Journal Pre-proof

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 , Sinan Hersek PhD ,
Joanna Fan , Erica Sander , Teresa De Marco MD ,
J. Alex Heller MS , Mozziyar Etemadi MD, PhD , Liviu Klein MD ,
Omer T. Inan PhD

Please cite this article as: Md Mobashir Hasan Shandhi MS , Sinan Hersek PhD , Joanna Fan ,
Erica Sander ,   Teresa De Marco MD ,   J. Alex Heller MS ,   Mozziyar Etemadi MD, PhD ,
Liviu Klein MD , Omer T. Inan PhD , Wearable Patch Based Estimation of Oxygen Uptake and
Assessment of Clinical Status during Cardiopulmonary Exercise Testing in Patients with Heart Failure,
Journal of Cardiac Failure (2020), doi: https://doi.org/10.1016/j.cardfail.2020.05.014

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[a] MS, Sinan Hersek[a] PhD, Joanna Fan[b], Erica Sander[b], Teresa De Marco[b] MD, J. Alex Heller[c] MS, Mozziyar Etemadi[c] MD, PhD, Liviu Klein[b*] MD and Omer T. Inan[a*] PhD

[a]Department of ECE, Georgia Institute of Technology, Atlanta, GA, USA,

[b]School of Medicine, University of California San Francisco, San Francisco, CA, USA,

[c]School of Medicine, Northwestern University, Chicago, IL, USA

* Senior authors

Brief Title: Wearable Monitoring of Cardiovascular Health in Heart Failure Patients

Address for correspondence:

Md Mobashir Hasan Shandhi,

Technology Square Research Building,

85 5th Street, Atlanta, Georgia 30308.

Phone: +1-801-809-4209

Email: mobashir.shandhi@gatech.edu

Abstract

Objective: 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.

Background: CPX is an important risk stratification tool for patients with heart failure (HF) due 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: We have conducted a total of 68 CPX tests on 59 subjects with HF with reduced ejection fraction (31% women, mean age 55±13 years, ejection fraction 0.27±0.11, 79% stage C). The subjects were fitted with a wearable sensing patch and underwent treadmill CPX. We divided the dataset into a training-testing (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.

Results: 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 (RMSE) of 2.51±1.12 ml.kg^{-1}.min^{-1} on the training-testing set, whereas R^2 and RMSE 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.

Conclusion: Wearable SCG and ECG can assess CPX oxygen uptake and thereby classify clinical status for patients with HF. These methods may provide value in 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.

Keywords: Cardiopulmonary Exercise Test; Cardiovascular Monitoring; Heart Failure; Seismocardiogram; Wearable Sensor.

Abbreviations list:

CPX = cardiopulmonary exercise testing

SCG = seismocardiogram

ECG = electrocardiogram

HFrEF = heart failure with reduced ejection fraction

VO_2 = oxygen uptake

HR= heart rate

RF = random forest

RMSE = root mean squared error

SVM = support vector machine

LOSO = leave-one-subject-out

Introduction:

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 (e.g. Kansas City Cardiomyopathy Questionnaire, Minnesota Living with Heart Failure questionnaire), and/or various objective exercise measures (e.g. 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 (VAD) implantation (1-4). Peak oxygen uptake (peak VO_2), slope of minute ventilation (VE) and carbon dioxide production (VCO_2) and VO_2 at anaerobic threshold (AT) are key CPX parameters that are used for this risk stratification and disease status quantification. While CPX is a valuable diagnostic and prognostic tool, it requires a specialized environment and trained professionals to conduct the study. Accordingly, while 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 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 following exercise via pre-ejection period (PEP) estimation and novel machine learning methodology (6,8,9). However, while these results were promising, no group has demonstrated to date that HF clinical state can be accurately classified using wearable SCG and ECG signals, nor 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 and 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. Supplemental Figure 1 shows a hypothetical system for longitudinal monitoring of patients with HF using our wearable patch.

Methods:

A. Experimental Protocol

The study was conducted under a protocol reviewed and approved by University of California, San Francisco (UCSF) and Georgia Institute of Technology Institutional Review Boards (IRBs). All subjects provided written consent before the procedure. We have conducted a

total of 68 CPX tests in 59 HFrEF patients (with 9 patients having two CPX tests separated by 253±117 days). All of the subjects were recruited from the cardiopulmonary stress test lab at UCSF. Only subjects with HFrEF and body mass index (BMI) < 40 were considered for this study. We have separated the CPX tests into two 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.

Figure 1(a) illustrates the experimental setup and placement of different sensors on each subject. Before starting the 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 subject. A finger pulse oximeter, a forehead pulse oximeter and a blood pressure (BP) cuff were placed and minimal baseline spirometry data was 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 subject 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 due 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 (PO_2) and partial pressure of carbon dioxide (PCO_2) 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 two HF physicians (TDM, LK), following standard guidelines (10,12,13). Patients were classified as stage D if they were recommended for heart transplant or VAD implant based on their peak VO_2 (< 14 mL/kg/min or < 50% predicted if women or obese) and VE/VCO_2 ratio (> 38 if respiratory exchange ratio was < 1.05).

B. 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 subjects, the wearable ECG and three axis SCG signals (head-to-foot, HtoF, dorso-ventral, DV, and lateral, LAT) were collected with a novel wearable patch as shown in Figure 1(b), which is an improvement upon our previous version described in (5). The patch has a diameter of 7 cm and weight of 39 gm. All the wearable signals were sampled at 1 kHz. Figure 1(c) shows representative ECG and tri-axial SCG signals from the wearable patch. Figure 2 illustrates the overall workflow used in this work.

C. Data Analytics Techniques for Reducing Noise and Extracting Features from the Wearable SCG and ECG signals

While the CPX equipment captures breath-by-breath VO_2 data, the wearable patch captures one data point every 0.001 sec (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 pre-processed using our existing data

analytics algorithms for SCG and ECG signals to reduce 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.

Pre-processing 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 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 three 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, ten consecutive heartbeats surrounding one VO_2 measurement from the CPX hardware were averaged time-point by time-point to obtain an *ensemble averaged* heartbeat (shown in Figure 2). Ensemble averaged heartbeats were computed across the whole recording with a step size of one heartbeat. Ensemble averaging was used to reduce noise and motion artifacts within each heartbeat (14). This 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 was 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 (i.e., 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 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.

D. 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 above. 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 subjects (thus leaving one 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 hyper-parameters on the data in the training-testing set only. For the validation of the global model, the regression model (with the optimized hyper-parameters) 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 employed 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-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 three 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 to 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 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 subjects with HF as stage C or stage D on a particular CPX procedure day using the wearable measurement alone and compared the estimated class to the true class based on the CPX outcome. Specifically, a support vector machine (SVM) classifier with a radial basis function (RBF) 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.

Similarly 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 subject. The majority vote (i.e. class) of the heartbeats was chosen as the predicted class for the subject 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 predicted class for all CPX instances. Similarly, for the validation set, we trained the classification model (with hyper-parameters 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 subjects 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$_2$: As peak VO$_2$ is one of the key parameters extracted from 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 to the true measured peak VO$_2$ from corresponding CPX procedure, in a correlation and a Bland-Altman analysis. We have calculated the percentage error (%Error) between estimated and true values of peak VO$_2$ and reported the average of the %Error. We have used values from all 68 CPX instances, including both training-testing and 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 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 the stage C and D were compared using Student *t* test. In this work, *p*-values below 0.05 were considered statistically significant.

Results

Subject demographics and clinical characteristics are detailed in Table 1 and CPX characteristics are provided in Table 2. Survival analysis using subsequent events (left ventricular assisted device implantation, heart transplant, or cardiovascular death) occurring over six months following the initial collection of data is provided in the supplementary materials.

A. Regression Model Comparison

Figure 3 (a) 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 whereas Figure 4 (a) 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 to the model using ECG features only: RMSE of 2.55±1.16 ml/kg/min vs. 3.75±1.68

ml/kg/min respectively ($p < 0.001$) and corresponding R^2 of 0.63 vs. 0.19. Combining SCG and ECG features improved the estimation accuracy slightly compared to SCG features only but the improvement was not significant ($p > 0.05$) with an RMSE of 2.50±1.12 ml/kg/min and 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/min vs. 3.52±1.5 ml/kg/min respectively ($p < 0.001$) and corresponding R^2 of 0.76 vs. 0.36. Similarly combining SCG and ECG features improved the estimation accuracy (RMSE of 2.28±0.93 ml/kg/min and R^2 of 0.76) slightly compared to SCG features only, though the improvement was not significant ($p > 0.05$).

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 to 0.31 using HR only for the training-testing set ($p < 0.05$), and correspondingly 0.76 compared to 0.25 using HR only in the validation set ($p < 0.05$). The corresponding RMSE values were 2.55±1.16 (SCG) versus 3.58±1.54 ml/kg/min (HR) for the training-testing set, and 2.28±1.04 (SCG) versus 3.66±1.74 ml/kg/min (HR) for the validation set.

B. Classification

Table 3 and Table 4 show the classification results using the SVM with an RBF kernel for the training-testing and validation set 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. Figure 3(b) and 4(b) shows the receiver operating characteristics (ROC) curve of the classifier with an area under the curve (AUC) of 0.74 and 0.92 for the training-testing and validation set respectively.

C. Peak VO_2 Estimation

Figure 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 %Error of 20.74% and an R^2 of 0.5.

D. Peak HR-based Regression and Classification

The correlation analysis between peak VO_2 and peak HR resulted in an R^2 of 0.23 for all 68 CPX instances. On the other hand, estimation of peak VO_2 using 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 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/min in this case. In the case of classifying the patients based on peak HR alone into stage C and D HF, the resultant AUC values for the ROC 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 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.

16

An important finding in this work was that the features from the SCG signal were more salient in estimating VO_2 as compared to the ECG signal. Many "Holter" type patches are currently available for ECG measurement, and have been used in studies for monitoring HF patients (22,23). Additionally, smartwatches are commercially available and can measure HR and possibly HR variability (provided there is minimal motion artifact). While such commercially available tools are convenient and readily applicable to studies in patients with HF, the results from this paper demonstrate that HR based features may not provide sufficient value in assessing cardiopulmonary health in patients with HF during exercise. Rather, approaches employing 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 six-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 two 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 MEMS accelerometer available, with a noise floor that is 2.5x lower than any other MEMS 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

17

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 is a well-known limitation of machine learning-based models as it will try to produce results close to the overall mean of the distribution rather than extreme values. Increasing the number of subjects with a broader spectrum of exercise capabilities may reduce 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 contributed to the comparatively lower performance of peak VO_2 estimation in our analysis compared to the estimation of the beat-by-beat estimation of VO_2.

While the measurement of VO_2 values less than peak may not currently be clinically relevant, one can imagine that with the capability of estimating VO_2 accurately for sub-maximal exercise tasks, such as walking upstairs or outdoors, the ability to assess patients with heart failure 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 sub-maximal 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 cardio-mechanical 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 stage C and D patients

as well. Future work should focus on improving the estimation accuracy of peak VO_2 from wearable SCG and ECG signals.

Peak VO_2 was used along with 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 sub-maximal exercise levels as well as maximal effort, whereas classification tasks can give one label to the whole CPX test. These preliminary findings, however, need verification in a larger patient population with a variety of exercise levels. As 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 sub-maximal exercise to relate to the severity of HF and investigate the underlying physiological relationship between them.

It should also be noted that, while 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 versus ECG-HR features does provide some insight into possible physiological 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 non-linearity (i.e., 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 (e.g., 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. As our data set had only 21% stage D subjects (25% in the training-testing and 13% in the validation set), resulting in higher peak VO_2 for stage D subjects. For a few cases of stage C subjects with a very high peak VO_2 compared to 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 subjects and incorporate subjects with a broader spectrum of exercise capabilities, which may reduce the estimation error for these extreme cases. Similarly, our classification model classified 30 out of 33 stage C CPX instances accurately whereas 7 out of 11 stage D CPX instances were accurately classified in the training-testing set. For the validation set, it classified 18 out of 21 stage C CPX instances accurately whereas 2 out of 3 stage D CPX instances were accurately classified. The comparatively poor performance in the classification of stage D subjects can be associated with a smaller number of stage D subjects (n=14) in our data set, shorter duration of exercise compared to stage C subjects, and larger pathophysiological differences among subjects due to various HF related diseases. Increasing the number of stage D subjects in future studies should increase the classification accuracy for stage D subjects 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 less stage D subjects by chance compared to the training-testing set, and our model performed well for the stage C patients as it has more stage C patients to learn from in the training phase. Incorporating more stage D patients 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, e.g., VCO_2, VE, tidal volume etc. 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 (HFpEF). 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 to 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.

<u>Conclusion</u>

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 / hospital settings such that treatment and management can be titrated and personalized based on physiological state. Since 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 heart failure patients 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 stratification of patients with heart failure based on the severity of their disease. Future work will investigate tracking physiological changes and responses to interventions during daily activities at home in this patient population.

24

References:
1. Malhotra R, Bakken K, D'Elia E, Lewis GD. Cardiopulmonary exercise testing in heart failure. JACC: Heart Failure 2016;4:607-616.
2. Myers J, Arena R, Cahalin LP, Labate V, Guazzi M. Cardiopulmonary exercise testing in heart failure. Current problems in cardiology 2015;40:322-372.
3. Balady GJ, Arena R, Sietsema K 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. Journal of Applied Physiology 2015;119:739-744.
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 transactions on biomedical circuits and systems 2016;10:280-288.
6. Inan Omer T, Baran Pouyan M, Javaid Abdul Q et al. Novel Wearable Seismocardiography and Machine Learning Algorithms Can Assess Clinical Status of Heart Failure Patients. Circulation: Heart Failure 2018;11:e004313.
7. Inan OT, Migeotte P-F, Park K-S et al. Ballistocardiography and seismocardiography: A review of recent advances. IEEE journal of biomedical and health informatics 2015;19:1414-1427.
8. Inan OT, Dorier A, Dowling S et al. Activity-contextualized wearable ballistocardiogram measurements can classify decompensated versus compensated heart failure patients. Am Heart Assoc, 2016.
9. Inan OT, Javaid AQ, Dowling S et al. Using ballistocardiography to monitor left ventricular function in heart failure patients. Journal of Cardiac Failure 2016;22:S45.
10. Guazzi M, Adams V, Conraads V et al. EACPR/AHA Scientific Statement. Clinical recommendations for cardiopulmonary exercise testing data assessment in specific patient populations. Circulation 2012;126:2261-2274.
11. Medicine ACoS. ACSM's guidelines for exercise testing and prescription: 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-e711.
13. Mehra MR, Canter CE, Hannan MM et al. The 2016 International Society for Heart Lung Transplantation listing criteria for heart transplantation: a 10-year update. The Journal of Heart and Lung Transplantation 2016;35:1-23.
14. Sörnmo L, Laguna P. Bioelectrical signal processing in cardiac and neurological applications: Academic Press, 2005.
15. Mahalanobis PC. On the generalized distance in statistics. National Institute of Science of India, 1936.
16. Breiman L. Random forests. Machine learning 2001;45:5-32.
17. Stone M. Cross-Validatory Choice and Assessment of Statistical Predictions. Journal of the Royal Statistical Society: Series B (Methodological) 1974;36:111-133.
18. Bot S, Hollander A. The relationship between heart rate and oxygen uptake during non-steady state exercise. Ergonomics 2000;43:1578-1592.

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 learning 1995;20:273-297.
22. Steinhubl SR, Waalen J, Edwards AM 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;320:146-155.
23. Levine DM, Ouchi K, Blanchfield B et al. Hospital-level care at home for acutely ill adults: A pilot randomized controlled trial. Journal of general internal medicine 2018;33:729-736.

Figure Titles and Legends:

Figure 1. Experimental Setup with Wearable Patch Hardware and Representative Cardiogenic Signals
(a) The experimental setup with the subject walking on a treadmill, with all the CPX measurement sensors and wearable patch attached to the body. (b) The wearable patch top and bottom view with snaps for ECG electrodes and internal hardware. (c) Representative ECG and tri-axial SCG signals (head-to-foot, HtoF, dorso-ventral, DV, and lateral, LAT) from one patient in the study.

Figure 2. Overview of the Regression and Classification Techniques
(a) Wearable ECG and SCG (only showing one 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. 10 heartbeat frames from the SCG signals were averaged to get ensemble averaged heartbeats corresponds to one 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 to 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 RBF kernel to estimate the clinical state of a subject and it was compared to the actual clinical state derived from CPX.

Figure 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 SVM classifier with RBF 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.

Figure 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 RBF 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.

Figure 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.

	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	0.53
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	0.59
Weight, kg	87.99±18.39	87.57±17.96	89.59±20.63	0.68
BMI, kg/m^2	29.53±5.26	29.27±4.85	30.51±6.73	0.37
Ejection fraction,%	27.25±10.64	26.21±9.29	31.29±14.46	0.13
NYHA 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%)	0.73

Table 1. Subject Demographics and Characteristics

Bilateral leg edema	12 (20%)	8 (18%)	4 (27%)	0.23
Systolic blood pressure, mmHg	105±15	105±14	102±19	0.41
Diastolic blood pressure, mmHg	68±10	68±9	68±15	0.85
BNP, pg/mL	568.4±722.5 (23*)	368±514 (17*)	1136.3±962.1 (6*)	0.02
NT-proBNP, pg/mL	1635.3±1671.2 (31*)	1783.4±1687.7 (25*)	1018.5±1587.5 (6*)	0.35
Serum Creatinine, mg/dL	1.40±1.43 (60*)	1.49±1.61 (46*)	1.13±0.43 (14*)	0.38
Loop Diuretics, Furosemide, mg/d	83.7±93.4 (68%)	64±71 (65%)	146.4±128.1 (79%)	0.01
B-blockers, Bisoprolol, mg/d	6.1±3.8 (94%)	5.9±3.9 (93%)	6.7±3.7 (100%)	0.54
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%)	0.28
ARNI, Sacubitril-Valsartan, mg/d	102.4±64.2 (58%)	101.2±64.8 (61%)	107.7±65.5 (50%)	0.91
MRA, Spironolactone, mg/d	29.8±16.7 (85%)	29.3±15.5 (81%)	31.6±20.7 (100%)	0.64
Subsequent Events (OHT/VAD/Death)[a]	11(16%)	7 (13%)	4 (29%)	0.16

Values shown are mean ± SD or n (% of population) or mean ± SD (% of population) unless otherwise indicated. Statistical significance between stage C and D subjects in values, where applicable, was evaluated using an unpaired t test or a chi-square test.; NYHA, New York Heart Association; BNP, b-type natriuretic peptide; NT-proBNP, N-terminal pro b-type natriuretic peptide; pg/mL, picogram per milliliter; mg/dL, milligram per deciliter; ACE, angiotensin converting enzyme; ARB, angiotensin receptor blocker; ARNI, angiotensin receptor blocker – neprilysin inhibitor; MRA, mineralocorticoid receptor antagonist; mg/d, milligram per day; OHT, orthotopic heart transplantation; VAD: ventricular assisted device implantation.

*Number of CPX test instances with available lab results.

[a]Subsequent events were recorded up to 6-months after the completion of the study. In the cases where one CPX subject had multiple events (e.g. VAD, followed by transplant later), only the first occurring event was counted as subsequent events for a particular subject.

29

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/VCO2 slope	33.35±6.65	32.44±6.48	36.82±6.34	0.04
VO$_2$ at AT, 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^{-1}	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, s	672±235	743±200	401±148	<0.001
Peak heart rate, beats.min^{-1}	120.06±23.8	124.57±22.79	102.64±19.77	0.002

Values shown are mean±SD. Statistical significance between stage C and stage D subjects in values, where applicable, was evaluated using an unpaired t test
*Number of CPX instances with detectable AT points, Modified V-slope method was used to detect the AT points.

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	

TP = True Positive, FN = False Negative, FP = False Positive, TN= True Negative
Accuracy = 0.84, Sensitivity = 0.91, Specificity = 0.64, Positive predictive value = 0.88 and 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 and Negative predictive value = 0.4

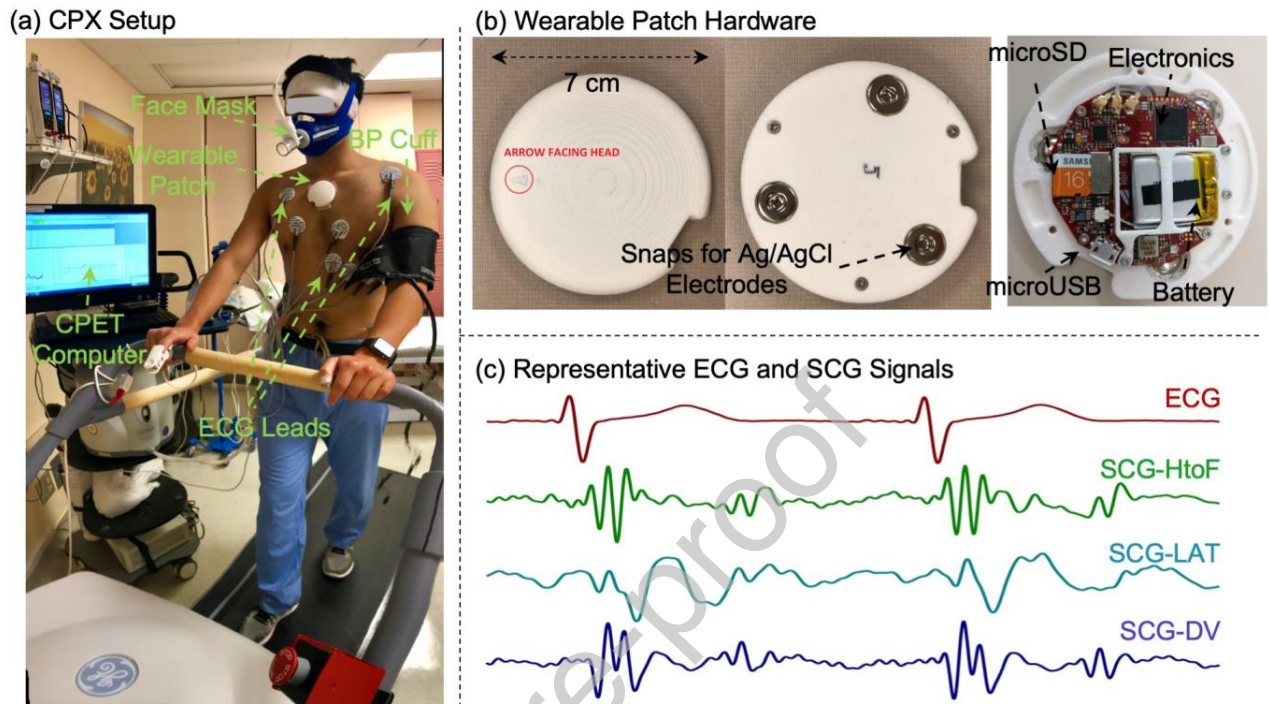**Figure 1: Experimental Setup with Wearable Patch Hardware and Representative Cardiogenic Signals**

Figure 1. (a) The experimental setup with the subject walking on a treadmill, with all the CPX measurement sensors and wearable patch attached to the body. (b) The wearable patch top and bottom view with snaps for ECG electrodes and internal hardware. (c) Representative ECG and tri-axial SCG signals (head-to-foot, HtoF, dorso-ventral, DV, and lateral, LAT) from one patient in the study.

Figure 2: Overview of the Regression and Classification Techniques
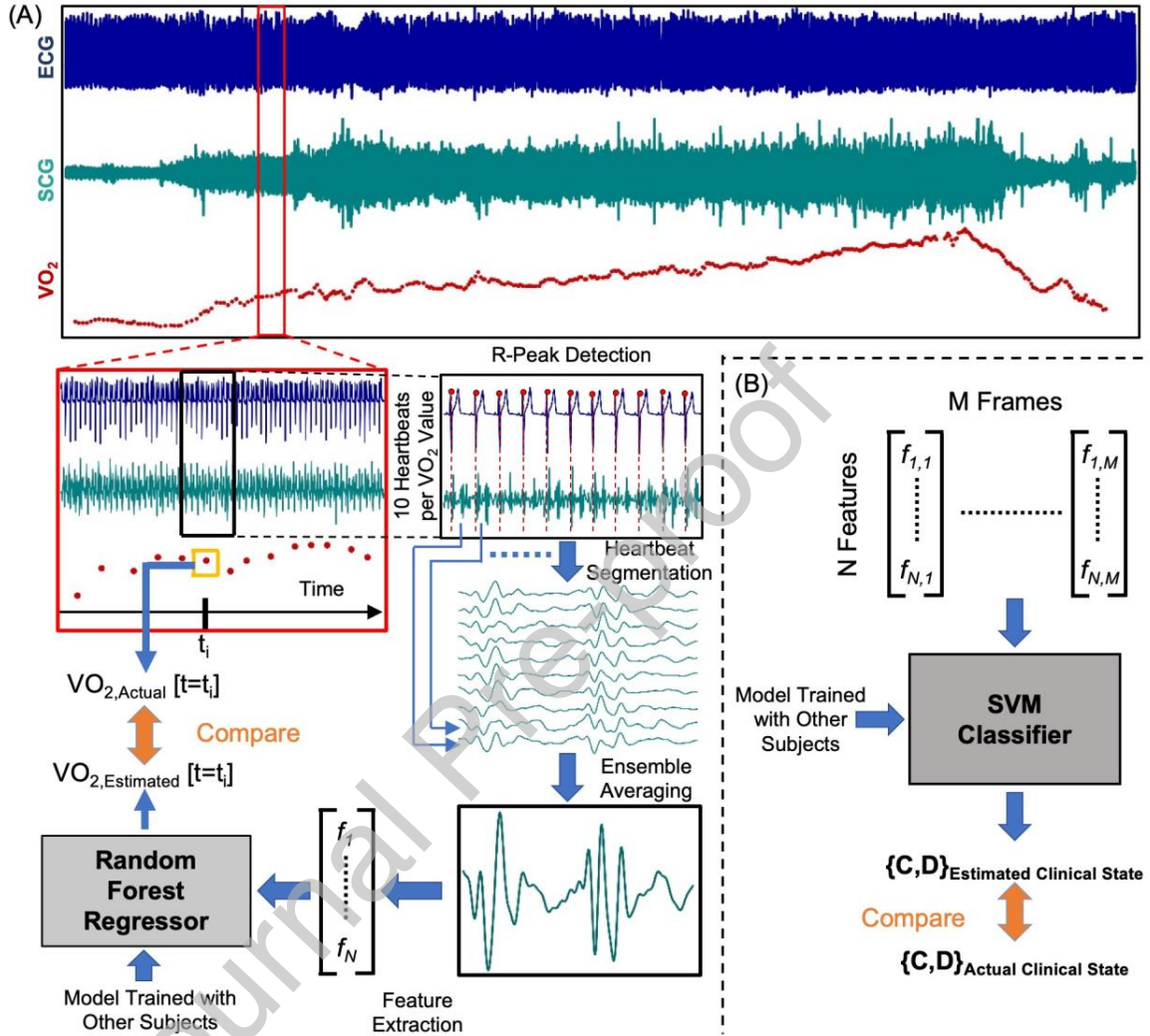
Figure 2. (a) Wearable ECG and SCG (only showing one axis of the signal for simplicity) signals were synchronized with breath-by-breath data from CPX computer. R-peaks of the ECG signal were detected and the SCG signals were segmented into heartbeats using corresponding R-peaks. 10 heartbeat frames from the SCG signals were averaged to get ensemble averaged heartbeats corresponds to one 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 to 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 RBF kernel to estimate the clinical state of a subject and it was compared to the actual clinical state derived from CPX.
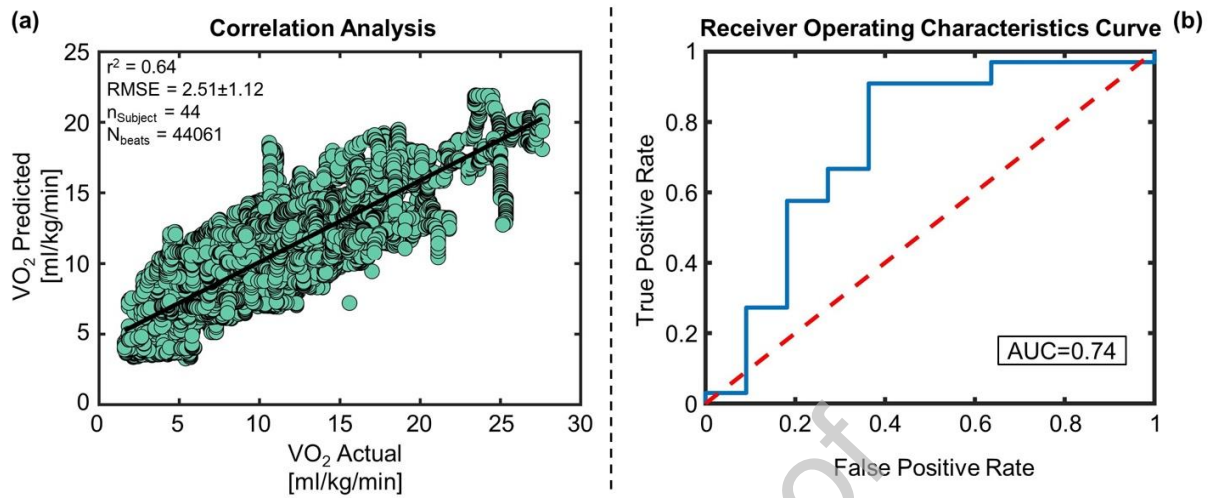
Figure 3: Regression and Classification Results on the Training-Testing Set

Figure 3. (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 SVM classifier with RBF 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.
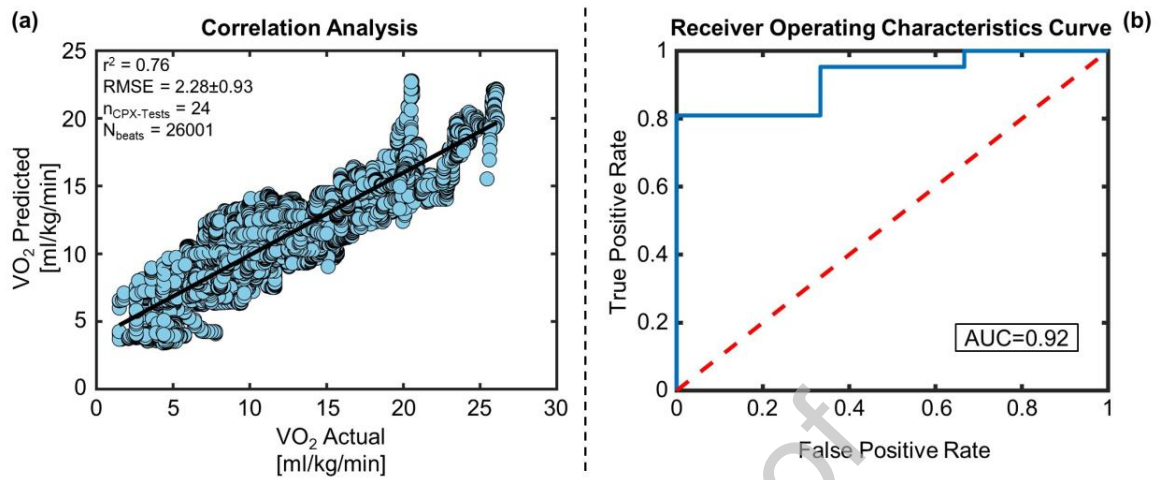
Figure 4: Regression and Classification Results on the Validation Set

Figure 4. (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 RBF 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.
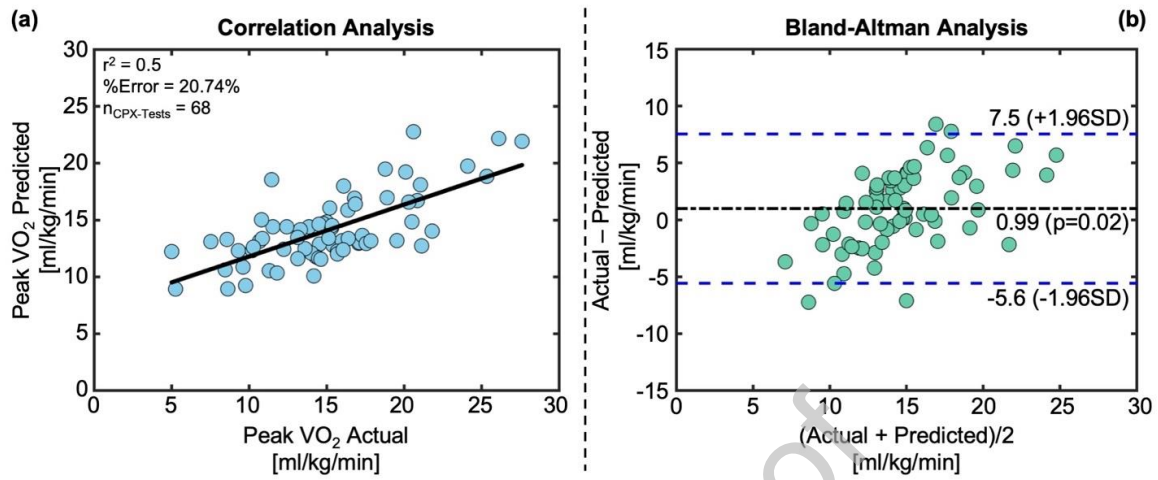
Figure 5: Results of Peak VO$_2$ Estimation

Figure 5. (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.