Journal of Cardiac Failure Vol. 00 No. 00 2020

# Wearable Patch-Based Estimation of Oxygen Uptake and Assessment of Clinical Status during Cardiopulmonary Exercise Testing in Patients With Heart Failure

MD MOBASHIR HASAN SHANDHI, MS,[1] SINAN HERSEK, PhD,[1] JOANNA FAN,[2] ERICA SANDER,[2] TERESA DE MARCO, MD,[2] J. ALEX HELLER, MS,[3] MOZZIYAR ETEMADI, MD, PhD,[3] LIVIU KLEIN, MD,[2,*] and OMER T. INAN, PhD[1,*]
*Atlanta, Georgia; San Francisco, California; and Chicago, Illinois*

## ABSTRACT

**Background:** To estimate oxygen uptake ($VO_2$) from cardiopulmonary exercise testing (CPX) using simultaneously recorded seismocardiogram (SCG) and electrocardiogram (ECG) signals captured with a small wearable patch. CPX is an important risk stratification tool for patients with heart failure (HF) owing to the prognostic value of the features derived from the gas exchange variables such as $VO_2$. However, CPX requires specialized equipment, as well as trained professionals to conduct the study.

**Methods and Results:** We have conducted a total of 68 CPX tests on 59 patients with HF with reduced ejection fraction (31% women, mean age 55 ± 13 years, ejection fraction 0.27 ± 0.11, 79% stage C). The patients were fitted with a wearable sensing patch and underwent treadmill CPX. We divided the dataset into a training−testing set ($n = 44$) and a separate validation set ($n = 24$). We developed globalized (population) regression models to estimate $VO_2$ from the SCG and ECG signals measured continuously with the patch. We further classified the patients as stage D or C using the SCG and ECG features to assess the ability to detect clinical state from the wearable patch measurements alone. We developed the regression and classification model with cross-validation on the training−testing set and validated the models on the validation set. The regression model to estimate $VO_2$ from the wearable features yielded a moderate correlation ($R^2$ of 0.64) with a root mean square error of 2.51 ± 1.12 mL · kg$^{-1}$ · min$^{-1}$ on the training−testing set, whereas $R^2$ and root mean square error on the validation set were 0.76 and 2.28 ± 0.93 mL · kg$^{-1}$ · min$^{-1}$, respectively. Furthermore, the classification of clinical state yielded accuracy, sensitivity, specificity, and an area under the receiver operating characteristic curve values of 0.84, 0.91, 0.64, and 0.74, respectively, for the training−testing set, and 0.83, 0.86, 0.67, and 0.92, respectively, for the validation set.

**Conclusions:** Wearable SCG and ECG can assess CPX $VO_2$ and thereby classify clinical status for patients with HF. These methods may provide value in the risk stratification of patients with HF by tracking cardiopulmonary parameters and clinical status outside of specialized settings, potentially allowing for more frequent assessments to be performed during longitudinal monitoring and treatment. (*J Cardiac Fail 2020;00:1−11*)

**Keywords:** Cardiopulmonary exercise test, Cardiovascular monitoring, Heart failure, Seismocardiogram, Wearable sensor.

A hallmark symptom of heart failure (HF) is exercise intolerance which often manifests through exertional dyspnea and fatigue. The degree of exercise intolerance is captured by subjective assessments (New York Heart Association functional class), quality of life questionnaires (eg, Kansas City Cardiomyopathy Questionnaire, Minnesota Living with Heart Failure questionnaire), and/or various objective exercise measures (eg, 6-minute walk distance). Cardiopulmonary exercise testing (CPX) is the most comprehensive exercise test performed in clinical settings to quantify the degree of myocardial impairment and pulmonary dysfunction.[1,2]

CPX has also evolved as an important diagnostic and prognostic tool to manage patients with HF by elucidating mechanisms of exercise intolerance, quantifying disease progression and facilitating recommendation for advanced therapies, such as heart transplantation or ventricular assist device implantation.[1−4] Peak oxygen uptake ($VO_2$), slope of minute ventilation (VE) and carbon dioxide production ($VCO_2$) and $VO_2$ at the anaerobic threshold are key CPX parameters that are used for this risk stratification and disease status quantification. Although CPX is a valuable

*From the [1]Department of ECE, Georgia Institute of Technology, Atlanta, Georgia; [2]School of Medicine, University of California San Francisco, San Francisco, California and [3]School of Medicine, Northwestern University, Chicago, Illinois.*

Reprint requests: Mobashir Hasan Shandhi, MS, Technology Square Research Building, 85 5th Street, Atlanta, Georgia 30308. Phone: +1-801-809-4209. E-mail: mobashir.shandhi@gatech.edu

*Senior authors.

See page 10 for disclosure information.

1071-9164/$ - see front matter



https://doi.org/10.1016/j.cardfail.2020.05.014

diagnostic and prognostic tool, it requires a specialized environment and trained professionals to conduct the study. Accordingly, although the information gained from CPX is valuable for patient assessment and titration of care, longitudinal CPX for patients with HF is cost prohibitive, inconvenient, and thus not feasible on a large scale. Using novel wearable technology, an unobtrusive and inexpensive alternative to the CPX, with the ability to potentially garner similar information as CPX from daily activities in home settings, could improve the remote monitoring and management of patients with HF.

Recently, our team has developed a wearable device[5] capable of measuring electrocardiogram (ECG) and seismocardiogram (SCG) signals and tested it in patients with HF.[6] The SCG represents the chest wall movements associated with the movement of blood in the heart, and includes features representing the ejection of blood through the aorta.[7] Our recent studies have shown that clinical status—degree of myocardial dysfunction and ability to augment cardiac output for patients with HF—can be assessed using SCG after exercise via pre-ejection period estimation and novel machine learning methodology.[6,8,9] However, although these results were promising, no group has demonstrated to date that an HF clinical state can be accurately classified using wearable SCG and ECG signals or that key parameters of cardiopulmonary function can be quantified from these signals.

In the current work, we recorded ECG and SCG signals using an updated version of the previously validated wearable patch[5] simultaneously with CPX for patients with HF with reduced ejection fraction (HFrEF). We extracted multiple features from these wearable signals and estimated $VO_2$ continuously throughout the course of exercise using state-of-the-art regression algorithms. We then classified the clinical state of the patients based on the changes in wearable signals associated with the exercise and compared the accuracy of this classification against gold standard clinical assessment based on CPX. Supplementary Fig. 1 shows a hypothetical system for longitudinal monitoring of patients with HF using our wearable patch.

## Methods

### Experimental Protocol

The study was conducted under a protocol reviewed and approved by the University of California, San Francisco, and the Georgia Institute of Technology Institutional Review Boards. All patients provided written consent before the procedure. We have conducted a total of 68 CPX tests in 59 patients with HFrEF (with 9 patients having 2 CPX tests separated by 253 ± 117 days). All of the patients were recruited from the cardiopulmonary stress test laboratory at the University of California, San Francisco. Only patients with HFrEF and a body mass index of less than 40 were considered for this study. We have separated the CPX tests into 2 groups of 44 CPX for a training−testing set and 24 CPX for a separate validation set. The 24 CPX tests for the validation set were obtained after the model was trained on the training−testing set.

Fig. 1A illustrates the experimental setup and placement of different sensors on each patient. Before starting the

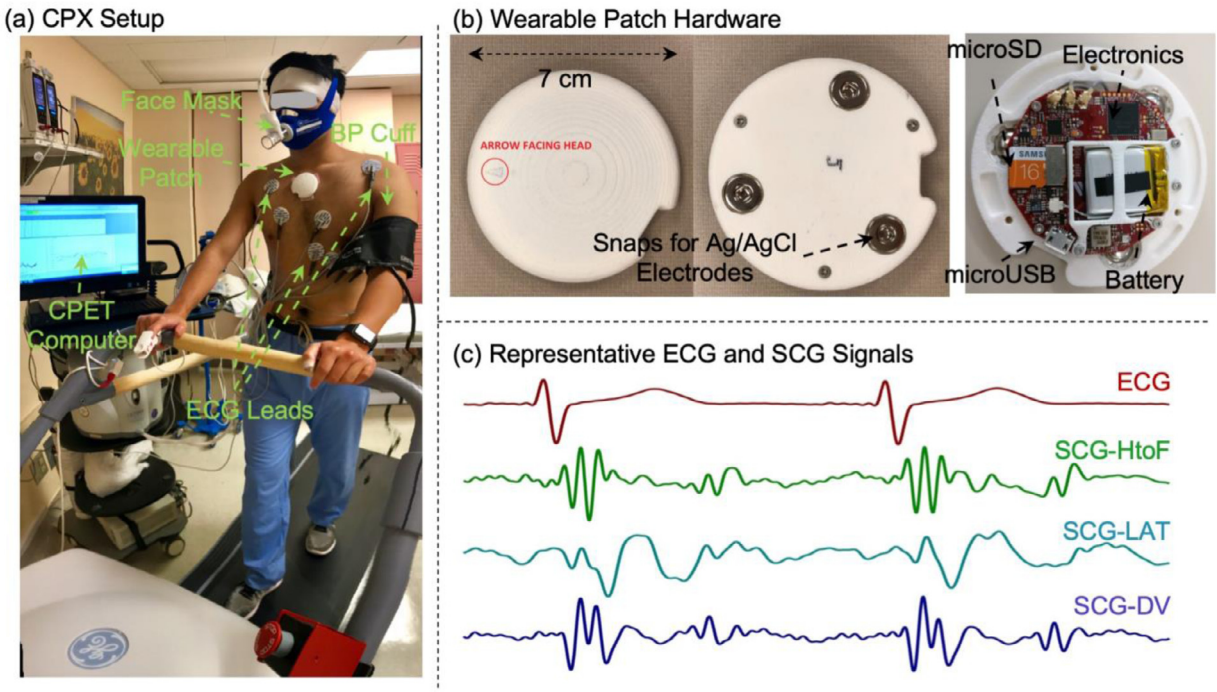


**Fig. 1.** Experimental setup with wearable patch hardware and representative cardiogenic signals. (A) The experimental setup with the patient walking on a treadmill, with all the cardiopulmonary exercise testing measurement sensors and wearable patch attached to the body. (B) The wearable patch top and bottom view with snaps for electrocardiogram (ECG) electrodes and internal hardware. (C) Representative ECG and triaxial SCG signals (head-to-foot [HtoF], dorsoventral [DV], and lateral [LAT]) from 1 patient in the study.

procedure, normal skin preparation methods were administered, and ECG leads were attached in a 12-lead ECG configuration. A gas exchange mask (Medgraphics) was placed on the patient. A finger pulse oximeter, a forehead pulse oximeter, and a blood pressure cuff were placed and minimal baseline spirometry data were collected to measure forced and slow vital capacity. The custom-built wearable device was placed just below the suprasternal notch. After placing all the sensors, all wires were taped down such that the patient could perform the protocol comfortably.

All CPX tests were performed on a treadmill (GE T2100) per the American College of Cardiology/American Heart Association Guidelines[10] and following the modified Naughton protocol.[11] Tests were terminated owing to general or leg fatigue, shortness of breath, angina, dizziness, or electrocardiographic evidence of ischemia or arrhythmia. Breath-by-breath measurements of respiratory rate, VE, $VO_2$, $VCO_2$, partial pressure of oxygen, and partial pressure of carbon dioxide were collected at rest, at zero grade low speed walk, during exercise, and during recovery. Heart rate (HR), rhythm, and oxygen saturations were continuously monitored with intermittent sphygmomanometry. ECG and SCG signals were obtained continuously using the wearable patch.

As an outcome of the CPX tests, patients were classified as American College of Cardiology/American Heart Association stage C HF ($n = 54$) or stage D HF ($n = 14$) based on the recommendations from 2 HF physicians (TDM, LK), following standard guidelines.[10,12,13] Patients were classified as stage D HF if they were recommended for heart transplant or ventricular assist device implant based on their peak $VO_2$ ($<14$ mL $\cdot$ kg$^{-1}$ $\cdot$ min$^{-1}$ or $<50\%$ predicted if women or obese) and $VE/VCO_2$ ratio ($>38$ if respiratory exchange ratio was $<1.05$).

### Sensing Hardware

Breath-by-breath data were collected using MGC Diagnostic/Medgraphics Ultima Series with Breeze suite 8.1.0.54 SP7 (software version number). ECGs (12-lead) were collected using GE Case V6.72. Pulse oximetry was measured using Radical 7 Masimo Rainbow Set.

For all patients, the wearable ECG and 3 axis SCG signals (head-to-foot [HtoF], dorsoventral [DV], and lateral [LAT]) were collected with a novel wearable patch as shown in Fig. 1B, which is an improvement on our previous version described in.[5] The patch has a diameter of 7 cm and weight of 39 g. All the wearable signals were sampled at 1 kHz. Fig. 1C shows representative ECG and triaxial SCG signals from the wearable patch. Fig. 2 illustrates the overall workflow used in this work.

### Data Analytics Techniques for Reducing Noise and Extracting Features from the Wearable SCG and ECG Signals

Whereas the CPX equipment captures breath-by-breath $VO_2$ data, the wearable patch captures one data point every 0.001 second (1 kHz sampling rate). A sliding window approach was used to combine all of the values from the SCG and ECG signals for the period in between breaths to estimate a single $VO_2$ value to compare against the gold standard. At a high level, the approach to estimating $VO_2$ was as follows: (1) the signals were preprocessed using our existing data analytics algorithms for SCG and ECG signals to decrease motion artifacts and other noise; (2) representative features, or signal characteristics, we hypothesized to be relevant for $VO_2$ estimation were extracted from the SCG and ECG signals; and (3) regression models were trained to mathematically estimate $VO_2$ from these SCG and ECG signal features for all CPX instances in the training−testing set and later validated in the validation set.

*Preprocessing and Noise Reduction.* All the signals from the wearable patch were synchronized with the breath-by-breath data from the CPX computer. The raw ECG and SCG signals from the wearable patch were digitally filtered (cut-off frequencies: 0.5−40.0 Hz for the ECG and 1−40 Hz for the SCG signals) to remove out-of-band noise. After filtering, a fourth SCG signal ($SCG_T$) was computed using vector summation on the 3 axes of the SCG. All the wearable signals were inspected for motion artifacts, and portions of the signals corrupted by motion artifacts were excluded from analysis. Details on the motion artifact removal algorithm are provided in the Supplementary Materials.

The ECG R-wave peaks were detected using a simple thresholding based peak detection method. The four SCG signals ($SCG_{HtoF}$, $SCG_{LAT}$, $SCG_{DV}$, and $SCG_T$) were segmented into individual heartbeats using the R-peaks from the ECG signals. Each heartbeat was windowed to a 600-ms duration from the R-peak. For each SCG signal, 10 consecutive heartbeats surrounding 1 $VO_2$ measurement from the CPX hardware were averaged time-point by time-point to obtain an ensemble averaged heartbeat (Fig. 2). Ensemble averaged heartbeats were computed across the whole recording with a step size of 1 heartbeat. Ensemble averaging was used to reduce noise and motion artifacts within each heartbeat.[14] This process resulted in a total of 46,673 ensemble-averaged heartbeats from 44 CPX instances in the training−testing set and 28,230 ensemble-averaged heartbeats from 24 CPX instances in the validation set. For each ECG signal, the R-to-R interval and instantaneous HR were calculated for each heartbeat and averaged in the same way as the ensemble averaged waveforms.

The average $VO_2$ measurements corresponding to each ensemble averaged heartbeat were computed to be used as the target variables for each ensemble averaged heartbeat (ie, the output variables against which the regression model was trained).

*Feature Extraction.* The next step toward estimating $VO_2$ from the measured signals involved extracting multiple features—or characteristics—that could then be input to a machine learning regression algorithm. A total of 17 features were automatically extracted (details in the

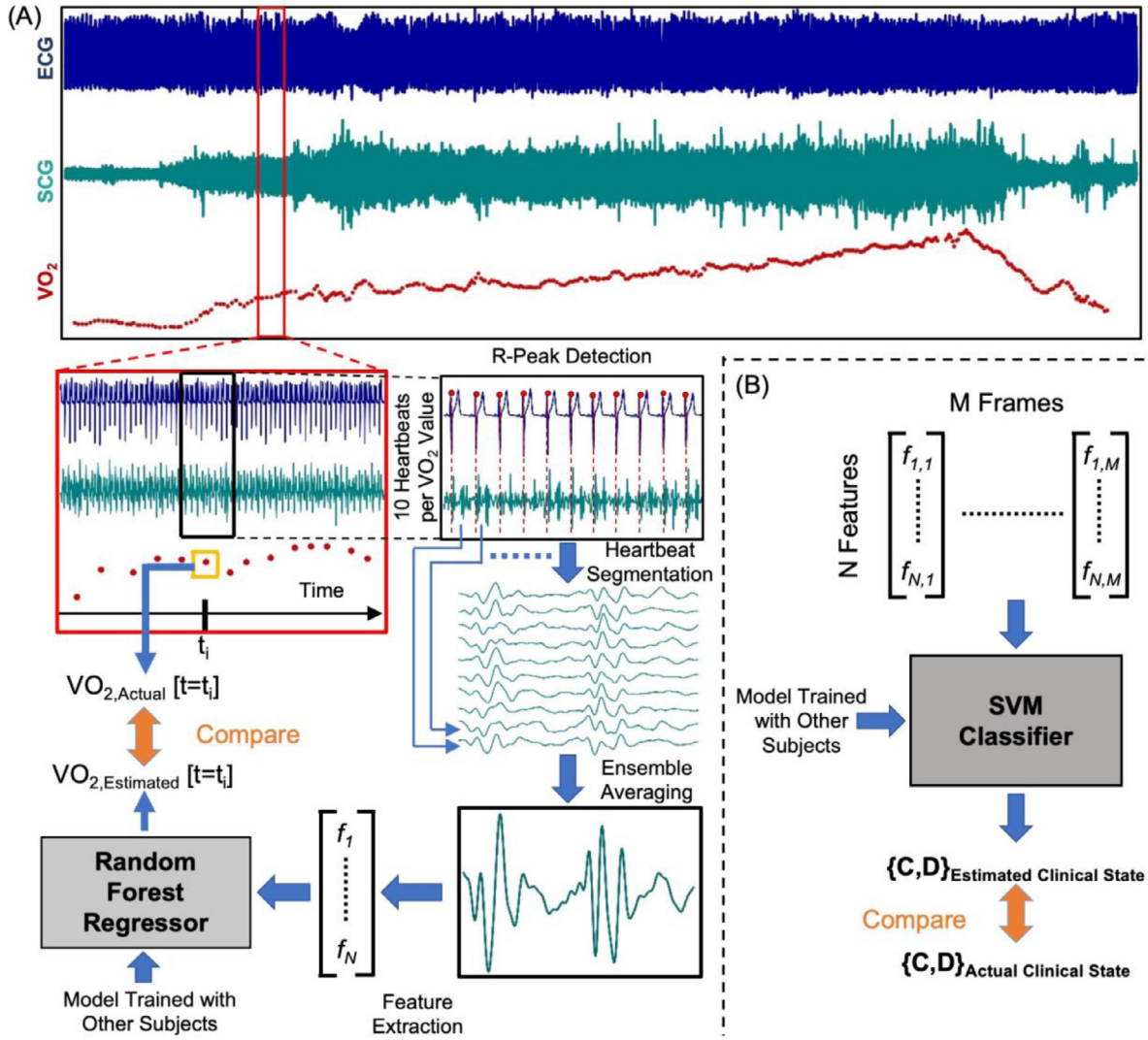


**Fig. 2.** Overview of the regression and classification techniques. (A) Wearable ECG and seismocardiogram (SCG) (only showing 1 axis of the signal for simplicity) signals were synchronized with breath-by-breath data from CPX computer. R-peaks of the ECG signal were detected and the SCG signals were segmented into heartbeats using corresponding R-peaks. Ten heartbeat frames from the SCG signals were averaged to get ensemble averaged heartbeats corresponds to 1 oxygen uptake ($VO_2$) value from breath-by-breath data from CPX and features were extracted from the averaged heartbeats. The features were fed into a random Forest regressor as estimators to estimate $VO_2$. Estimated $VO_2$ was compared with actual $VO_2$ to see the estimation accuracy. (b) The features from SCG and ECG were fed into a support vector machine (SVM) classifier with radial basis function kernel to estimate the clinical state of a patient and it was compared with the actual clinical state derived from CPX. Abbreviations as in Fig. 1.

Supplementary Materials) from each of the four SCG signals resulting in a total of 68 SCG features per ensemble averaged heartbeat. The feature extraction process was visually verified for each of the ensemble averaged heartbeats. The averaged R-to-R interval and instantaneous HR for each averaged heartbeat were used as ECG features.

Before training, a regression model to estimate $VO_2$, we removed outlier beats from the ensemble averaged SCG heartbeats using the Mahalanobis distance.[15] Details on this are provided in the Supplementary Material. The distance calculated (based on the feature set used) for each frame was added to the feature set and used for regression. The signal processing and feature extraction were performed in Matlab 2018a.

### Regression and Classification

*Regression Model.* For each $VO_2$ measurement recorded by the CPX equipment, a corresponding set of features from the SCG signals was derived using methods described elsewhere in this article. A regression algorithm was then designed and trained on the training set to mathematically estimate $VO_2$ from this set of features using part of the recorded data as a training set and the remainder of the data as a testing set. Specifically, we trained a random Forest (RF)[16] regression algorithm to estimate $VO_2$ from the wearable signal features and used leave-one-subject-out (LOSO) cross-validation[17] to evaluate the estimation accuracy. For all 44 CPX instances in the training−testing set, at each fold—or iteration of the cross-validation process—a

RF regression model was trained on the data from 43 patients (thus leaving 1 CPX instance out) to learn the relationship between features from the wearable sensors and the target variable $VO_2$. The resulting trained model was then used to estimate corresponding $VO_2$ values for the heartbeat frames from the left out CPX instance. This procedure was repeated 43 more times, leaving a different CPX instance out each time. This cross-validation method was used to develop a global regression model with optimized hyperparameters on the data in the training−testing set only. For the validation of the global model, the regression model (with the optimized hyperparameters) was trained on the whole training−testing set (all 44 CPX instances) and tested on the separate validation set (with 24 CPX instances). As a result, we obtained predictions of all target variables from all ensemble averaged heartbeats, from all 68 CPX instances.

Two figures of merit that are commonly used in the existing literature were used to evaluate the regression model and approach. First, the root mean squared error (RMSE) was calculated for each left out CPX instance: specifically, the error between the estimated $VO_2$ values and the CPX equipment measured $VO_2$ values across all breaths. The cross-validated RMSE was then calculated as the average of the RMSE scores from 44 folds in the training−testing set and 24 CPX instances in the validation set. Second, the coefficient of determination ($R^2$) between the true values and the cross-validated predictions of $VO_2$ across all CPX instances were calculated for the training−testing set and the validation set separately.

To assess the benefit of using a combined SCG/ECG approach for predicting $VO_2$, the RF regression approach was repeated for 3 different feature sets: the SCG features only, the ECG features only, and the combined SCG and ECG features. We compared the resulting cross-validated RMSE scores to assess the performance of each feature set to estimate $VO_2$. We performed statistical analysis on the cross-validation results from the different feature sets.

To understand the value of the information provided by SCG signals and our machine learning algorithm compared with the ECG-derived HR for estimating instantaneous $VO_2$, we trained an RF regression model using SCG signal features alone and a second model with HR alone using a simple linear regression model as used in literature to investigate the $VO_2$−HR relationship.[18,19] We performed the same LOSO cross-validation and calculated the cross-validated RMSE. We performed statistical analysis on the cross-validation results to compare the SCG signal feature-based model with the HR-based model.

*Classification.* In addition to estimating $VO_2$ using regression, we aimed to assess the ability to classify each patient's clinical status based on the wearable sensing data measured during treadmill exercise using classification. We used a machine learning classification technique to classify the patients with HF as stage C or stage D on a particular CPX procedure day using the wearable measurement alone and compared the estimated class with the true class based on the CPX outcome. Specifically, a support vector machine classifier with a radial basis function kernel[20,21] was used and classification performance was evaluated using LOSO cross-validation in the training−testing set and later validated on the separate validation set similarly as described in the regression model section. Details on the preprocessing of the wearable features for classifier are given in the Supplementary Materials.

Similar to the regression analysis approach with the training−testing set, for the classification task the classifier was trained on the features from 43 of the 44 CPX instances to map the features into an output of stage C and D state. We then used this classifier to predict the class of each heartbeat frame for the left-out patient. The majority vote (ie, class) of the heartbeats was chosen as the predicted class for the patient on that particular CPX procedure day. We repeated these steps 43 more times leaving a different CPX instance out each time. In this way, we obtained a predicted class for all CPX instances. Similarly, for the validation set, we trained the classification model (with hyperparameters tuned in the training−testing set of the classification task) on all 44 CPX instances in the training−testing set and estimated the class of each CPX instances in the validation set. Finally, we compared the estimated class to the true class of the patients from corresponding CPX outcome to calculate classification performance for the training−testing and validation set separately. The machine learning techniques for regression and classification were performed using Python 3.6.

*Estimation of Peak $VO_2$.* Because the peak $VO_2$ is among the key parameters extracted from a CPX procedure to assess the clinical status of the patients, we tried to see how our regression model, which estimates instantaneous $VO_2$, can be used in estimating peak $VO_2$ as well. The maximum of the estimated $VO_2$ values for a particular CPX instance was used as the estimated peak $VO_2$ value for that CPX and compared with the true measured peak $VO_2$ from corresponding CPX procedure, in a correlation and a Bland−Altman analysis. We have calculated the percentage error between estimated and true values of peak $VO_2$ and reported the average of the percentage error. We have used values from all 68 CPX instances, including both the training−testing and the validation CPX instances.

*Peak HR-Based Regression and Classification.* To understand the potential added value from SCG signals and our machine learning approach beyond peak HR alone, we have directly studied peak HR-based correlation and classification for the same dataset. We performed a simple correlation analysis (without any cross-validation) between peak $VO_2$ and peak HR. Further, we also applied exactly the same methodology (regression model with cross-validation) as for SCG-based peak $VO_2$ estimation and formed a model for estimating peak $VO_2$ from peak HR alone. In addition to the regression analysis, we classified the patients based on peak HR alone into stage C and stage D, in exactly the same manner we applied to our SCG-based features.

*Statistical Analysis.* We performed statistical analysis on the cross-validated RMSE results to compare regression results from different feature sets. Multiple comparison tests were performed on the RMSE results from the cross-validation. The Friedman test was performed to detect if statistical differences exist and the Wilcoxon signed-rank test was performed in post hoc testing for pairwise comparison. Additionally, for the post hoc testing, Benjamini−Hochberg correction for multiple comparison was performed on the *P* values. The demographics of patients in stage C and stage D were compared using the Student *t* test. In this work, *P* values of less than .05 were considered statistically significant.

## Results

Patient demographics and clinical characteristics are detailed in Table 1 and CPX characteristics are provided in Table 2. A survival analysis using subsequent events (left ventricular assisted device implantation, heart transplant, or cardiovascular death) occurring over 6 months after the initial collection of data is provided in the Supplementary Materials.

### Regression Model Comparison

Fig. 3A shows the correlation analysis between the actual (measured) $VO_2$ and the estimated $VO_2$ using the combined features from SCG and ECG for the training−testing set and Fig. 4A shows the corresponding analysis for the validation set. For the training−testing set, the regression model with the SCG features only performed better in estimating $VO_2$compared with the model using ECG features only: RMSE of 2.55 ± 1.16 mL · kg$^{-1}$ · min$^{-1}$ vs 3.75 ± 1.68 mL · kg$^{-1}$ · min$^{-1}$, respectively ($P < .001$) and a corresponding $R^2$ of 0.63 vs 0.19. Combining SCG and ECG features improved the estimation accuracy slightly compared with SCG features only, but the improvement was not significant ($P > .05$) with an RMSE of 2.50 ± 1.12 mL · kg$^{-1}$ · min$^{-1}$ and an $R^2$ of 0.64.

In the case of the validation set, similar results were obtained using SCG and ECG features separately: RMSE of 2.28 ± 1.04 mL · kg$^{-1}$ · min$^{-1}$ vs 3.52 ± 1.5 mL · kg$^{-1}$ · min$^{-1}$, respectively ($P < .001$) and a corresponding $R^2$ of 0.76 vs 0.36. Similarly, combining the SCG and ECG features improved the estimation accuracy (RMSE of 2.28 ± 0.93 mL · kg$^{-1}$ · min$^{-1}$ and $R^2$ of 0.76) slightly compared with SCG features only, although the improvement was not significant ($P > .05$).

**Table 1.** Patient Demographics and Characteristics

| | All CPX Instances ($N = 68$) | Stage C ($n = 54$) | Stage D ($n = 14$) | *P* Value |
|---|---|---|---|---|
| Age, years | 54.53 ± 12.68 | 54.81 ± 12.88 | 53.43 ± 12.28 | .53 |
| Sex | | | | |
| Male | 47 (69%) | 40 (74) | 7 (50) | |
| Female | 21 (31) | 14 (26) | 7 (50) | |
| Height, cm | 172.4 ± 9.14 | 172.67 ± 9.34 | 171.4 ± 8.57 | .59 |
| Weight, kg | 87.99 ± 18.39 | 87.57 ± 17.96 | 89.59 ± 20.63 | .68 |
| Body mass index, kg/m$^2$ | 29.53 ± 5.26 | 29.27 ± 4.85 | 30.51 ± 6.73 | .37 |
| Ejection fraction,% | 27.25 ± 10.64 | 26.21 ± 9.29 | 31.29 ± 14.46 | .13 |
| NYHA functional class | | | | |
| I | 12 (13) | 12 (18) | 0 (0) | |
| II | 24 (30) | 22 (36) | 2 (0) | |
| III | 32 (57) | 20 (45) | 12 (100) | |
| Orthopnea | 17 (27) | 13 (27) | 4 (27) | .73 |
| Bilateral leg edema | 12 (20) | 8 (18) | 4 (27) | .23 |
| Systolic blood pressure, mm Hg | 105 ± 15 | 105 ± 14 | 102 ± 19 | .41 |
| Diastolic blood pressure, mm Hg | 68 ± 10 | 68 ± 9 | 68 ± 15 | .85 |
| BNP, pg/mL | 568.4 ± 722.5 (23*) | 368 ± 514 (17*) | 1136.3 ± 962.1 (6*) | .02 |
| NT-proBNP, pg/mL | 1635.3 ± 1671.2 (31*) | 1783.4 ± 1687.7 (25*) | 1018.5 ± 1587.5 (6*) | .35 |
| Serum creatinine, mg/dL | 1.40 ± 1.43 (60*) | 1.49 ± 1.61 (46*) | 1.13 ± 0.43 (14*) | .38 |
| Loop diuretics, furosemide, mg/d | 83.7 ± 93.4 (68) | 64 ± 71 (65) | 146.4 ± 128.1 (79) | .01 |
| $\beta$-Blockers, bisoprolol, mg/d | 6.1 ± 3.8 (94) | 5.9 ± 3.9 (93) | 6.7 ± 3.7 (100) | .54 |
| ACE inhibitors, lisinopril, mg/d | 18.6 ± 15.5 (10) | 18.6 ± 15.5 (13) | 0 (0) | |
| ARB, losartan, mg/d | 54.8 ± 30.4 (19) | 61.1 ± 30.9 (17) | 40.6 ± 27.7 (29) | .28 |
| ARNI, sacubitril−valsartan, mg/d | 102.4 ± 64.2 (58) | 101.2 ± 64.8 (61) | 107.7 ± 65.5 (50) | .91 |
| MRA, spironolactone, mg/d | 29.8 ± 16.7 (85) | 29.3 ± 15.5 (81) | 31.6 ± 20.7 (100) | .64 |
| Subsequent events (OHT/VAD/death)† | 11 (16) | 7 (13) | 4 (29) | .16 |

Values shown are mean ± standard deviation or n (% of population) or mean ± standard deviation (% of population) unless otherwise indicated. Statistical significance between stage C and D patients in values, where applicable, was evaluated using an unpaired t test or a $\chi 2$ test.

ACE, angiotensin converting enzyme; ARB, angiotensin receptor blocker; ARNI, angiotensin receptor blocker—neprilysin inhibitor; BNP, B-type natriuretic peptide; CPX, cardiopulmonary exercise testing; MRA, mineralocorticoid receptor antagonist; NT-proBNP, N-terminal pro b-type natriuretic peptide; NYHA, New York Heart Association; OHT, orthotopic heart transplantation; VAD, ventricular assisted device implantation.

*Number of CPX test instances with available laboratory results.

†Subsequent events were recorded up to 6 months after the completion of the study. In the cases where 1 cardiopulmonary exercise testing patient had multiple events (eg, VAD, followed by transplant later), only the first occurring event was counted as subsequent events for a particular patient.

**Table 2.** Cardiopulmonary Exercise Test Responses

| | All CPX Instances ($N = 68$) | Stage C ($n = 54$) | Stage D ($n = 14$) | $P$ Value |
|---|---|---|---|---|
| Peak $VO_2$, mL · kg$^{-1}$ · min$^{-1}$ | 15.58 ± 4.82 | 17.21 ± 3.92 | 9.32 ± 1.93 | <0.001 |
| Percent predicted peak $VO_2$, % | 58 ± 21 | 63 ± 20 | 37 ± 9 | <0.001 |
| VE/$VCO_2$ slope | 33.35 ± 6.65 | 32.44 ± 6.48 | 36.82 ± 6.34 | 0.04 |
| $VO_2$ at anaerobic threshold, mL · kg$^{-1}$ · min$^{-1}$ | 11.79 ± 3.95 (62*) | 12.92 ± 3.33 (50*) | 7.08 ± 2.69 (12*) | <0.001 |
| Peak oxygen pulse, mL/beat | 12.02 ± 3.68 | 12.91 ± 3.47 | 8.59 ± 2.24 | <0.001 |
| Peak respiratory exchange ratio | 1.05 ± 0.12 | 1.07 ± 0.11 | 0.96 ± 0.12 | 0.002 |
| Exercise duration, seconds | 672 ± 235 | 743 ± 200 | 401 ± 148 | <0.001 |
| Peak heart rate, beats/min | 120.06 ± 23.8 | 124.57 ± 22.79 | 102.64 ± 19.77 | 0.002 |

CPX, cardiopulmonary exercise testing; VO2, oxygen uptake; VCO2, carbon dioxide production.

Values shown are mean ± standard deviation. Statistical significance between patients with stage C and stage D HF in values, where applicable, was evaluated using an unpaired t test

*Number of CPX instances with detectable anaerobic threshold points, Modified V-slope method was used to detect the anaerobic threshold points.

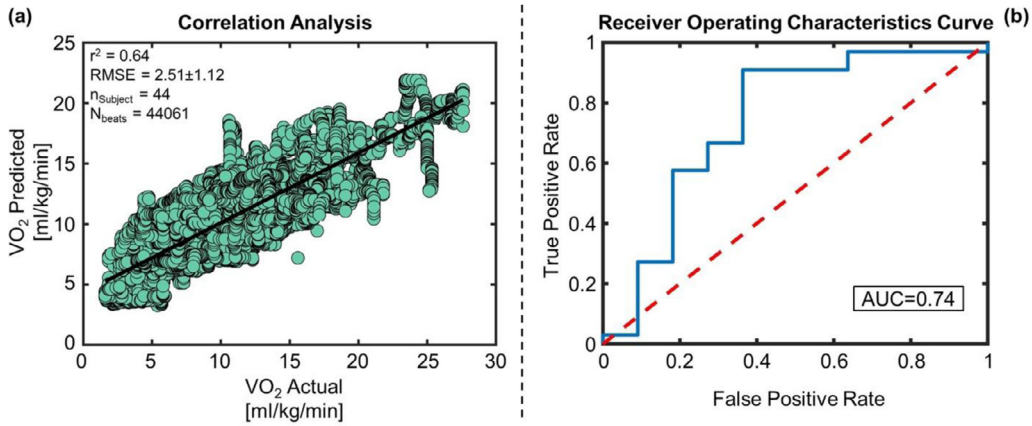


**Fig. 3.** Regression and classification results on the training−testing set. (A) Correlation analysis between $VO_2$ predicted vs $VO_2$ actual for the training and testing set. (B) The blue curve is showing the receiver operating characteristic (ROC) curve for the support vector machine (SVM) classifier with radial basis function kernel for the training and testing set. The red line is the ROC curve for classification based on random chance. The area under the blue ROC curve (AUC) is 0.74. Abbreviations as in Figs. 1 and 2.

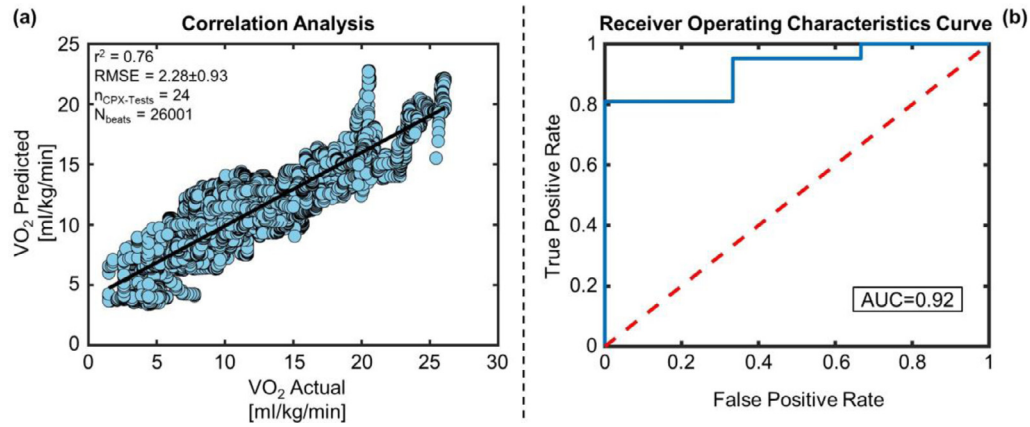


**Fig. 4.** Regression and classification results on the validation set. (A) Correlation analysis between $VO_2$ predicted vs $VO_2$ actual for the validation set. (B) The blue curve is showing the ROC curve for the SVM classifier with radial basis function kernel for the validation set. The red line is the ROC curve for classification based on random chance. The AUC of the blue ROC curve is 0.92. Abbreviations as in Figs. 2 and 3.

In the case of comparing SCG features with ECG-derived HR in estimating instantaneous $VO_2$, SCG features resulted in a significantly higher $R^2$ of 0.63 compared with 0.31 using HR only for the training−testing set ($P < .05$), and correspondingly 0.76 compared with 0.25 using HR only in the validation set ($P < .05$). The corresponding RMSE values were 2.55 ± 1.16 (SCG) vs 3.58 ± 1.54 mL · kg$^{-1}$ · min$^{-1}$ (HR) for the training−testing set and 2.28 ± 1.04 (SCG) vs 3.66 ± 1.74 mL · kg$^{-1}$ · min$^{-1}$ (HR) for the validation set.

**Table 3.** Confusion Matrix for Classification on the Training−Testing Set ($n = 44$)

| | Predicted Stage C | Predicted Stage D | |
|---|---|---|---|
| Actual stage C | 30 (TP) | 3 (FN) | 33 |
| Actual stage D | 4 (FP) | 7 (TN) | 11 |
| | 34 | 10 | |

FN, false negative; FP, false positive; TP, true positive; TN, true negative.
Accuracy = 0.84; sensitivity = 0.91; specificity = 0.64; positive predictive value = 0.88; negative predictive value = 0.7.

**Table 4.** Confusion Matrix for Classification on the Validation Set ($n = 24$)

| | Predicted Stage C | Predicted Stage D | |
|---|---|---|---|
| Actual stage C | 18 (TP) | 3 (FN) | 21 |
| Actual stage D | 1 (FP) | 2 (TN) | 3 |
| | 19 | 5 | |

Accuracy = 0.83; sensitivity = 0.86; specificity = 0.67; positive predictive value = 0.95; negative predictive value = 0.4.
Abbreviations as in Table 3.

### Classification

Table 3 and Table 4 show the classification results using the support vector machine with a radial basis function kernel for the training−testing and validation sets, respectively. Accuracy, sensitivity, and specificity obtained for the training−testing set were 0.84, 0.91, and 0.64, respectively, whereas for the validation set, they were 0.83, 0.86, and 0.67 respectively. Fig. 3B and Fig. 4B show the receiver operating characteristics curve of the classifier with an area under the curve of 0.74 and 0.92 for the training−testing and validation sets, respectively.

### Peak $VO_2$ Estimation

Fig. 5 shows the correlation analysis and Bland−Altman analysis between measured and estimated peak $VO_2$ values using SCG and ECG features for all 68 CPX instances, with a percentage error of 20.74% and an $R^2$ of 0.5.

### Peak HR-Based Regression and Classification

The correlation analysis between the peak $VO_2$ and the peak HR resulted in an $R^2$ of 0.23 for all 68 CPX instances. In contrast, estimation of the peak $VO_2$ using the peak HR using the same regression model and LOSO cross-validation approach used with SCG features resulted in an $R^2$ of 0.19 between the measured and estimated peak $VO_2$ values for all 68 CPX instances. The Bland−Altman confidence interval was calculated to be 17.1 mL · $kg^{-1}$ · $min^{-1}$ in this case. In the case of classifying the patients based on peak HR alone into stage C and stage D HF, the resultant area under the curve values for the receiver operating characteristics curve were 0.59 for the training−testing set and 0.54 for the validation set.

## Discussion

With this proof-of-concept study, we have shown the potential of a small, lightweight, wearable patch capable of measuring SCG and ECG to estimate beat-by-beat $VO_2$ estimation throughout a standard CPX procedure. Our results have shown that features from the wearable patch may capture the changes in cardiopulmonary demand during exercise and may be used to differentiate between stage C and stage D HFrEF. These promising initial results provide a foundation for determining cardiopulmonary variables and clinical status of patients with HF in their daily life and activities using wearable sensors. With further research, this approach could enable remote monitoring of these patients outside clinical settings.

An important finding in this work was that the features from the SCG signal were more salient in estimating $VO_2$ as compared with the ECG signal. Many Holter-type patches are currently available for ECG measurement, and have been used in studies for monitoring patients with HF.[22,23] Additionally, smartwatches are commercially available and can measure HR and possibly HR variability (provided there is minimal motion artifact). Although such commercially available tools are convenient and readily applicable to studies in patients with HF, the results from

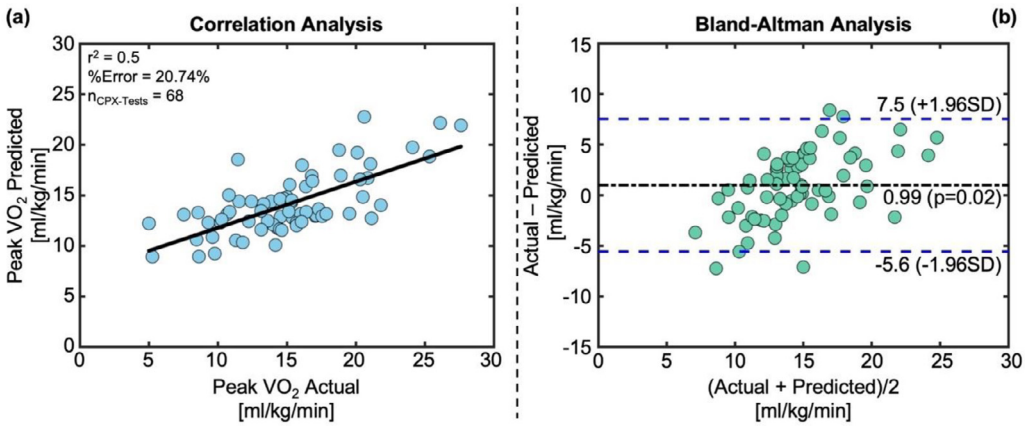


**Fig. 5.** Results of peak $VO_2$ estimation. (A) Correlation analysis and (B) Bland−Altman analysis between predicted peak $VO_2$ vs actual peak $VO_2$ for all 68 CPX instances used in the study. Abbreviations as in Figs. 1 and 2.

this article demonstrate that HR-based features may not provide sufficient value in assessing cardiopulmonary health in patients with HF during exercise. Rather, approaches using a combination of ECG- and SCG-based sensing are needed such that $VO_2$ and a patient's clinical status can be accurately determined during exercise. This result is consistent with our prior work where changes in the SCG signal in response to a 6-minute walk test were found to be more salient in assessing clinical state for patients with HF than ECG or HR features alone.[6]

Another important, and perhaps surprising, finding in this work is that the signal quality of the SCG signals measured during treadmill exercise in patients with low signal levels overall (patients with HF) was sufficiently high to enable accurate estimation of $VO_2$. The 2 main factors allowing such high signal quality to be obtained during exercise from a signal that has typically been limited to low motion/vibration environments only were the following: (1) the improved wearable patch we have developed that was used in this work employs the lowest noise microelectromechanical system accelerometer available, with a noise floor that is 2.5 times lower than any other microelectromechanical system accelerometer used in prior studies to the best of our knowledge; and (2) the direct coupling of the patch to the chest wall at the sternum with a triangular configuration of ECG electrodes provides a rigid and robust mechanical interface to the body from which SCG signals can be reliably recorded, even in the presence of motion artifacts. Thus, the results of this work may form a foundation upon which future efforts focused on assessing the mechanical aspects of left ventricular function during movement can be designed and realized.

From the result with peak $VO_2$ estimation, it is apparent that the model underestimated and overestimated peak $VO_2$ for very high and low values of measured peak $VO_2$, respectively. This limitation is well-known in machine learning-based models, because it will try to produce results close to the overall mean of the distribution rather than extreme values. Increasing the number of patients with a broader spectrum of exercise capabilities may decrease the estimation accuracy for the extreme peak $VO_2$ values in future studies. Also, a point to note here is that the regression model presented here was trained to learn the underlying relationship of SCG and ECG features with beat-by-beat $VO_2$, not only peak $VO_2$. Maximal effort covers only a small portion of the CPX protocol. This can be attributed to the comparatively lower performance of peak $VO_2$ estimation in our analysis compared with the estimation of the beat-by-beat estimation of $VO_2$.

Although the measurement of $VO_2$ values at less than peak may not currently be clinically relevant, one can imagine that with the capability of estimating $VO_2$ accurately for submaximal exercise tasks, such as walking upstairs or outdoors, the ability to assess patients with HF outside of clinical settings may be enhanced. Thus, in future clinical care scenarios where digital data collection methodologies are being leveraged, the measurement of $VO_2$ in submaximal tasks could potentially become an important and clinically relevant capability.

Comparing the results of peak $VO_2$ estimation using our method with peak HR-based method demonstrates that augmenting HR with cardiomechanical features may result in a higher correlation coefficient and smaller confidence interval for estimating peak $VO_2$. The SCG signal features resulted in more robust classification performance for separating patients with stage C and D HF as well. Future work should focus on improving the estimation accuracy of peak $VO_2$ from wearable SCG and ECG signals.

The peak $VO_2$ was used along with the $VE/VCO_2$ ratio to determine the severity of HF (stage C and D) in these patients. In our regression analysis, the algorithm was trained to learn the underlying features of the SCG and ECG signals to estimate instantaneous $VO_2$ throughout the CPX protocol, whereas the classification algorithm was trained to learn the underlying features of the SCG and ECG signals to determine the severity (stage C vs stage D) of HF for these patients. The regression model can be used to estimate $VO_2$ during submaximal exercise levels as well as maximal effort, whereas classification tasks can give 1 label to the whole CPX test. These preliminary findings, however, need verification in a larger patient population with a variety of exercise levels. Because peak $VO_2$ played a key role in determining the true class of the patients, there can be some common SCG and ECG features that were used by both regression and classification models. Future work should examine both SCG and ECG features from both maximal and submaximal exercise to relate to the severity of HF and investigate the underlying physiological relationship between them.

It should also be noted that, although the regression and classification approaches used in this work are "black box," as is the case for any machine learning technique, the relative importance of SCG frequency domain features vs ECG-HR features does provide some insight into possible physiologic mechanisms behind the relationship between SCG signals and $VO_2$. Specifically, the changes in the frequency domain characteristics of the signals might suggest the presence of nonlinearity (ie, harmonics) in the vibrations of the chest in response to the heartbeat at higher levels of exercise and $VO_2$. Another potential mechanistic link could be in the relationship between some frequencies of the SCG signal and stroke volume, which is an important factor constituting $VO_2$. Nevertheless, these mechanistic links are conjecture at this point and should be investigated in the future using studies with direct hemodynamic measurements (eg, right heart catheterization) taken simultaneously with SCG signals to characterize the origin and characteristics of the signal in the context of left ventricular function and health.

This study also has several limitations that should be noted. Our dataset had only 21% patients with stage D HF (25% in the training−testing set and 13% in the validation set), resulting in higher peak $VO_2$ for patients with stage D HF. For a few cases of patients with stage C HF with a very

high peak $VO_2$compared with the rest of the population, our model underestimated their $VO_2$ and corresponding peak $VO_2$ estimation. In future studies, we will increase the number of patients and incorporate patients with a broader spectrum of exercise capabilities, which may decrease the estimation error for these extreme cases. Similarly, our classification model classified 30 of 33 stage C CPX instances accurately, whereas 7 of 11 stage D CPX instances were accurately classified in the training−testing set. For the validation set, it classified 18 of 21 stage C CPX instances accurately, whereas 2 of 3 stage D CPX instances were accurately classified. The comparatively poor performance in the classification of patients with stage D HF can be associated with a smaller number of patients with stage D HF ($n = 14$) in our dataset, the shorter duration of exercise compared with patients with stage C HF, and greater pathophysiologic differences among patients owing to various HF-related diseases. Increasing the number of patients with stage D HF in future studies should increase the classification accuracy for patients with stage D HF as well.

This preliminary study demonstrated the potential of using advanced machine learning algorithms to estimate continuous $VO_2$ throughout the CPX procedure and clinical status of patients with HF, both in a training−testing set and a separate validation set. Results in the validation set were comparatively better than the training−testing set. One reason can be that our validation set had fewer patients with stage D HF by chance compared with the training−testing set, and our model performed well for the patients with stage C HF because it has more patients with stage C HF to learn from in the training phase. Incorporating more patients with stage D HF in future studies should verify these initial findings in a large set of population pool.

In this work, we have only estimated $VO_2$. Future work should focus on estimating other gas exchange variables (eg, $VCO_2$, VE, and tidal volume) from the CPX and to investigate the underlying mechanisms. Additionally, we have collected data only from patients with HFrEF. Future studies can assess the efficacy of this sensor in patients with HF with preserved ejection fraction. In addition, these tests were performed in a controlled clinical setting with trained professionals. The data from home or an unsupervised setting may be of lower quality compared with the data obtained here. Future studies can elucidate whether wearable SCG and ECG parameters measured during normal activities of daily living can be predictive of the parameters measured during extensive CPX.

## Conclusions

We have demonstrated that a wearable chest patch-based sensor capable of recording ECG and SCG may be used to estimate $VO_2$ from CPX for patients with HF using a global regression model and may facilitate determination of clinical state of the patient. We thus demonstrated that wearable sensors can potentially be used to monitor cardiopulmonary health and to stratify disease risk for patients with HF. The approach described in this work may thus provide the capability to perform longitudinal CPX testing for patients with HF in clinical and hospital settings such that treatment and management can be titrated and personalized based on physiologic state. Because CPX testing has been established as a valuable technique in assessing patient state for HF, broadening the ability to perform such testing in longitudinal patient management may improve the quality of care and life for patients with HF. Future studies should verify these preliminary findings in a larger patient population with a wider spectrum of exercises, in both a clinical environment and normal daily living activities.

## Clinical Perspectives

Wearable technologies have the potential to allow monitoring of patients with HF in the ambulatory setting. In this work, we have shown that a wearable patch can estimate oxygen consumption during cardiopulmonary stress testing and can assist in the stratification of patients with HF based on the severity of their disease. Future work will investigate tracking physiologic changes and responses to interventions during daily activities at home in this patient population.

## Financial Support

Supported in part by the National Heart, Lung and Blood Institute under R01HL130619. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## Disclosures

O. T. Inan is a Scientific Advisor for Physiowave, Inc.

## Acknowledgments

Dr Klein would like to acknowledge the research support from gifts from Joyce and Roger Isaacs and George Doubleday.

## Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.cardfail.2020.05.014.

## References

1. Malhotra R, Bakken K, D'Elia E, Lewis GD. Cardiopulmonary exercise testing in heart failure. JACC Heart Failure 2016;4:607–16.
2. Myers J, Arena R, Cahalin LP, Labate V, Guazzi M. Cardiopulmonary exercise testing in heart failure. Curr Problems Cardiol 2015;40:322.
3. Balady GJ, Arena R, Sietsema K, Myers J, Coke L, Fletcher GF, et al. Clinician's guide to cardiopulmonary exercise testing in adults: a scientific statement from the American Heart Association. Circulation 2010;122:191–225.
4. Haykowsky MJ, Tomczak CR, Scott JM, Paterson DI, Kitzman DW. Determinants of exercise intolerance in patients

with heart failure and reduced or preserved ejection fraction. J Appl Physiol 2015;119:739–44.
5. Etemadi M, Inan OT, Heller JA, Hersek S, Klein L, Roy S. A wearable patch to enable long-term monitoring of environmental, activity and hemodynamics variables. IEEE Transact Biomed Circuits Syst 2016;10:280–8.
6. Inan Omer T, Baran Pouyan M, Javaid Abdul Q, Dowling S, Etemadi M, Dorier A, et al. Novel wearable seismocardiography and machine learning algorithms can assess clinical status of heart failure patients. Circ Heart Fail 2018;11:e004313.
7. Inan OT, Migeotte P-F, Park K-S, Etemadi M, Tavakolian K, Casanella R, et al. Ballistocardiography and seismocardiography: a review of recent advances. IEEE J Biomed Health Inform 2015;19:1414–27.
8. Inan OT, Dorier A, Dowling S, Javaid AQ, Etemadi M, Roy S, Roy S, Marco TD, Klein L, et al. Activity-contextualized wearable ballistocardiogram measurements can classify decompensated versus compensated heart failure patients. Circulation 2016;134:A17525.
9. Inan OT, Javaid AQ, Dowling S, Ashouri H, Etemadi M, Heller JA, Roy S, Klein L, et al. Using ballistocardiography to monitor left ventricular function in heart failure patients. J Cardiac Fail 2016;22:S45.
10. Guazzi M, Adams V, Conraads V, Halle M, Mezzani A, Vanhees L, et al. EACPR/AHA Scientific Statement. Clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. Circulation 2012;126:2261–74.
11. American College of Sports Medicine (ACSM). ACSM's guidelines for exercise testing and prescription. Philadelphia: Lippincott Williams & Wilkins; 2013.
12. Guazzi M, Arena R, Halle M, Piepoli MF, Myers J, Lavie CJ. 2016 focused update: clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. Circulation 2016;133:e694–711.
13. Mehra MR, Canter CE, Hannan MM, Semigran MJ, Uber PA, Baran DA, et al. The 2016 International Society for Heart Lung Transplantation listing criteria for heart transplantation: a 10-year update. J Heart Lung Transplant 2016;35:1–23.
14. Sörnmo L, Laguna P. Bioelectrical signal processing in cardiac and neurological applications. New York: Academic Press; 2005.
15. Mahalanobis PC. On the generalized distance in statistics. Proceedings of National Institute of Sciences (India) 1936;2:49–55.
16. Breiman L. Random forests. Machine Learn 2001;45:5–32.
17. Stone M. Cross-validatory choice and assessment of statistical predictions. J R Stat Soc B Methodol 1974;36:111–33.
18. Bot S, Hollander A. The relationship between heart rate and oxygen uptake during non-steady state exercise. Ergonomics 2000;43:1578–92.
19. Loe H, Rognmo Ø, Saltin B, Wisløff U. Aerobic capacity reference data in 3816 healthy men and women 20−90 years. PloS One 2013;8:e64319.
20. Platt J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in Large Margin Classifiers 1999;10:61–74.
21. Cortes C, Vapnik V. Support-vector networks. Machine Learn 1995;20:273–97.
22. Steinhubl SR, Waalen J, Edwards AM, Ariniello LM, Mehta RR, Ebner GS, et al. Effect of a home-based wearable continuous ECG monitoring patch on detection of undiagnosed atrial fibrillation: the mSToPS randomized clinical trial. JAMA 2018;320:146–55.
23. Levine DM, Ouchi K, Blanchfield B, Diamond K, Licurse A, Pu CT, et al. Hospital-level care at home for acutely ill adults: a pilot randomized controlled trial. J Gen Intern Med 2018;33:729–36.