg	142 ± 14	134 ± 16	<0.001
Diastolic BP, mm Hg	84 ± 9	79 ± 10	<0.001
Heart rate, beats/min	65 ± 9	64 ± 10	0.48
BMI, g/m^2	32 ± 6	31 ± 5	0.23
ARIC HF risk score	6.3 (3.8-10.6)	1.2 (0.8-2.6)	<0.001
Standard ECG abnormalities			
Atrial fibrillation	5 (1.7)	0 (0.0)	0.16
LBBB	3 (1)	0 (0.0)	0.28
LV hypertrophy	15 (5.0)	2 (1.8)	0.13
Abnormal ECG (per Glasgow analysis)	81 (28)	16 (14)	0.004
Echocardiographic measures			
LV mass index, g/m^2	68 ± 20	69 ± 17	0.53
LV hypertrophy	17 (6)	4 (4)	0.35
LV ejection fraction, %	63 ± 7	62 ± 5	0.65
GLS, %	19 (17-20)	20 (18-21)	<0.001
GLS ≤16%	54 (19)	7 (6)	0.002
E/A ratio	0.83 ± 0.22	1.03 ± 0.33	<0.001
Average e′, cm/s	7.8 ± 1.9	7.4 ± 1.7	0.03
Average E/e′	8 (7-10)	9 (8-11)	0.001
LAVI, ml/m^2	35 (30-41)	25 (22-31)	<0.001
Diastolic abnormality	77 (27)	14 (13)	0.002
LV dysfunction	146 (51)	25 (23)	<0.001

Values are median (interquartile range), n (%), or mean ± SD.
GLS = global longitudinal strain; other abbreviations as in Table 1.

The ARIC HF risk score was selected during feature selection along with 18 CWT features to train an RF model (Table 3). At a probability threshold of 0.51 (optimized for sensitivity), the sensitivity and specificity of the model for prediction of LVD on the test dataset were 85% and 72%, respectively (AUC: 0.83 [95% CI: 0.74 to 0.92]) (Table 4). With ARIC removed from the model, the optimized sensitivity and specificity for detection of LVD were 88% and 70%, respectively (AUC: 0.78 [95% CI: 0.67 to 0.88]; $p = 0.32$ for difference between models) (Table 3 and Figure 2). For a prevalence of 43%, this corresponded to a negative predictive value of 89% (95% CI: 78% to 95%) and positive predictive value of 69% (95% CI: 60% to 77%).

Incremental improvements in prediction were seen for both RF models compared with the ARIC HF risk score alone, as assessed by cNRI and IDI. For the RF

TABLE 3 Model Features and Relative Importance (as a Proportion of 1) for Predicting LV Dysfunction

CWT Feature Description	ECG Lead(s)	Variable Importance
Repolarization late measure (RV) below specified threshold		0.01
Repolarization early measure (RV) below specified threshold		0.01
Repolarization late measure ratio exceeded for RV/LV		0.002
Sum of Depolarization measures from Q, R, and S waves divided by heart rate	aVF	0.1
Early repolarization is too low (below specified threshold)	V5	0.098
Frequency during minimum energy in early repolarization	V5, V6	0.056, 0.055
Minimum energy at peak repolarization	II	0.095
Frequency during maximum energy in early repolarization	V6	0.09
Frequency during minimum energy in late repolarization	V4, II	0.051, 0.065
Frequency during Repolarization late measure	I, aVF	0.086, 0.092
Polarity of R wave	aVR	0.012
Power spectrum (harmonic) amplitude of the first 4 harmonic peaks is too low	II, V1	0.008, 0.01
Power spectrum (harmonic) amplitudes of the first 4 peaks is too high	II, III, V4, aVR	0.005, 0.006, 0.005, 0.003
Power spectrum (harmonic) 4^{th} peak amplitude is greater than the 3^{rd} peak	aVF, V5	0.018, 0.015
Power spectrum (harmonic) 5^{th} peak amplitude is greater than the 1^{st} peak	II, III, V1, V3	0.013, 0.014, 0.017, 0.013
Power spectrum (harmonic) 5^{th} peak amplitude is greater than the 3^{rd} peak	V4, V6	0.011, 0.009
Power spectrum (harmonic) 1^{st} peak amplitude is too low	V2, V3	0.007, 0.012
Power spectrum (harmonic) 3^{rd} peak amplitude is too low	I, aVL	0.011, 0.015

CWT = continuous wave transform; ECG = electrocardiogram; LV = left ventricular; RV = right ventricular.

model incorporating ARIC, cNRI was 0.79 (95% CI: 0.23 to 1.17) and IDI 0.09 (95% CI: 0.012 to 0.24). For the model incorporating only ewECG features, cNRI was 0.94 (95% CI: 0.46 to 1.29) and IDI 0.11 (95% CI: 0.017 to 0.255).

The RF classifiers were inspected to reveal their node features. For the LVD predictive model, features were temporally associated with both depolarization and repolarization and included several features derived from the energy/power spectrum, that is, certain ratios of harmonics within the power spectrum throughout cardiac cycles (**Table 3**).

PREDICTION OF LOW GLS BY ewECG USING THE RF CLASSIFIER. When the features from the predictive model for LVD were used to train a predictive model for low GLS (≤16%), this was not able to identify any cases of low GLS (Supplemental Table 2). After repeating feature selection, 16 features were found to confer peak predictive power (Supplemental Table 3), and performance on the test dataset showed a sensitivity of 57% and a specificity of 90% (**Table 4**). However, the proportion with low GLS in the test dataset (6%) was significantly lower than in the training dataset (19%; p = 0.002), which is of significance in interpreting model performance. The CWT features selected for the low GLS model were predominantly power spectrum and repolarization features (ARIC HF risk score was not selected). Nine out of the 16 features were chosen for the LVD model.

PREDICTION OF DIASTOLIC ABNORMALITIES BY ewECG USING THE RF CLASSIFIER. Overall, 91 subjects (23%) exhibited diastolic abnormalities, with a significantly lower proportion in the test versus the training dataset (13% vs. 27%; p = 0.002) (**Table 2**). Again, features from the LVD predictive model trained for diastolic abnormalities were unable to discriminate (Supplemental Table 2). After repeated feature selection, a model with 14 features produced a sensitivity of 50% and a specificity of 90% (**Table 4**). Features selected for inclusion in the model included power spectrum and repolarization features as well as 1 depolarization-related feature that also occurred in the LVD model (ARIC HF risk score was not selected) (Supplemental Table 4). Ten out of the 14 features were also chosen for the LVD model.

IMPACT OF ewECG ON A SCREENING PROCESS FOR LVD. On the basis of greatest sensitivity for LVD identification, the RF ewECG model without the ARIC HF risk score was considered most optimal. Use of this ewECG-based model to select people for echocardiography could reduce the number of echocardiograms by 45% [(true negatives + false negatives)/total screened × 100]. However, 12% of cases of LVD would be missed. Alternatively, the RF model incorporating ARIC score and ewECG would result in a 47% reduction in echocardiograms but would miss 14% of cases of LVD. In comparison, using the best of the conventional methods, that is, ARIC HF risk score ≥2.6, echocardiograms would be reduced by

only 27% and 9% of LVD cases would be missed (Central Illustration).

DISCUSSION

The application of signal processing (CWT) to a conventionally acquired ECG signal provides a viable screening test that can be integrated with echocardiographic screening for LVD. In our cohort of people at risk for HF, our machine-learning RF algorithm provided 88% sensitivity and 70% specificity for detection of LVD, outperforming a clinical risk score, biomarkers, and an established automated ECG analysis algorithm. Furthermore, we highlighted that CWT features required for identification of LVD differed slightly according to LV abnormality, for example, reduced systolic function or diastolic abnormality. The best-performing model was for prediction of a composite measure of LVD. There was no significant difference between a machine-learning model that incorporated clinical information, namely, the ARIC HF risk score, compared with ewECG features alone. If implemented in combination with echocardiography as a screening test, ewECG could reduce echocardiograms by 45% in screening for LVD. This is important given the 82% prevalence of subclinical HF in the community in those >67 years of age (16). For the United States this would mean that approximately 40 million people would be eligible for screening and ewECG could reduce the number of echocardiograms by around 18 million (17). Even if such a screening program did not eventuate, ewECG may serve as a gatekeeper to the echocardiography laboratory in high-risk but asymptomatic individuals.

ECG AND MYOCARDIAL DYSFUNCTION. Structural and metabolic cardiac pathology manifesting as electrocardiographic abnormalities is well accepted. However, abnormalities may be too subtle for either a human reader or standard analytics to detect (9). Accordingly, a recent study used artificial intelligence (AI) (convolutional neural networks) applied to standard ECG digital data to predict LVD (defined as LV ejection fraction ≤35%) with a sensitivity of 89%, a specificity of 83%, and an AUC of 0.93 (18). Interestingly, those with a false positive result were 4 times as likely to develop LVD during 4-year follow-up, suggesting that the AI could recognize early abnormalities. Another approach extracted advanced ECG (A-ECG) parameters (3-dimensional ECG parameters, QRS/T-wave complexity parameters) including variability analysis (5-min high-fidelity ECG recording) as well as conventional ECG measures to devise a prediction score for myocardial disease (19). The authors demonstrated that a 5-parameter A-ECG score (derived with the use of a feature selection technique and logistic regression) had 83% sensitivity and 93% specificity for LVD (LV ejection fraction <50%) in a group of predominantly male subjects with either CAD or LVH. Interestingly, none of the features used in this score required an extended duration or high sample rate recording and as such could be attained from a conventionally recorded ECG. These works and ours demonstrate a growing body of evidence supporting the feasibility of electrocardiographic identification of LVD.

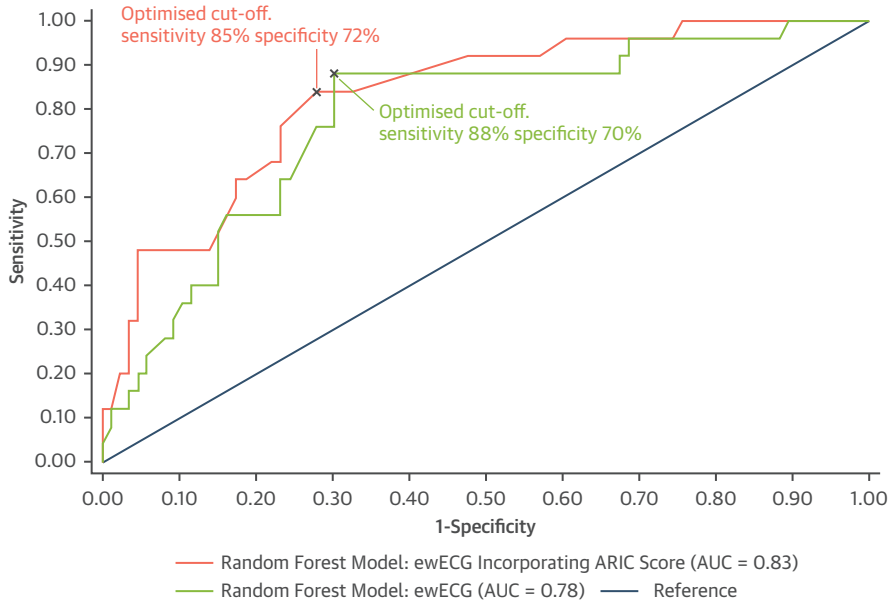
CWT PROCESSING AND CARDIAC DISEASE. Wavelet transforms have been applied to the ECG signal for measurement of intervals, noise reduction, and importantly identification of abnormalities (20). One of the first applications was identification of ventricular late potentials, microvoltage deflections after (and sometimes within) the QRS complex that are often obscured by noise (4,20). The detection of ventricular late potentials with the use of CWT improved prediction of post-infarction ventricular arrhythmias from 52% to 72% and from 64% to 76% for inferior and anterior infarctions, respectively, compared with standard signal filtering (21). Wavelet transforms have also revealed electric similarities

TABLE 4 Performance of Random Forest Classifier Models for Predicting LV Dysfunction, Low GLS Alone, and Diastolic Abnormalities

	Training (n = 287)	Test (n = 111)
Prediction target: LV dysfunction		
Model components: ewECG features + ARIC HF risk score		
Sensitivity, %	67	85
Specificity, %	68	72
AUC (95% CI)	0.71 (0.64-0.77)	0.83 (0.74-0.92)
F-score	0.68	0.60
Model components: ewECG features		
Sensitivity, %	66	88
Specificity, %	60	70
AUC (95% CI)	0.66 (0.59-0.72)	0.78 (0.67-0.88)
F-score	0.64	0.59
Prediction target: Low GLS		
Sensitivity, %	35	57
Specificity, %	95	90
AUC (95% CI)	0.67 (0.58-0.76)	0.65 (0.37-0.93)
F-score	0.45	0.36
Prediction target: Diastolic abnormalities		
Sensitivity, %	36	50
Specificity, %	94	90
AUC 95% (CI)	0.69 (0.62-0.76)	0.62 (0.42-0.82)
F-score	0.47	0.45

AUC = area under the receiver-operating characteristic curve; CI = confidence interval; other abbreviations as in Tables 1 and 2.

FIGURE 2 Performance of the Random Forest Classifier Model Utilizing ewECG Features for Prediction of Left Ventricular Dysfunction

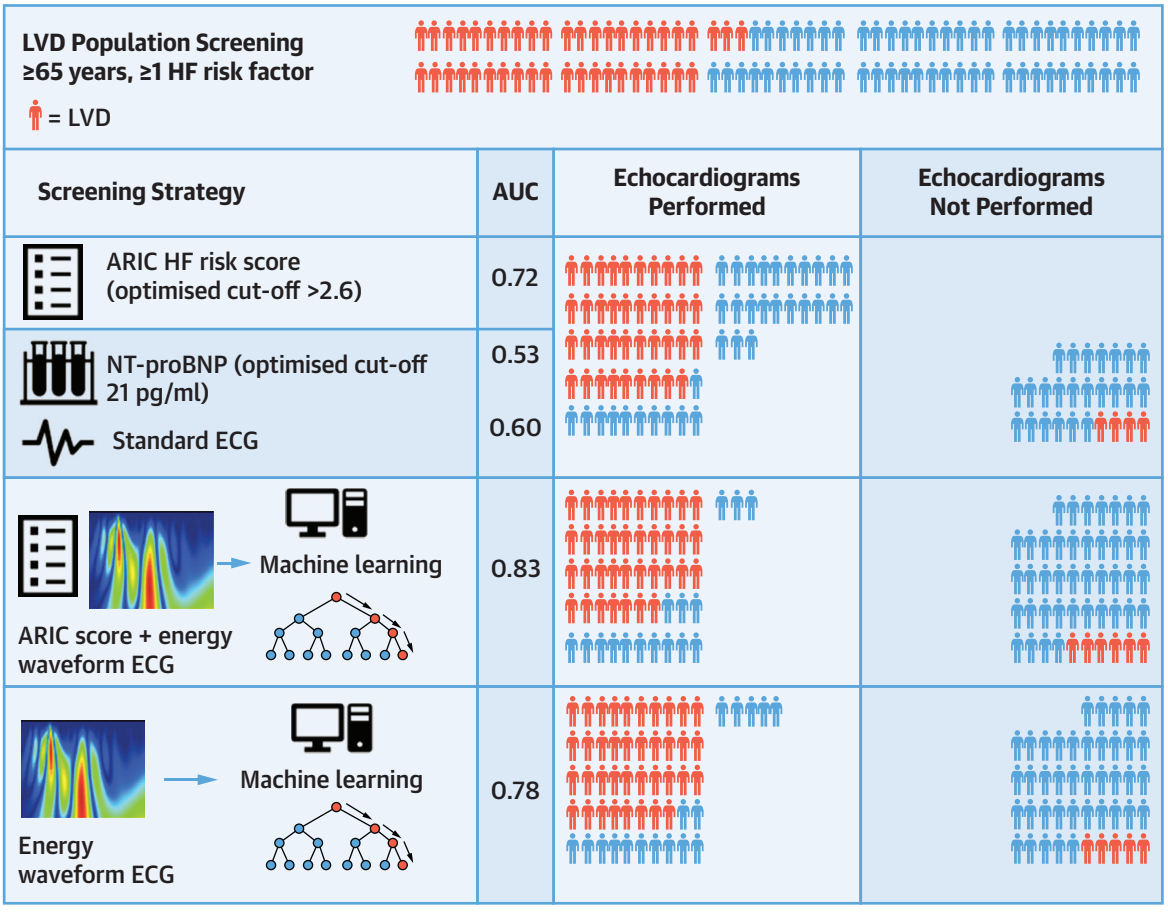


The area under the receiver operator characteristic curves (AUCs) were similar for ewECG features combined with the ARIC heart failure risk score (**(red)**; AUC: 0.83; sensitivity: 85%; specificity: 72%) and with ewECG alone (**(green)**; AUC: 0.78; sensitivity: 88%: specificity: 70%). ARIC = Atherosclerosis Risk in Communities; other abbreviations as in Figure 1.

(abnormal frequency content) within the QRS complex in congenital and acquired long QT syndrome, providing insight into shared electropathophysiology (22).

Although CWT-processed ECG has demonstrated value by revealing known or suspected electric abnormalities, there are limited data on a direct association between CWT-processed ECG features and cardiac function. Associations between standard ECG features and cardiac dysfunction has focused on long QT syndrome, which is associated with increased isovolumetric relaxation time, altered tissue Doppler velocity profiles, and mechanical dispersion (23,24). Furthermore, the interval from T-wave peak to T-wave end is increased in DD assessed according to mitral inflow and tissue Doppler velocities (7). At the molecular level, DD is partly related to low-amplitude calcium transients secondary to reduced calcium uptake into, and leakage from, the sarcoplasmic reticulum (25,26). Given that calcium is a key modulator of the action potential duration, disturbances in the electric signal on the surface ECG may be apparent in LVD. It follows that detailed decomposition of the ECG signal from a diseased myocardium may reveal characteristic abnormalities, and indeed, the machine-learning model using only ewECG features provided accurate detection of LVD. Recently, CWT-processed ECG (as used in our study) has shown 80% sensitivity and 84% specificity for abnormal relaxation, assessed by low e′ (AUC: 0.9) in a cohort of patients presenting with symptoms of CAD (5). As in our study, a machine-learning approach (RF classifier) was used but with far more features (n = 257) because of different methodologies. Furthermore, in a larger patient cohort with similar characteristics (e.g., suspected CAD or indication for LV function evaluation) (n = 1,202), a machine-learning algorithm was trained with ewECG features to quantitatively predict e′ (6). This algorithm was able to discriminate guideline-defined thresholds with an AUC of 0.84, and because the model generated a continuous output for e′, inaccuracies associated with age-based declines could be avoided. Although we chose cutoffs, there is no suggestion that GLS declines in normal aging (27), and our definition of diastolic dysfunction would be inclusive of signs of early disease before elevations in LA pressure. We also think that our study population is the most appropriate choice for testing and application of this technology, that is, where echocardiography may not be strictly indicated.

CENTRAL ILLUSTRATION **Comparison of Strategies for LVD Screening**



Potter, E.L. et al. J Am Coll Cardiol Img. 2021;■(■):■–■.

The use of an echo screening strategy in all people with risk factors would identify almost 50% of the population as having LVD. A combination of clinical scoring and NT-proBNP would reduce echocardiography by about one-third and miss few people with LVD. However, the use of ewECG would reduce the need for echocardiography and miss even fewer patients with LVD. ARIC = Atherosclerosis Risk in Communities; AUC = area under the receiver-operating characteristic curve; ECG = electrocardiography; ewECG = energy waveform ECG; HF = heart failure; LVD = left ventricular dysfunction; NT-proBNP = N-terminal pro-B-type natriuretic peptide.

The benefit of our machine-learning method, as opposed to an AI approach (e.g., neural networks), is the potential for interrogation of the model to provide mechanistic insight. We were interested to see whether systolic dysfunction was exclusively temporally associated with depolarization features, which it was not. This may not be surprising for 2 reasons: 1) the surface ECG is a simplification of electric activity spreading across the complex 3-dimensional structure of the heart and body; and 2) early LV systolic dysfunction and diastolic dysfunction often coexist (28,29). The predictive model for diastolic abnormalities included measures from depolarization as well as repolarization, and most of the features within the low GLS and diastolic models also appeared in the composite LVD model. Clearly, investigation concerning the association between LVD and specific CWT signatures is in its infancy and is likely to be facilitated by machine-learning algorithms.

SCREENING FOR LVD. The detection of subclinical LVD fulfills some but not all criteria for screening (30). On both the individual and the population health levels, HF is burdensome, and its natural history involves an early asymptomatic stage that is readily detected as abnormal GLS and DD, which carry

risk of symptomatic HF and mortality (31–34), analogous to standard markers of impaired LV function (34,35). In terms of treatment, guideline-advocated therapies (angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, and beta-blockers) significantly improve outcomes (36,37), not only in populations with ischemic cardiomyopathy with reduced ejection fraction, but also in subjects with reduced GLS and diastolic abnormalities with preserved ejection fraction, where intensification of cardioprotective therapies may reduce progression to symptomatic HF (38).

Although echocardiography is safe and accurate, cost and access may be problematic. In this setting, the high sensitivity of the ewECG cutoffs that we have developed minimizes the number of patients going to echocardiography, while at the same time minimizing the numbers of false negatives who do not proceed. The test is low risk and acceptable to patients. The sensitivity of ewECG in our study is superior to the fecal occult blood test for colorectal cancer screening, although the specificity is lower (39). However, the risks associated with further testing after a positive ewECG (i.e., echocardiography) are far lower than for colonoscopy, for example. Nonetheless, further work with ewECG will need to include integration of machine-learning algorithms into the device's software to enable immediate interpretation and guide decision making and integration of ewECG into clinical workflows.

The alternative is the use of natriuretic peptides (NPs) (e.g., BNP ≥50 pg/ml) to guide therapy. Intensification of renin-angiotensin-aldosterone system inhibition and beta-blockade in diabetics with NT-proBNP >125 pg/ml has been shown to reduce cardiovascular hospitalizations compared with usual care (relative risk: 0.52; 95% CI: 0.4 to 0.68) (40). However, although previous work has shown that NP-based therapy reduces asymptomatic LVD in individuals >40 years of age with cardiovascular risk factors (adjusted odds ratio: 0.6; 95% CI: 0.39 to 0.93), that study showed no significant difference in HF hospitalization over the 4.2-year mean follow-up (41). Indeed, we found that NT-proBNP had poor screening performance. An inherent problem of BNP in this setting is that levels are artefactually reduced in the setting of obesity. Thus, the role of NT-proBNP in a screening role in this population remains unclear.

STUDY LIMITATIONS. Machine-learning models are inherently limited by the amount of data available to train the algorithm. Continued acquisition of ewECG data will continue to improve our models. We demonstrated that models differ between targets; performance for one cardiac abnormality in one population should not be extrapolated to others. The poor performance of ewECG for diastolic and early systolic dysfunction abnormalities is likely due to the small number of abnormal studies available for the algorithm to train with, as well as the fact that the over-represented group (in this case, normal studies) is, by chance, more likely to be predicted. More developmental work is needed to apply ewECG in these settings. Our study is cross-sectional and therefore we do not know what proportion will go on to develop symptomatic HF or whether ewECG varies between those who do or do not progress. Furthermore, it is unknown whether ewECG can reveal abnormalities before the onset of early LVD or in the subset of patients who fail to exhibit resting echocardiographic abnormalities before manifesting HF.

The definition of diastolic dysfunction used in this paper is not conventional. We chose not to use the standard criteria because many subjects are identified as indeterminate. To create a definition suitable for screening, we used clear criteria of raised LA pressure (E/e′ with LAE), or if ambiguous (e.g., isolated LAE) partnered that with another diastolic dysfunction marker. Therefore there were 3 criteria: E/e′ >15, E/e′ >10 with LAE, and impaired relaxation and LAE.

CONCLUSION

Patients with subclinical LVD are at increased risk of HF, which may be prevented by initiation of cardioprotective therapy. However, there is currently no consensus as to whether (or how) subclinical LVD should be detected. Advanced analysis of a routinely acquired ECG using CWT signal processing and machine learning would be a suitably sensitive first step in a selective echocardiographic screening process for detection of LVD. Should such screening be adopted, it could reduce the number undergoing echocardiography by almost one-half.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

The work was partially supported by a partnership grant (1149692) from the National Health and Medical Research Council, Canberra, the Ian Potter Foundation, Melbourne, and the Baker Heart and Diabetes Institute, Melbourne. Dr. Potter is supported by a Monash University postgraduate scholarship. All authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Prof. Tom Marwick, Baker Heart and Diabetes Institute, 75 Commercial Road, Melbourne, Victoria 3004, Australia. E-mail: tom.marwick@baker.edu.au.

PERSPECTIVES

COMPETENCY IN SYSTEMS-BASED PRACTICE: The detection of LV dysfunction (LVD) by echocardiography in asymptomatic people with risk factors for heart failure (HF) identifies a group who are at increased risk of HF. However, echocardiographic screening of the population provides logistical and financial challenges. A selection process for echocardiography would make detection more feasible.

TRANSLATIONAL OUTLOOK: This study provides data to support the use of "energy waveform" electrocardiography (ewECG) to identify people at low risk. This could reduce the need for echocardiography by >50% while at the same time missing a minimal number of patients with LVD. Further evaluation of ewECG is warranted for selection of patients for screening for LVD.

REFERENCES

1. Hunt SA, Abraham WT, Chin MH, et al. 2009 focused update incorporated into the ACC/AHA 2005 guidelines for the diagnosis and management of heart failure in adults. Circulation 2009;119:e391–479.

2. Garbi M, Edvardsen T, Bax J, et al. EACVI appropriateness criteria for the use of cardiovascular imaging in heart failure derived from European National Imaging Societies voting. Eur Heart J Cardiovasc Img 2016;17:711–21.

3. Crowe J, Gibson N, Woolfson M, Somekh M. Wavelet transform as a potential tool for ECG analysis and compression. J Biomed Eng 1992;14:268–72.

4. Meste O, Rix H, Caminal P, Thakor N. Ventricular late potentials characterization in time-frequency domain by means of a wavelet transform. IEEE Trans Biomed Eng 1994;41:625–34.

5. Sengupta P, Kulkarni H, Narula J. Prediction of abnormal myocardial relaxation from signal processed surface ECG. J Am Coll Cardiol 2018;71:1650–60.

6. Kagiyama N, Piccirilli M, Yanamala N, et al. Machine learning assessment of left ventricular diastolic function based on electrocardiographic features. J Am Coll Cardiol 2020;76:930–41.

7. Sauer A, Wilcox J, Andrei A, Passman R, Goldberger J, Shah S. Diastolic electromechanical coupling. Circ Arrhythm Electrophysiol 2012;5:537–43.

8. Yang H, Wang Y, Nolan M, Negishi K, Okin P, Marwick T. Community screening for nonischemic cardiomyopathy in asymptomatic subjects ≥65 years with stage B heart failure. Am J Cardiol 2016;117:1959–65.

9. Willems J, Abreu-Lima C, Arnaud P, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. N Engl J Med 1991;325:1767–73.

10. Clark E, Sejersten M, Clemmensen P, Macfarlane P. Automated electrocardiogram interpretation programs versus cardiologists' triage decision making based on teletransmitted data in patients with suspected acute coronary syndrome. Am J Cardiol 2010;106:1696–702.

11. The MathWorks. Five easy steps to a continuous wavelet transform. Available at: http://matlab.izmiran.ru/help/toolbox/wavelet/ch01_i15.html. Accessed May 10, 2021.

12. Potter E, Marwick T. Detection of stage B heart failure in the community using energy waveform ECG. J Am Coll Cardiol 2019;73 11 Suppl 1:1894.

13. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. J Machine Learn Res 2011;12:2825–30.

14. Pencina MJ, d'Agostino RB Sr., Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. Stat Med 2012;31:101–13.

15. Pencina MJ, d'Agostino RB Sr., d'Agostino RB Jr., Vasan RS. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. Stat Med 2008;27:157–72.

16. Shah A, Claggett B, Loehr L, Chang P, Matsushita K, Kitzman D. Heart failure stages among older adults in the community. Circulation 2017;135:224–40.

17. Administration for Community Living, US Department of Health and Human Services. Profile of older americans. Available at: https://acl.gov/aging-and-disability-in-america/data-and-research/profile-older-americans. Accessed May 10, 2021.

18. Attia Z, Kapa S, Lopez-Jimenez F, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nature Med 2019;25:70–4.

19. Schlegel T, Kulecz W, Feiveson A, Greco E, DePalma J, Starc V. Accuracy of advanced versus strictly conventional 12-lead ECG for detection and screening of coronary artery disease, left ventricular hypertrophy and left ventricular systolic dysfunction. BMC Cardiovasc Disord 2010;10:28.

20. Addison P. Wavelet transforms and the ECG: a review. Physiol Meas 2005;26:R155.

21. Reinhardt L, Mäkijärvi M, Fetsch T, Montonen J, Sierra G. Predictive value of wavelet correlation functions of signal-averaged electrocardiogram in patients after anterior versus inferior myocardial infarction. J Am Coll Cardiol 1996;27:53–9.

22. Chevalier P, Rodriguez C, Bontemps L, et al. Noninvasive testing of acquired long QT syndrome: evidence for multiple arrhythmogenic substrates. Cardiovasc Res 2001;50:386–98.

23. Savoye C, Klug D, Denjoy I, et al. Tissue Doppler echocardiography in patients with long QT syndrome. Eur J Echocardiogr 2003;4:209–13.

24. Haugaa K, Edvardsen T, Leren T, et al. Left ventricular mechanical dispersion by tissue Doppler imaging: a novel approach for identifying high-risk individuals with long QT syndrome. Eur Heart J 2009;30:330–7.

25. Piacentino V, Weber C, Chen X, et al. Cellular basis of abnormal calcium transients of failing human ventricular myocytes. Circ Res 2003;92:651–8.

26. Hobai I, O'Rourke B. Decreased sarcoplasmic reticulum calcium content is responsible for defective excitation-contraction coupling in canine heart failure. Circulation 2001;103:1577–84.

27. Potter E, Wright L, Yang H, Marwick T. Normal range of global longitudinal strain in the elderly: the impact of subclinical disease. J Am Coll Cardiol Img 2021;14:300–2.

28. Yu C-M, Lin H, Yang H, Kong S-L, Zhang Q, Lee S-L. Progression of systolic abnormalities in patients with "isolated" diastolic heart failure and diastolic dysfunction. Circulation 2002;105:1195–201.

29. Kraigher-Krainer E, Shah A, Gupta D, Santos A, Claggett B, Pieske B. Impaired systolic function by strain imaging in heart failure with preserved ejection fraction. J Am Coll Cardiol 2014;63:447–56.

30. Wilson J, Junger G. Principles and practice of screening for disease. Geneva: World Health Organisation; 1968. Available at: https://apps.who.int/iris/handle/10665/37650. Accessed May 10, 2021.

31. Biering-Sorensen T, Biering-Sorensen S, Olsen F, Sengelov M, Jorgensen P, Mogelvang R. Global longitudinal strain by echocardiography predicts long-term risk of cardiovascular morbidity and mortality in a low-risk general population: the

Copenhagen City Heart Study. Circ Cardiovasc Imaging 2017;10:e005521.

32. Stanton T, Leano R, Marwick T. Prediction of all-cause mortality from global longitudinal speckle strain: comparison with ejection fraction and wall motion scoring. Circ Cardiovasc Imaging 2009;2:356–64.

33. From A, Scott C, Chen H. The development of heart failure in patients with diabetes mellitus and pre-clinical diastolic dysfunction. J Am Coll Cardiol 2010;55:300–5.

34. Wang Y, Yang H, Huynh Q, Nolan M, Negishi K, Marwick TH. Diagnosis of nonischemic stage B heart failure in type 2 diabetes mellitus: optimal parameters for prediction of heart failure. J Am Coll Cardiol Cardiovasc Img 2018;11:1390–400.

35. Ammar KA, Jacobsen SJ, Mahoney DW, et al. Prevalence and prognostic significance of heart failure stages. Circulation 2007;115:1563–70.

36. SOLVD Investigators. Effect of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left ventricular ejection fractions. N Engl J Med 1992;327:685–91.

37. CAPRICORN Investigators. Effect of carvedilol on outcome after myocardial infarction in patients with left-ventricular dysfunction: the CAPRICORN randomised trial. Lancet 2001;357:1385–90.

38. Yang H, Negishi K, Wang Y, Nolan M, Marwick T. Imaging-guided cardioprotective treatment in a community elderly population of stage B heart failure. J Am Coll Cardiol Imag 2017;10:217–26.

39. Park D, Ryu S, Kim Y. Comparison of guaiac-based and quantitative immunochemical fecal occult blood testing in a population at average risk undergoing colorectal cancer screening. Am J Gastroenterol 2010;105:2017–25.

40. Sweeney C, Ryan F, Ledwidge M, et al. Natriuretic peptide-guided treatment for the prevention of cardiovascular events in patients without heart failure. Cochrane Database Syst Rev 2019;10:CD013015.

41. Ledwidge M, Gallagher J, Conlon C, et al. Natriuretic peptide-based screening and collaborative care for heart failure: the STOP-HF randomized trial. JAMA 2013;310:66–74.

KEY WORDS electrocardiography, left ventricular dysfunction, machine-learning, screening

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