ARTICLE OPEN

Check for updates

Detection of left ventricular systolic dysfunction from single-lead electrocardiography adapted for portable and wearable devices

Akshay Khunte [1], Veer Sangha[1], Evangelos K. Oikonomou[2], Lovedeep S. Dhingra [2], Arya Aminorroaya[2], Bobak J. Mortazavi [3,4], Andreas Coppi [4,5], Cynthia A. Brandt[6,7], Harlan M. Krumholz [2,4,8] and Rohan Khera [2,4,6,9 ✉]

Artificial intelligence (AI) can detect left ventricular systolic dysfunction (LVSD) from electrocardiograms (ECGs). Wearable devices could allow for broad AI-based screening but frequently obtain noisy ECGs. We report a novel strategy that automates the detection of hidden cardiovascular diseases, such as LVSD, adapted for noisy single-lead ECGs obtained on wearable and portable devices. We use 385,601 ECGs for development of a standard and noise-adapted model. For the noise-adapted model, ECGs are augmented during training with random gaussian noise within four distinct frequency ranges, each emulating real-world noise sources. Both models perform comparably on standard ECGs with an AUROC of 0.90. The noise-adapted model performs significantly better on the same test set augmented with four distinct real-world noise recordings at multiple signal-to-noise ratios (SNRs), including noise isolated from a portable device ECG. The standard and noise-adapted models have an AUROC of 0.72 and 0.87, respectively, when evaluated on ECGs augmented with portable ECG device noise at an SNR of 0.5. This approach represents a novel strategy for the development of wearable-adapted tools from clinical ECG repositories.

INTRODUCTION

Left ventricular systolic dysfunction (LVSD) is associated with more than 8-fold increase in heart failure risk and to a nearly 2-fold increase in the risk of dying prematurely[1]. Early diagnosis can effectively mitigate this risk[2–4], but LVSD is frequently diagnosed only after patients develop symptomatic disease due to the lack of effective screening strategies[5–7]. Artificial intelligence (AI) can detect left ventricular systolic dysfunction (LVSD) from electrocardiograms (ECGs), a diagnosis that has traditionally relied on comprehensive echocardiography or other cardiac imaging, which is resource-intensive and challenging to use for generalized screening strategies[8,9]. Even though AI-ECG is a promising screening tool for detecting LVSD, the algorithms have been designed in clinically obtained 12-lead ECGs. Advances in wearable and handheld technologies now enable the point-of-care acquisition of single-lead ECG signals, paving the path for more efficient and scalable screening tools with these AI-ECG technologies[10,11]. This improved accessibility could enable broader AI-based screening for LVSD, but the reliability of such tools is limited by the presence of noise in data collected from wearable and handheld devices[12,13]. Consequently, the performance of models for detecting LVSD from portable device ECGs may degrade in the real-world setting, with lower performance than observed in the original single-lead derivatives of the clinical development studies[14,15].

In the absence of large, labelled datasets of wearable ECGs, the development of algorithms that can detect underlying structural heart disease on wearable devices relies on single-lead information specifically adapted from 12-lead ECGs extracted from clinical ECG libraries. However, this process does not specifically account for the unique data acquisition challenges encountered with wearable ECG, possibly contributing to their inconsistent diagnostic performance. Indeed, several sources of noise exist in wearable data, arising from factors such as poor electrode contact with the skin, movement and muscle contraction during the ECG, and external electrical interference[16–19]. This noise has practical implications, as models demonstrate poorer performance when tested on all available wearable ECG data as opposed to selected high-quality subsets[15]. This marked difference in performance based on noise has limited wearable device-based screening programs, with a wearable device-based atrial fibrillation screening study disqualifying 22% of patients due to insufficient signal quality[12]. Accounting for this noise is a prerequisite to develop broadly accessible models that will form the basis of effective screening programs for LVSD in the community.

In the present study, we hypothesize that a novel, noise-enhanced training approach can boost the performance of wearable-adapted, single-lead ECG models for accurate and noise-agnostic identification of LVSD. Our method, which relies on training on single-lead ECG data derived from clinical ECGs and augmented with custom noise patterns developed in key frequency ranges relevant for specific ECG noise signatures, explicitly accounts for, and generalizes to multiple real-world noise patterns, including noise isolated from a portable ECG device.

[1]Department of Computer Science, Yale University, New Haven, CT, USA. [2]Section of Cardiovascular Medicine, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA. [3]Department of Computer Science & Engineering, Texas A&M University, College Station, TX, USA. [4]Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, New Haven, CT, USA. [5]Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA. [6]Section of Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT, USA. [7]VA Connecticut Healthcare System, West Haven, CT, USA. [8]Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA. [9]Section of Health Informatics, Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA. ✉email: rohan.khera@yale.edu

RESULTS

Study population

There were 2,135,846 consecutive 12-lead ECGs performed at the Yale-New Haven Hospital (YNHH) between 2015 and 2021, 440,072 of which had accompanying transthoracic echocardiograms (TTEs) acquired within 15 days of the ECG. We developed the models on 385,601 of the ECG-TTE pairs, representing 116,210 unique patients, who had a complete 12-lead ECG recording (Fig. S1). The signal from Lead I, the standard lead obtained from wearable devices[11], was then isolated from each 12-lead ECG. All selected single-lead ECG recordings contained 10 s of Lead I signal at 500 Hz. The single-lead ECGs were then split at the patient level into training, validation, and test datasets (85%-5%-10%).

Of these ECGs, 56,894 (14.8%) were from patients with LVSD, defined as having a paired TTE recording of LV ejection fraction (LVEF) below 40%. Additionally, 40,240 (10.4%) had an LVEF between 40% and 50%, and the remaining 288,467 (74.8%) had an LVEF of 50% or higher. Patients had a median age of 68 years (IQR 56, 78) at the time of ECG recording and 50,776 (43.7%) of the patients were women. A total of 75,928 (65.3%) patients were non-Hispanic white individuals, 14,000 (12.0%) were non-Hispanic Black, 9,349 (8.0%) were Hispanic, and 16,843 (14.5%) were from other racial backgrounds (Table S1).

Detection of LV systolic dysfunction

The noise-adapted model was trained on a noise-augmented development set. High and low pass filtering was used to isolate five-minute samples of random gaussian noise within four different frequency ranges encompassing real-world ECG noises, including 3–12 Hz, 12–50 Hz, 50–100 Hz, and 100–150 Hz. The first of these four ranges, 3–12 Hz, was selected to emulate motion artifact noises due to tremors[20,21], while the 12–50 Hz and 100–150 Hz frequency ranges encompass more frequently occurring lower- and higher-frequency muscle activation artifacts, respectively[17,21]. The 50–100 Hz domain was selected to represent electrode motion noises[21]. Both this domain and the 100–150 Hz frequency range, which contain multiples of 50 and 60 Hz, the two mains frequencies used in ECG acquisition[17], also serve to emulate powerline interference noise[17,21]. These noise samples were then used to generate the noise-augmented development set, in which ECGs were selectively noised with random, 10-second sequences of one of the four frequency-banded gaussian noises at one of four signal-to-noise ratios (SNRs) each time an ECG was loaded. The standard model was trained on the original training set (described in Methods, Isolation of Frequency-Banded Gaussian Noise and Methods, Noise Augmentation).

Both models were trained to detect LVEF below 40%, a threshold present in most heart failure diagnosis guidelines[4], and consistent with prior work[8,22]. Using DeLong's test to compare area under the receiver operating characteristic curves (AUROCs) of the noise-adapted and standard models for detection of LVEF < 40%[23,24], the noise-adapted and standard models, with an AUROC of 0.90 (95% CI 0.89–0.91) and 0.90 (95% CI 0.88–0.91), respectively, performed similarly on a held-out test set without added noise (p-value = 0.60). Area under the precision-recall curve (AUPRC) on this held-out test set was 0.46 and 0.48, respectively. With separate threshold values selected that achieved sensitivity of 0.90 in the validation set without added noise for each model, the noise-adapted model had specificity and sensitivity of 0.68 and 0.92, respectively, and a PPV and NPV of 0.20 and 0.99, respectively, in the held-out test set. The standard model had sensitivity and specificity of 0.69 and 0.91, respectively, and a PPV and NPV of 0.21 and 0.99, respectively. The noise-adapted model's performance was comparable to the standard model across subgroups of age, sex, and race (Table 1).

Standard and noise-adapted model performance on noised ECGs

The performance of each model was also tested on four distinct real-world noise recordings, including three half-hour recordings containing electrode motion, muscle artifact, and baseline wander noise sourced from the publicly available MIT-BIH noise stress test database[17]. Both models were tested on separate versions of the held-out test set augmented with each of these 3 real-world noises at seven different signal-to-noise ratios (described in Methods, Acquisition of Real-World, Public ECG Noise Recordings, and Methods, Noise Augmentation).

For the noise-adapted model, model performance was comparable across all SNRs for each MIT-BIH noise with AUROC between 0.86–0.89, 0.87–0.89, and 0.88–0.89 for electrode motion, muscle artifact, and baseline wander noise, respectively, for SNRs from 0.5 to 2.0. The standard model had lower performance across all SNRs for every noise, with AUROC between 0.79–0.86, 0.81–0.86, and 0.80–0.86 for electrode motion, muscle artifact, and baseline wander noise, respectively (Table 2 and Fig. 1).

Standard and noise-adapted performance with portable ECG device noise

Real-world portable ECG device noise, isolated using a Fast Fourier Transform-based technique on a 30-second, single-lead ECG recording from a KardiaMobile 6 L portable ECG device, was also used to evaluate both models (Methods, Noise Extraction from a Portable Device ECG). Each model was used to generate predictions for noise-augmented ECGs across all SNRs, with the noise-adapted model's AUROC ranging from 0.87 to 0.89. The standard model's performance was significantly lower at each

Table 1. Performance of noise-adapted model on test ECGs without noise across demographic subgroups.

Labels	PPV	NPV	Specificity	Sensitivity	AUROC (95% CI)	AUPRC	OR
All	0.203	0.990	0.676	0.924	0.896 (0.886–0.905)	0.455	25.195
Male	0.231	0.986	0.653	0.916	0.881 (0.867–0.894)	0.478	20.508
Female	0.163	0.994	0.716	0.923	0.913 (0.898–0.927)	0.404	30.318
White	0.204	0.990	0.675	0.922	0.894 (0.883–0.906)	0.445	24.544
Black	0.211	0.990	0.596	0.945	0.880 (0.854–0.906)	0.461	25.407
Hispanic	0.208	0.992	0.740	0.923	0.917 (0.881–0.953)	0.535	34.158
Other	0.218	0.995	0.772	0.941	0.932 (0.900–0.964)	0.545	54.122
65 or older	0.198	0.989	0.606	0.936	0.879 (0.865–0.893)	0.432	22.611
Under 65	0.209	0.990	0.731	0.910	0.908 (0.895–0.921)	0.481	27.543

PPV positive predictive value, *NPV* negative predictive value, *AUROC* area under receiver operating characteristic curve, *CI* confidence interval, *AUPRC* area under precision-recall curve, *OR* odds ratio.

Table 2. Performance of noise-adapted and standard model on noise-augmented test set ECGs across different types of noise.

Model	Noise	SNR	PPV	NPV	Specificity	Sensitivity	AUROC (95% CI)	AUPRC
Noise-adapted	None	N/A	0.203	0.990	0.676	0.924	0.896 (0.886–0.905)	0.455
	Portable ECG	0.5	0.152	0.993	0.523	0.956	0.871 (0.861–0.882)	0.392
	Portable ECG	2	0.188	0.992	0.636	0.940	0.889 (0.880–0.899)	0.439
	Electrode motion	0.5	0.126	0.993	0.395	0.971	0.858 (0.846–0.869)	0.361
	Electrode motion	2	0.167	0.991	0.579	0.945	0.885 (0.875–0.895)	0.426
	Muscle artifact	0.5	0.175	0.989	0.608	0.927	0.871 (0.861–0.882)	0.389
	Muscle artifact	2	0.194	0.990	0.654	0.930	0.891 (0.881–0.900)	0.449
	Baseline wander	0.5	0.186	0.990	0.634	0.932	0.883 (0.873–0.893)	0.427
	Baseline wander	2	0.201	0.991	0.669	0.929	0.892 (0.883–0.902)	0.457
Standard	None	N/A	0.207	0.988	0.688	0.910	0.895 (0.884–0.905)	0.475
	Portable ECG	0.5	0.091	0.981	0.132	0.972	0.723 (0.706–0.739)	0.200
	Portable ECG	2	0.125	0.991	0.398	0.958	0.834 (0.822–0.847)	0.329
	Electrode motion	0.5	0.088	0.988	0.079	0.990	0.792 (0.779–0.806)	0.239
	Electrode motion	2	0.152	0.988	0.537	0.925	0.855 (0.843–0.866)	0.333
	Muscle artifact	0.5	0.100	0.991	0.210	0.979	0.807 (0.795–0.820)	0.245
	Muscle artifact	2	0.181	0.988	0.631	0.912	0.864 (0.853–0.875)	0.344
	Baseline wander	0.5	0.095	0.993	0.158	0.987	0.802 (0.789–0.816)	0.236
	Baseline wander	2	0.169	0.986	0.599	0.908	0.855 (0.843–0.866)	0.322

SNR signal-to-noise ratio, *PPV* positive predictive value, *NPV* negative predictive value, *AUROC* area under receiver operating characteristic curve, *CI* confidence interval, *AUPRC* area under precision-recall curve.

SNR, ranging from 0.72 to 0.83. This difference was most pronounced at an SNR of 0.5, at which the noise-adapted model retained an AUROC of 0.87 (95% CI 0.86–0.88) and the standard model had an AUROC of 0.72 (95% CI 0.71–0.74, p-value < 0.001) (Table 2 and Fig. 1).

Standard and noise-adapted performance with multiple simultaneous noise types

Three unique, multiple-noise recordings were generated by adding the MIT-BIH electrode motion, baseline wander, and muscle artifact noise recordings to the portable ECG device noise at an SNR of 1.0. These generated noise recordings were then used to evaluate each model's performance on ECGs augmented with multiple-noise signatures simultaneously. For the noise-adapted model, model performance was comparable across all SNRs for the combinations of portable ECG device noise and electrode motion, muscle artifact, and baseline wander noise with AUROC between 0.86–0.89, 0.87–0.89, and 0.88–0.89, respectively, for SNRs from 0.5 to 2.0. The standard model had lower performance across all SNRs for every noise combination, with AUROC between 0.77–0.85, 0.77–0.85, and 0.78–0.86 for portable ECG device noise combined with electrode motion, muscle artifact, and baseline wander noise, respectively (Table S3).

Explaining performance differences in standard and noise-adapted models

To gain mechanistic insights into the variation in performance for standard and noise-augmented data by different models, we visually and quantitatively assessed differences in the embedding outputs of both noised and standard ECGs for both standard and noise-adapted models. For this, we focused on the output of the 320-dimensional last fully connected layer of each model before generating final predictions. These predictions were collected for five different versions of the same 1000 ECGs—once with the original ECG, and once for each of the four real-world noises. The variation in these predictions due to the addition of noise was visualized by using uniform manifold approximation and

projection (UMAP)[25], which constructs a two-dimensional representation of the 320-dimension prediction vectors for each noise and model combination. The magnitude of the shift in the UMAP projection due to noise augmentation, which approximates the shift in the underlying prediction vectors, corresponds to each models' resilience to a shift in predicted probabilities due to noise and artifacts in the signal. This shift was also quantitatively assessed using scaled Euclidean distances between prediction vectors for the same ECG with and without each type of noise for both models.

For the standard model, the predictions for each of the noised ECGs were visually distinct from those of the standard ECGs, despite being for the same set of 1000 ECGs, and differing only on the added noise. However, for the noise-adapted model, there was no visual separation in the model's predictions between standard and noised ECGs, indicating that the predictions of the noise-adapted model are more resilient than those of the standard model (Fig. 2).

Quantitatively, the scaled Euclidean distances between predictions for noised and standard versions of the ECGs were lower for the noise-adapted model across all four noises. In the ECG with noise derived from portable ECGs, for instance, the average scaled Euclidean distance for the standard model was 0.50 (95% CI 0.49–0.51) and for the noise-adapted model was 0.41 (95% CI 0.40–0.42). Similarly for the baseline wander it was 0.52 (0.51–0.53) and 0.36 (0.36–0.37), respectively (Table 3).

DISCUSSION

We developed a novel strategy that automates the detection of hidden signatures of cardiovascular diseases, such as LVSD, adapted for noisy single-lead ECGs obtained on wearable and portable devices. Using this approach, we developed a noise-adapted deep-learning algorithm that accurately identifies LV systolic dysfunction from single-lead ECG data and was resilient to significant noisy artifacts, despite not encountering the specific noises in the model development process. Specifically, the algorithm demonstrates excellent discriminatory performance
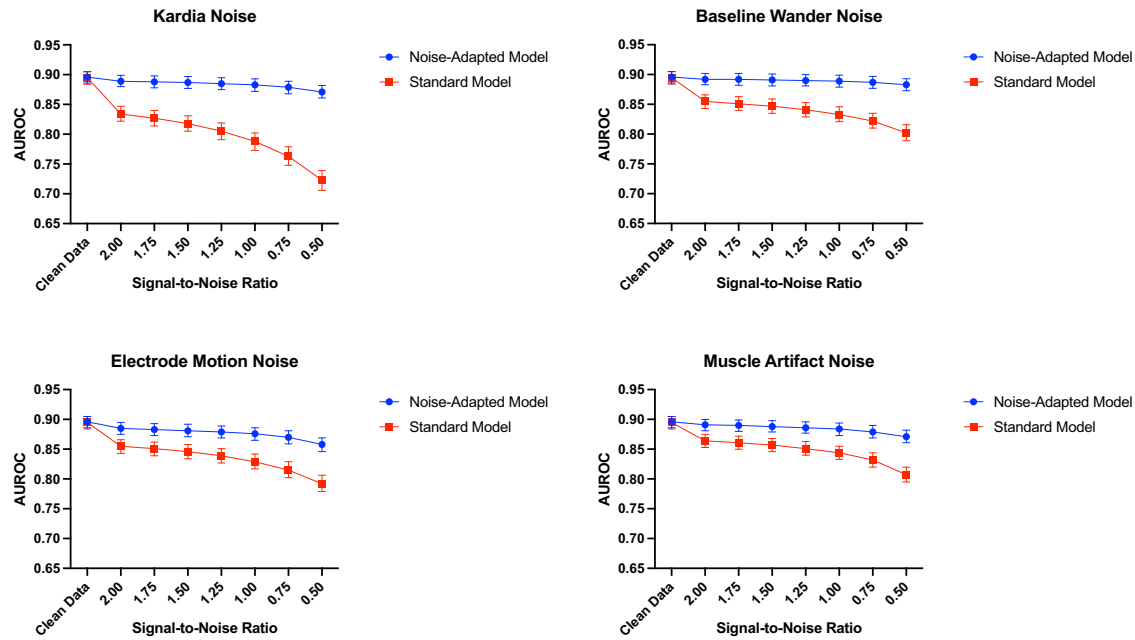
Fig. 1 Standard and noise-adapted model performance at increasing levels of noise. AUROC area under receiver operating characteristic curve, ECG electrocardiogram.

even on ECGs containing twice as much noise as signal, features that make it ideal for wearable device-based screening strategies. Notably, the algorithm was developed and validated in a diverse population and demonstrates consistent performance across age, sex, and race subgroups. The noise-adapted approach defines a novel paradigm on how to build robust, wearable-ready, single-lead ECG cardiovascular screening models from clinical ECG repositories, with significant potential to expand the screening of LV structural cardiac disorders to low-resource settings with limited access to hospital-grade equipment.

Noise-adapted training of deep-learning algorithms represents a relatively novel field of AI research, focused on expanding the use of AI tools to everyday life by accounting for noise and artifacts that may preclude their reliable use in this setting. Models trained on clinical ECGs have been applied to wearable device ECGs, but have traditionally shown significantly lower performance on wearable device ECGs than on held-out clinical ECG test sets[14,15]. Additionally, models which have successfully generalized to wearable device ECGs have relied on automated algorithms to select high-quality subsets of available ECG data, with reduced performance when evaluated on all collected ECGs[15,26]. While such algorithms can improve performance, this approach has limitations. Studies have shown high rates of poor signal quality among data recorded by wearable devices, and significant variation in quality across different manufacturers[13]. Excluding large portions of collected data may limit the scope of community-wide screening programs, which may use multiple different devices to collect ECGs and may not be able to collect multiple ECGs per person until a certain quality threshold has been met. A model which generalizes to ECGs with varying types and levels of noise may be necessary for implementation in such settings.

Due to the lack of publicly available wearable device ECG datasets, training models using wearable ECG data directly is challenging. The current 12-lead ECG-based models are limited to investments by health systems to incorporate tools into digital ECG repositories, and thereby limited to individuals who already seek care in those systems. In addition to the clinically indicated ECGs limiting the scope of screening, even this technology may not be available or cost-effective for smaller hospitals and clinics with limited access to digital ECGs.

Wearable devices allow obtaining ECGs that are more accessible and allow for community-wide screening, an important next step in the early detection of common and rare cardiomyopathies. On this note, our approach represents a major advancement from a methodological and clinical standpoint. First, it augments clinical ECG datasets in such way that it enables reliable modelling of noisy, wearable-derived, single-lead ECG signals. Second, it demonstrates that through noise-augmentation, single-lead ECG models can retain the prognostic performance of 12-lead ECG models, as shown here for the task of predicting LV systolic dysfunction. Additionally, this approach avoids the unnecessary exclusion of collected data, which increases its generalizability across different device platforms and is particularly important for community-wide screening programs, which may not be able to collect multiple ECGs per person and consistently meet high signal quality thresholds for each participant.

Our approach also offers a strategy to examine the process by which the models achieve better performance. Compared with the current approach of developing models, our noise-adapted approach resulted in selective removal of noise from signal, even for noises the model had not encountered before, while preserving the model's robustness in discerning complex hidden labels. This strategy is particularly important for a model intended for use on wearable devices, which capture ECGs in unpredictable settings with varying types and magnitudes of noise.

This study has several limitations. First, this model was developed using ECGs from patients who had both an ECG and a clinically indicated echocardiogram. Though this population differs from the intended broader real-world use of this algorithm as a screening method for LV systolic dysfunction among individuals with no clinical disease, the consistent performance across demographic subgroups suggests robustness and generalizability of the model's performance. Nevertheless, prospective assessments in the intended screening setting are warranted. Second, the model performance may vary by the severity of LV systolic dysfunction. Though the LVEF threshold of 40% was selected due to its therapeutic implications[4], it is possible that the model performance among patients with an LVEF near to this cutoff differed compared with those individuals with LVEF significantly higher or lower than 40%. This might also be
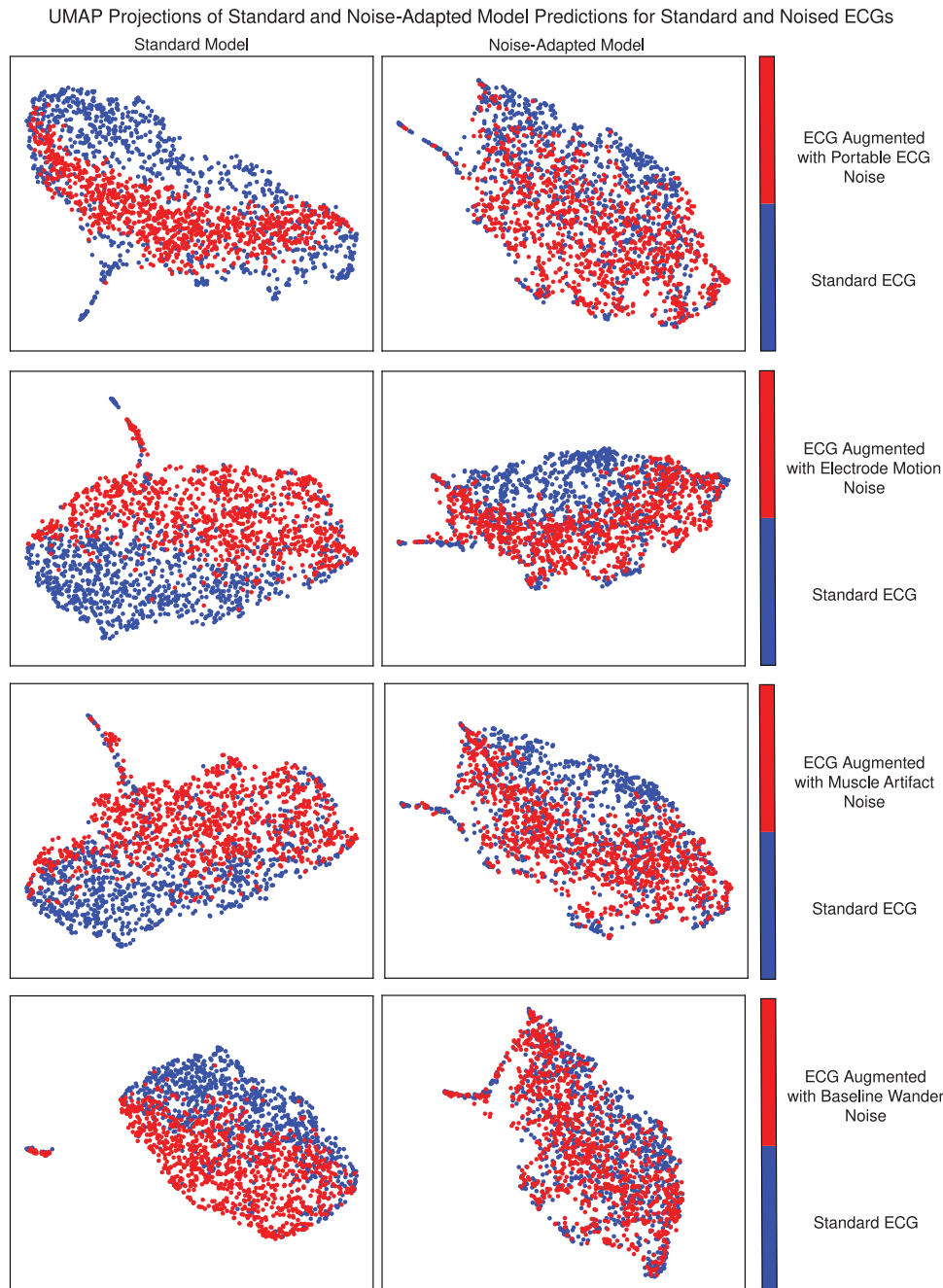
UMAP Projections of Standard and Noise-Adapted Model Predictions for Standard and Noised ECGs

Fig. 2 UMAP projections of last-layer predictions of standard and noise-adapted model for standard and noise-augmented ECGs. UMAP uniform manifold approximation and projection, ECG electrocardiogram.

attributable to a lack of precision in LVEF measurement by echocardiography, which has shown to be less precise relative to other approaches, such as magnetic resonance imaging[27,28]. Finally, four distinct types of randomly generated noise were used during training and randomly selected sequences of four real-world noises were used at multiple signal-to-noise ratios in the evaluation of performance on the test set. Though this suggests that the model performance generalizes well to unseen noise, we cannot ascertain whether it maintains performance on every type and magnitude of noise possible on wearable devices, including any device-specific noise signatures.

We developed a novel strategy that automates the detection of hidden signatures of cardiovascular diseases, such as LVSD, adapted for noisy single-lead ECGs obtained on wearable and portable devices. Using this approach, we developed a single-lead ECG algorithm that accurately identifies LV systolic dysfunction despite significant noisy artifacts, suggesting a novel approach for the development of wearable-adapted tools from clinical ECG repositories.

METHODS
Study design
The study was reviewed by the Yale Institutional Review Board, which approved the study protocol and waived the need for informed consent as the study represents secondary analysis of existing data. The data cannot be shared publicly.

Table 3. Scaled Euclidean distance between last-layer predictions of noise-adapted and standard models for noise-augmented and standard ECGs across different noise types at an SNR of 0.5.

Noise type	Standard model scaled euclidean distance (95% CI)	Noise-adapted model scaled euclidean distance (95% CI)	P-value
Portable ECG	0.50 (0.49–0.51)	0.41 (0.40–0.42)	<0.001
Electrode motion	0.56 (0.55–0.58)	0.49 (0.48–0.50)	<0.001
Muscle artifact	0.53 (0.52–0.54)	0.40 (0.39–0.41)	<0.001
Baseline wander	0.52 (0.51–0.53)	0.36 (0.36–0.37)	<0.001

SNR signal-to-noise ratio, ECG electrocardiogram, CI confidence interval, P-value probability value calculated using the paired t-test.

The study was designed as a retrospective analysis of a cohort of 116,210 patients, 18 years of age or older, who underwent clinically indicated ECG with paired echocardiograms within 15 days of the index ECG at the Yale-New Haven Health hospital. To ensure the generalizability of our models, we applied no exclusion criteria, including patients of all sexes, races, and ethnicities (Table S1).

Data source and population

Raw voltage data for lead I was isolated from 12-lead ECGs collected at the Yale-New Haven Hospital (YNHH) between 2015 and 2021. Lead I was chosen as it represents the standard lead obtained from wearable devices[11]. Each clinical ECG was recorded as a standard 10-second, 12-lead recording with a sampling frequency of 500 Hz. These ECGs were predominantly recorded using Philips PageWriter and GE MAC machines. Patient identifiers were used to link ECGs with an accompanying transthoracic echocardiogram within 15 days of the ECG. These echocardiograms had been evaluated by expert cardiologists, and the LVEF defined in their interpretation was identified. If multiple echocardiograms were performed within the 15-day window, the one nearest to each ECG was used to define the LVEF for the model development and evaluation.

Data preprocessing

A standard preprocessing strategy was used to isolate signal from lead I of 12-lead ECGs, that included median pass filtering and scaling to millivolts. The Lead I signal was then isolated from each ECG, and a one second median filter was calculated for and subtracted from each single-lead ECG to remove baseline drift. The amplitudes of each sample in each recording were then divided by a factor of 1000 to scale the voltage recordings to millivolts.

Isolation of frequency-banded Gaussian noise

Random gaussian noise within four distinct frequency ranges was isolated to train the noise-adapted model. High pass and low pass filters were applied to five minutes of random gaussian noise to isolate noise within each of the frequency ranges, which included 3–12 Hz, 12–50 Hz, 50–100 Hz, and 100–150 Hz. Each frequency range was specifically selected to model an element of real-world ECG noises. 3–12 Hz models the motion artifact noises attributable to tremors, which occur within this frequency range[20,21]. The 50–100 Hz domain reflects consistent electrode contact noise[21], while the 12–50 Hz and 100–150 Hz ranges contain the lower and higher-frequency muscle noises, respectively[17,21]. Additionally, the 50–100 and 100–150 Hz ranges each contain multiples of 50 and 60 Hz, the two mains frequencies used in ECG acquisition[17], which make up powerline interference noise[17,21] (Fig. 3).

Acquisition of real-world, public ECG noise recordings

Four real-world noise records, not included during training of either model, were used to test the models. These included three half-hour noise recordings from the MIT-BIH Noise Stress Test Database. The MIT-BIH dataset noises, each obtained at a sampling frequency of 360 Hz, represent three types of noises frequently encountered in ECGs: baseline wander noise, a low-frequency noise produced by lead or subject movement[17], muscle artifact, caused by muscle contractions[21], and electrode motion artifact, which is caused by irregular movement of the electrodes during ECG recordings[17,21]. Each of these recordings were obtained using a standard 12-lead ECG recorder by positioning the electrodes on patient limbs such that the patients' ECG signals were not visible in the recordings (Fig. 4)[17].

Noise extraction from a portable device ECG

Real-world noise was also isolated from a 30-s, 300 Hz, single-lead ECG recording obtained using a KardiaMobile 6 L portable ECG device. The noise was extracted from the recording by applying a modified version of the Fourier transform-based approach previously used to denoise ECGs[29]. First, a fast Fourier transform (FFT) was applied to a 30-second ECG recording and the result was plotted in the frequency domain. Then, a threshold was manually selected to separate the high- and low-amplitude frequencies, which contained signal and noise, respectively. Finally, instead of computing the inverse FFT on the frequencies with amplitudes greater than this threshold, an inverse FFT was applied to all frequencies with amplitudes below the selected threshold, yielding the noise (Fig. 5).

Noise augmentation

While training the noise-adapted model, each ECG was included in the training dataset twice, with one of the four frequency ranges for the generated noise randomly selected each time an ECG was loaded. A 10-s sequence was then randomly selected from the full five-minute length of the selected noise recording. This noise sequence was then added to the standard ECG at a signal-to-noise ratio (SNR) randomly selected from a set of SNRs, including 0.50, 0.75, 1.00, and 1.25[30].

During model evaluation, the noised versions of the test set were generated by following the same procedure as with the noised training set with several key modifications. First, the noise added to ECGs in the test set was sourced from either the baseline wander, electrode motion, or muscle artifacts noise recordings from the MIT-BIH dataset or from the 30-s noise sample from the KardiaMobile 6 L portable ECG device. Second, the MIT-BIH noises and the portable ECG noise were all up-sampled from their original sampling frequencies, 360 Hz and 300 Hz, respectively, to a 500 Hz sampling frequency to match that of the clinical device ECGs. Third, the specific 10-second sequence of noise added to each ECG in the test set was randomly selected once and defined for each ECG, ensuring that every time any model was tested at any SNR for any specific noise, each individual ECG was always loaded with the same randomly selected sequence of noise. Model performance was evaluated separately for ECGs augmented with each noise type on a larger set of SNRs, including all the SNRs used in training and SNRs of 1.50, 1.75, and 2.00. Three
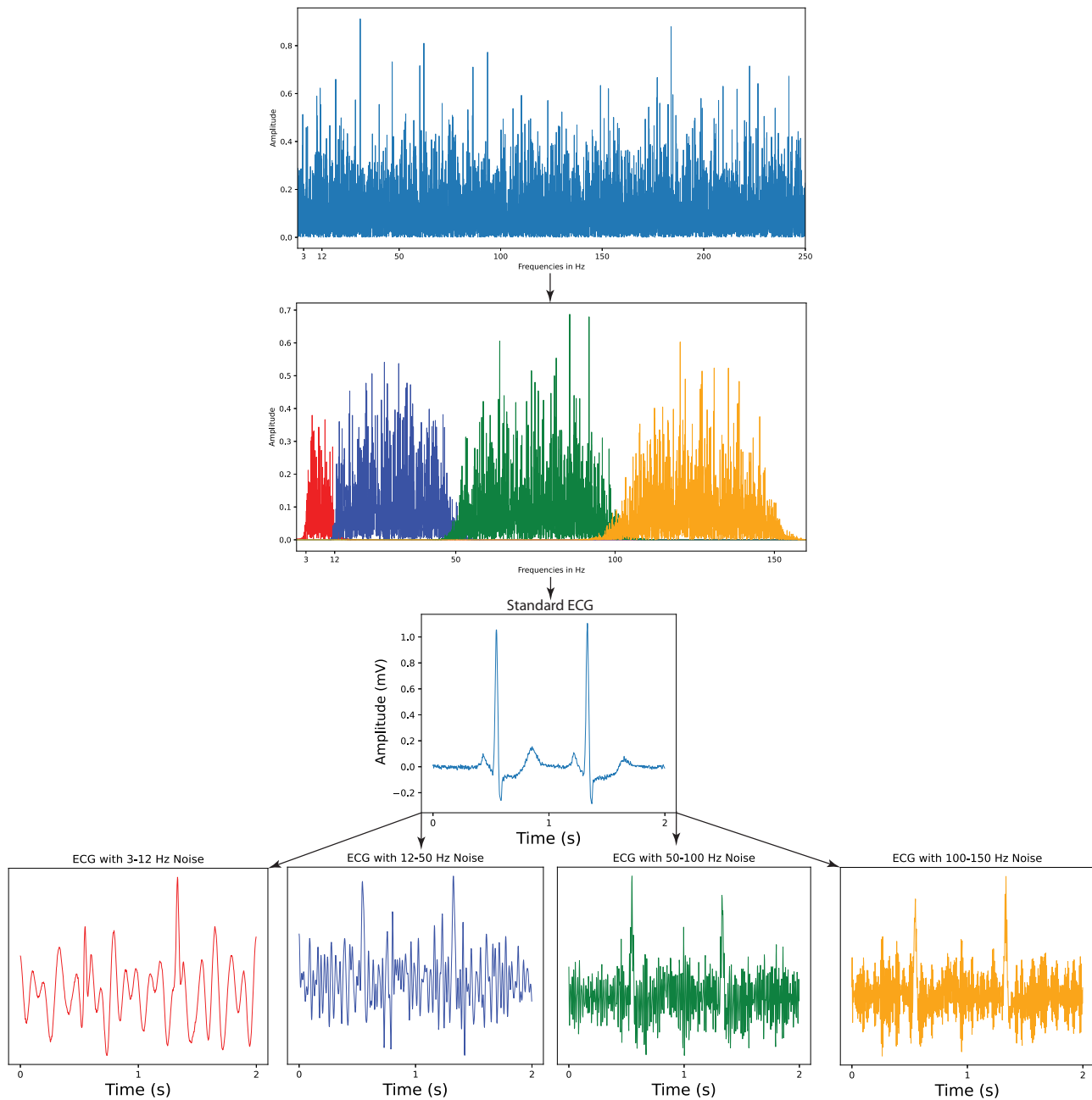
Fig. 3 Noise extraction and ECG augmentation for noise-adapted training. ECG electrocardiogram, Hz hertz, s seconds.

additional multiple-noise recordings were generated by separately combining a randomly selected, 10-second segment of portable ECG noise with randomly selected, 10-second segments of each of the three MIT-BIH noise recordings at a 1:1 ratio. These multiple-noise recordings were then used to evaluate both models' performance on ECGs augmented with multiple simultaneous noise signatures with the same larger set of SNRs.

Outcome label
Each ECG included in the dataset had a corresponding LVEF value from a paired echocardiogram within 15 days of the ECG. The cutoff for low LVEF was set as LVEF < 40%, a threshold present in most heart failure diagnosis guidelines[4], and consistent with prior work in this space[8,22].

Model training
All unique patients represented in the set of ECGs were then randomly subset on the patient level into training, validation, and held-out test sets (85%, 5%, 10%). We built and tested multiple convolutional neural network (CNN) models with varying numbers and sizes of convolutional layers and total model parameters. We selected the architecture yielding the highest area under the receiver operator characteristic (AUROC) curve on the validation set with the fewest number of parameters for the standard model. This architecture consisted of a (5000, 1, 1) input layer, corresponding to a 10-s, 500 Hz, Lead I ECG, followed by seven two-dimensional convolutional layers, each of which were followed by a batch normalization layer[31], ReLU activation layer, and a two-dimensional max-pooling layer. The output of the seventh convolutional layer was then taken as input into a fully connected network consisting of two dense layers, each of which
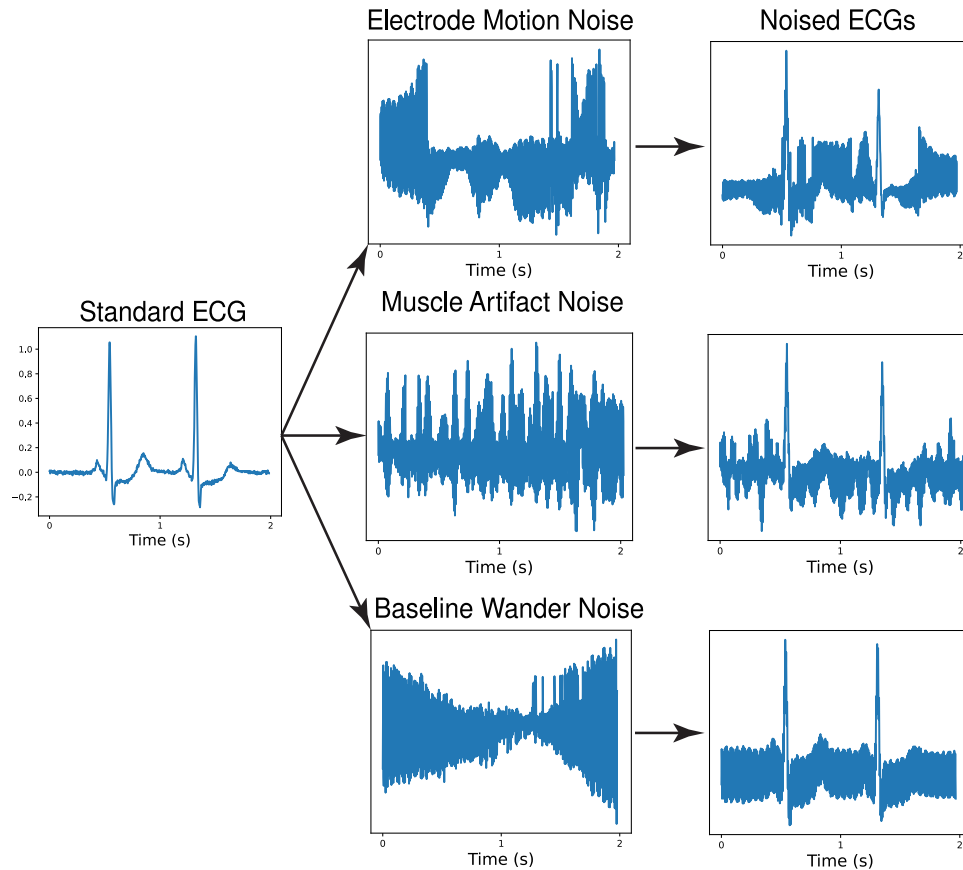
Fig. 4 Noise augmentation of test set with MIT-BIH real-world noises. ECG electrocardiogram, s seconds.

were followed by a batch normalization layer, ReLU activation layer, and a dropout layer with a dropout rate of 0.5[32]. The output layer was a dense layer with one class and a sigmoid activation function. Model weights were calculated for the loss function such that learning was not affected by the lower frequency of LVEF < 40% compared to the incidence of LVEF ≥ 40% using the effective number of samples class re-weighting scheme[33] (Fig. S2).

Both models were trained on the Keras framework in Tensor-Flow 2.9.1 and Python 3.9 using the Adam optimizer. First, the models were trained at a learning rate of 0.001 for one epoch. The learning rate was then lowered to 0.0001 and training was continued until performance on the validation set did not improve for three consecutive epochs. The epoch with the highest performance on the validation set was selected for each model.

Learning representation assessments for noise-adapted model

To visualize the variation between predictions for standard and noise-augmented data for each model, we first modified both the standard and noise-adapted models by removing their final output layers so both models instead produced the 320-dimensional vector output of the model's final fully connected layer. We then randomly selected a 1000 ECG subset from the held-out test set. For each of the two models, we generated the 320-dimension prediction vectors five times for each of the 1000 ECGs—once without noise, and once augmented at an SNR of 0.5 for each of the four noises used for testing. We then visualized the variation in the predictions separately for the standard and noise-adapted models using uniform manifold approximation and projection (UMAP), which constructs a two-dimensional representation of the 320-dimension prediction vectors[25]. The variation in predictions

was numerically assessed by the pair-wise calculation of the Euclidean distances between the 320-dimensional prediction vectors for the standard and noised data for each of the four noises. These Euclidean distances were then scaled on a per-model and per-noise basis by dividing by the total range of pair-wise Euclidean distances for each model and noise combination. The average scaled Euclidean distance and a 95% confidence interval was then calculated for each model and noise combination.

Statistical analysis

Summary statistics are presented as counts (percentages) and median (interquartile range, IQR), for categorical and continuous variables, respectively. Model performance was evaluated in the held-out test set both with and without added real-world noises. We used area under receiving operation characteristics (AUROC) to measure model discrimination. We also assessed area under precision-recall curve (AUPRC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and diagnostic odds ratio, and chose threshold values based on cutoffs that achieved sensitivity of 0.90 in validation data. We used a DeLong's test to compare AUROCs of the noise-adapted and standard models for each noise at each signal-to-noise ratio[23]. 95% confidence intervals for AUROC were calculated using DeLong's algorithm[23,24]. A paired t-test was used to compute the probability of overlap for the scaled Euclidean distances between last-layer projections of the noise-adapted and standard models. All analyses were performed using Python 3.9 and level of significance was set at an alpha of 0.05.
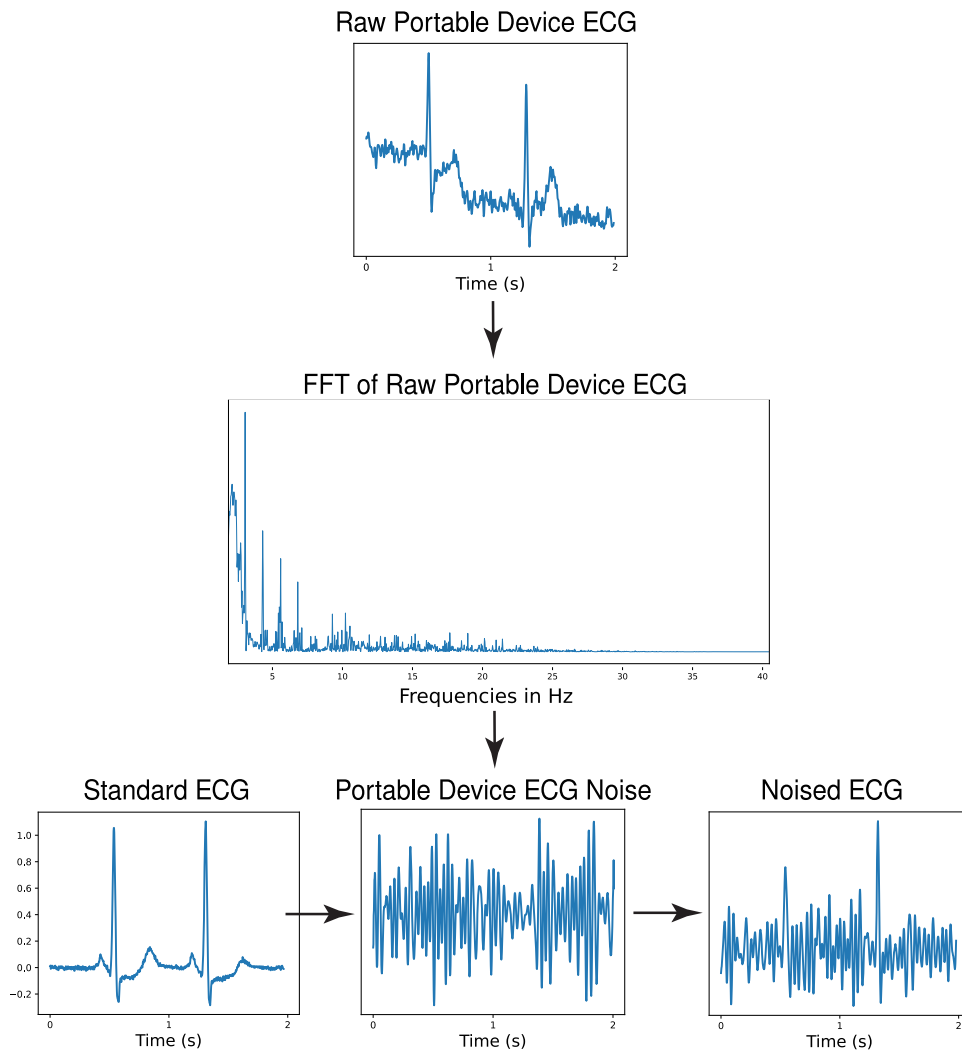
Raw Portable Device ECG

FFT of Raw Portable Device ECG

Frequencies in Hz

Standard ECG	Portable Device ECG Noise	Noised ECG

Fig. 5 Portable ECG device noise isolation and ECG noise augmentation. ECG electrocardiogram, s seconds, FFT Fast Fourier Transform, Hz hertz.

REFERENCES

1. Wang, T. J. et al. Natural history of asymptomatic left ventricular systolic dysfunction in the community. *Circulation* **108**, 977–982 (2003).
2. Srivastava, P. K. et al. Heart failure hospitalization and guideline-directed prescribing patterns among heart failure with reduced ejection fraction patients. *JACC Heart Fail* **9**, 28–38 (2021).
3. Wolfe, N. K., Mitchell, J. D. & Brown, D. L. The independent reduction in mortality associated with guideline-directed medical therapy in patients with coronary artery disease and heart failure with reduced ejection fraction. *Eur. Heart J. Qual Care Clin Outcomes.* **7**, 416–421 (2021).
4. Heidenreich, P. A. 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* **145**, e895–e1032 (2022).
5. Wang, T. J., Levy, D., Benjamin, E. J. & Vasan, R. S. The epidemiology of "asymptomatic" left ventricular systolic dysfunction: implications for screening. *Ann. Intern. Med.* **138**, 907–916 (2003).
6. Vasan, R. S. et al. Plasma natriuretic peptides for community screening for left ventricular hypertrophy and systolic dysfunction: the Framingham heart study. *JAMA* **288**, 1252–1259 (2002).
7. McDonagh, T. A. & McDonald, K. Screening for asymptomatic left ventricular dysfunction using B-type natriuretic peptide. *Congest. Heart Fail* **14**, 5–8 (2008).
8. Attia, Z. I. et al. Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram. *Nat. Med.* **25**, 70–74 (2019).
9. Galasko, G. I. W., Barnes, S. C., Collinson, P., Lahiri, A. & Senior, R. What is the most cost-effective strategy to screen for left ventricular systolic dysfunction: natriuretic peptides, the electrocardiogram, hand-held echocardiography, traditional echocardiography, or their combination? *Eur. Heart J.* **27**, 193–200 (2006).
10. Kamga, P., Mostafa, R. & Zafar, S. The use of wearable ECG devices in the clinical setting: a review. *Curr. Emerg. Hosp. Med. Rep.* **10**, 67–72 (2022).
11. Duarte, R. et al. Lead-I ECG for detecting atrial fibrillation in patients attending primary care with an irregular pulse using single-time point testing: A systematic review and economic evaluation. *PLoS ONE* **14**, e0226671 (2019).
12. Dörr, M. et al. The WATCH AF Trial: SmartWATCHes for detection of atrial fibrillation. *JACC Clin. Electrophysiol.* **5**, 199–208 (2019).

13. Mannhart, D. et al. Clinical validation of 5 direct-to-consumer wearable smart devices to detect atrial fibrillation: BASEL wearable study. *JACC Clin. Electrophysiol.* **9**, 232–242 (2023).

14. Bachtiger, P. et al. Point-of-care 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* **4**, e117–e125 (2022).

15. Attia, Z. I. 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* https://doi.org/10.1093/ehjdh/ztac030 (2022).

16. Satija, U., Ramkumar, B. & Manikandan, M. S. A review of signal processing techniques for electrocardiogram signal quality assessment. *IEEE Rev. Biomed. Eng.* **11**, 36–52 (2018).

17. Moody, G. B., Muldrow, W. & Mark, R. A noise stress test for arrhythmia detectors. *Comput. Cardiol.* **11**, 381–384 (1984).

18. Satija, U., Ramkumar, B. & Manikandan, M. S. A new automated signal quality-aware ecg beat classification method for unsupervised ECG diagnosis environments. *IEEE Sens. J.* **19**, 277–286 (2019).

19. Satija, U., Ramkumar, B. & Manikandan, M. S. A simple method for detection and classification of ECG noises for wearable ECG monitoring devices. in *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)* 164–169 (2015).

20. Charles, P. D., Esper, G. J., Davis, T. L., Maciunas, R. J. & Robertson, D. Classification of tremor and update on treatment. *Am. Fam. Physician* **59**, 1565–1572 (1999).

21. Friesen, G. M. et al. A comparison of the noise sensitivity of nine QRS detection algorithms. *IEEE Trans. Biomed. Eng.* **37**, 85–98 (1990).

22. Sangha, V. et al. Detection of left ventricular systolic dysfunction from electrocardiographic images. Preprint at https://www.medrxiv.org/content/10.1101/2022.06.04.22276000v2 (2022).

23. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).

24. Sun, X. & Xu, W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Process. Lett.* **21**, 1389–1393 (2014).

25. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. https://arxiv.org/abs/1802.03426 (2018).

26. Attia, Z. I. et al. Prospective evaluation of smartwatch-enabled detection of left ventricular dysfunction. *Nat. Med.* **28**, 2497–2503 (2022).

27. Pellikka, P. A. et al. Variability in ejection fraction measured by echocardiography, gated single-photon emission computed tomography, and cardiac magnetic resonance in patients with coronary artery disease and left ventricular dysfunction. *JAMA Netw. Open* **1**, e181456 (2018).

28. Jenkins, C. et al. Left ventricular volume measurement with echocardiography: a comparison of left ventricular opacification, three-dimensional echocardiography, or both with magnetic resonance imaging. *Eur. Heart J.* **30**, 98–106 (2009).

29. Kumar, A., Ranganatham, R., Komaragiri, R. & Kumar, M. Efficient QRS complex detection algorithm based on Fast Fourier Transform. *Biomed. Eng. Lett.* **9**, 145–151 (2019).

30. Johnson, D. Signal-to-noise ratio. *Scholarpedia J.* **1**, 2088 (2006).

31. Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. in *Proceedings of the 32nd International Conference on Machine Learning* (eds. Bach, F. & Blei, D.) Vol. 37, 448–456 (PMLR, 07–09, Jul 2015).

32. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. R. Improving neural networks by preventing co-adaptation of feature detectors. https://arxiv.org/abs/1207.0580 (2012).

33. Cui, Y., Jia, M., Lin, T.-Y., Song, Y. & Belongie, S. Class-balanced loss based on effective number of samples. https://arxiv.org/abs/1901.05555 (2019).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

R.K. conceived the study and accessed the data. A.K. and R.K. developed the model. A.K., V.S., and R.K. pursued the statistical analysis. A.K. and E.K.O. drafted the manuscript. All authors provided feedback regarding the study design and made critical contributions to writing of the manuscript. R.K. supervised the study, procured funding, and is the guarantor.

COMPETING INTERESTS

A.K., V.S., and R.K. are the coinventors of U.S. Provisional Patent Application No. 63/428,569, "Articles and methods for detection of hidden cardiovascular disease from portable electrocardiographic signal data using deep-learning". V.S. and R.K. are the coinventors of U.S. Pending Patent Application No. 63/346,610, "Articles and methods for format independent detection of hidden cardiovascular disease from printed electrocardiographic images using deep-learning". E.K.O. and R.K. are the coinventors of U.S. Provisional Patent Application No. 63/177,117 (unrelated to current work) and are co-founders of Evidence2Health, a precision health platform for evidence-based care. B.J.M. reported receiving grants from the National Institute of Biomedical Imaging and Bioengineering, National Heart, Lung, and Blood Institute, US Food and Drug Administration, and the US Department of Defense Advanced Research Projects Agency outside the submitted work; in addition, B.J.M. has a pending patent on predictive models using electronic health records (US20180315507A1). H.M.K. works under contract with the Centers for Medicare & Medicaid Services to support quality measurement programs, was a recipient of a research grant from Johnson & Johnson, through Yale University, to support clinical trial data sharing; was a recipient of a research agreement, through Yale University, from the Shenzhen Center for Health Information for work to advance intelligent disease prevention and health promotion; collaborates with the National Center for Cardiovascular Diseases in Beijing; receives payment from the Arnold & Porter Law Firm for work related to the Sanofi clopidogrel litigation, from the Martin Baughman Law Firm for work related to the Cook Celect IVC filter litigation, and from the Siegfried and Jensen Law Firm for work related to Vioxx litigation; chairs a Cardiac Scientific Advisory Board for UnitedHealth; was a member of the IBM Watson Health Life Sciences Board; is a member of the Advisory Board for Element Science, the Advisory Board for Facebook, and the Physician Advisory Board for Aetna; and is the co-founder of Hugo Health, a personal health information platform, and co-founder of Refactor Health, a healthcare AI-augmented data management company. R.K. is an associate editor at JAMA, and received support from the National Heart, Lung, and Blood Institute of the National Institutes of Health (under award K23HL153775) and the Doris Duke Charitable Foundation (under award, 2022060). He also receives research support, through Yale, from Bristol-Myers Squibb and Novo Nordisk.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41746-023-00869-w.

Correspondence and requests for materials should be addressed to Rohan Khera.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.