



Published in final edited form as:

Physiol Meas. ; 41(4): 044001. doi:10.1088/1361-6579/ab7f93.

Detecting Heart Failure Using Wearables: A Pilot Study

Amit J. Shah^{1,2,3},

Nino Isakadze⁴,

Oleksiy Levantsevych²,

Adriana Vest^{1,5},

Gari Clifford^{5,6},

Shamim Nemati⁷

¹Dept of Medicine, Emory University School of Medicine, Atlanta, GA 30322.

²Dept of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA 30322.

³Atlanta Veterans Affairs Medical Center, Decatur, GA 30033.

⁴Dept of Medicine, Johns Hopkins University, Baltimore, MD 21218.

⁵Dept of Biomedical Informatics, Emory University, Atlanta, GA 30322

⁶Dept of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA 30332

⁷Dept of Biomedical Informatics, University of California San Diego Health, San Diego, CA 92093

Abstract

Objective: Heart failure (HF) can be difficult to diagnose by physical examination alone. We examined whether wristband technologies may facilitate more accurate bedside testing.

Approach: We studied on a cohort of 97 monitored in-patients and performed a cross-sectional analysis to predict HF with data from the wearable and other clinically available data. We recorded photoplethysmography (PPG) and accelerometry data using the wearable at 128 samples per second for 5 minutes. HF diagnosis was ascertained via chart review. We extracted four features of beat-to-beat variability and signal quality, and used them as inputs to a machine learning classification algorithm.

Main Results: The median [interquartile] age was 60 [51 68] years, 65% were men, and 54% had heart failure; in addition, 30% had acutely decompensated HF. The best 10-fold cross-validated testing performance for the diagnosis of HF was achieved using a support vector machine. The waveform-based features alone achieved a pooled test area under the curve (AUC) of 0.80; when a high-sensitivity cut-point (90%) was chosen, the specificity was 50%. When adding demographics, medical history, and vital signs, the AUC improved to 0.87, and specificity improved to 72% (90% sensitivity).

Corresponding author: Amit Shah, MD, MSCR, Assistant Professor, Department of Epidemiology, Assistant Professor, Department of Medicine, Division of Cardiology, Staff Physician, Division of Cardiology, Atlanta VA Medical Center, Emory University, 1518 Clifton Road NE, Room 3053, Atlanta, GA 30322, (T) 404-727-8712, (C) 404-647-4351, (F) 404-727-8737, ajshah3@emory.edu.

Significance: In a cohort of monitored in-patients, we were able to build an HF classifier from data gathered on a wristband wearable. To our knowledge, this is the first study to demonstrate an algorithm using wristband technology to classify HF patients. This supports the use of such a device as an adjunct tool in bedside diagnostic evaluation and risk stratification.

Keywords

heart failure; mobile health technology; biomedical informatics

Introduction

Heart failure (HF) is broadly defined as a complex clinical syndrome in which the heart's ability to fill or eject blood is impaired. There are an estimated 38 million people with HF worldwide [1], and the affected population is growing steadily. Although in many cases HF can cause symptoms such as shortness of breath and leg swelling, other times the symptoms can be nonspecific and difficult to diagnose without an echocardiogram, which is expensive and can have limited availability (for example, in primary care clinics and developing countries). Other conditions such as anemia, depression, emphysema, and chronic venous insufficiency can have similar clinical presentations to HF. Therefore, the need to have accurate bedside diagnostic tools is critically important when diagnosing and treating patients with symptoms suggestive of HF. Physical examination is helpful, but can vary by provider and offer limited sensitivity [2].

Wearables have the potential to measure physiology in a way that is potentially more accurate, objective, and repeatable than the clinical exam in diagnosing HF. A photoplethysmography (PPG) signal provides a wealth of information, including heart rate and rhythm, vascular dynamics, and oxygenation [3]. One of the most notable differences between HF and controls is impaired autonomic function and increased propensity for arrhythmia. Generally, HF patients have less heart rate variability, higher overall heart rates, and increased arrhythmia risk [4]. In this study, we seek to examine the ability of heart rate variability and entropy (as an indirect arrhythmia measure) derived from a wristband wearable with PPG sensor to classify HF in a pilot cohort of inpatient subjects. We hypothesize that such metrics, with or without clinical characteristics, can aid in the classification of HF.

Methods

Description of cohort:

Subjects were recruited randomly from an inpatient sample undergoing bedside ECG monitoring at Emory University Hospital, Emory Midtown Hospital, and Grady Memorial Hospital from October 2015 until March 2016. Most of these subjects were admitted as inpatients for coronary artery disease, heart failure, or arrhythmia management, while a minority of patients were admitted for non-cardiac reasons, including pulmonary disease and infection. Because an additional aim of the study was AF detection, those with AF were oversampled with approximately 50% prevalence [5]. Otherwise, recruitment was blind to HF status and comorbidities. ECG rhythms were reviewed by 3 physicians (A.S., O.L., and

N.I.) to adjudicate AF vs. non-AF with 12-lead ECG's that were scanned in the medical records. The study was IRB approved and all subjects underwent informed consent.

Description of data collection device and acquisition of data:

The Samsung Simband was used for capture of reflective PPG and 3-axis accelerometry at 128 Hz each. Subjects were instructed to sit still, and they wore the device for 5 minutes while resting in a chair. Approximately 3 minutes into the monitoring period, they were asked to do 5 Valsalva maneuvers over 1 minute. This was done in order to evaluate the baroreflex and other pressure-related mechanisms of the heart.

Description of HF adjudication:

Chronic HF was based on physician-adjudicated chart review by a single individual (N.I.). HF was classified based on whether or not the patient had any history of heart failure with reduced (HFrEF) or preserved ejection fraction (HFpEF). This was based on clinical notes, echocardiographic data, and cardiac catheterization data. If ejection fraction was found to be less than or equal to 40% on echocardiogram or left ventriculogram, then the HF was classified as HFrEF. Relevant clinical history included previous admission of acutely decompensated heart failure and/or current symptoms of heart failure (e.g. dyspnea and orthopnea) with corresponding physical examination (rales, increased jugular venous pressure, and edema) and elevated brain natriuretic peptide levels.

In addition to HF, a detailed chart review was performed to ascertain clinical characteristics that could be risk factors with HF such as age, diabetes, hypertension, coronary artery disease, and others. ECG at the time of examination was evaluated and adjudicated for AF independently by 2 clinicians (N.I. and A.S.). Disagreements were discussed and a consensus was obtained in each applicable case.

Description of Data Analysis

Waveform Preprocessing: We first performed outlier rejection and amplitude normalization. We then standardized the PPG signal data to have a mean amplitude of zero, and cropped the 10% highest amplitude waveforms, including below the lower 5% (most negative waveforms) and above the upper 95%. We then performed bandpass filtering to remove frequencies outside of the range 0.2 – 10 Hz with a FIR filter of order 41. PPG pulse onset detection was performed using the slope sum function approach (see Fig. 2), and the signal quality index was assessed using the Hjorth's purity quality metric (purity-SQI: 0=random noise, 1=sinusoidal signal) [6, 7]. Additionally, accelerometer data was evaluated as part of the signal quality assessment [8].

Feature measurement: The entire 5-minute interval was analyzed. Beat-to-beat intervals were assessed from the PPG signal, which allowed for the calculation of heart rate variability (HRV). Such metrics have been found to associate strongly with heart failure due to generalized sympathetic activation [9]. We then calculated several standard HRV metrics, including standard deviation of the RR intervals (RR-rSTD), as well as sample entropy (SampEn). PPG signal quality index (Purity-SQI) and standard deviation of accelerometry (ACC-STD) were added feature of the model as well, given that a significant amount of data

was noisy. RR-rSTD is considered as an important variable for two reasons; in heart failure, the value is reduced, while in AF, it is elevated. Because AF and HF are comorbid, both low and high extremes in RR-rSTD are important to consider for the purposes of classifying HF.

Technical study design: Our goal was to build an accurate HF classifier, and to assess the contribution of wearable-based features as well as clinical variables to the overall classification performance. Therefore, we considered four groups of variables and constructed four separate models, based on wearable-based features (Model 1: Purity-SQI, SampEn, RR-rSTD, ACC-STD), socio-demographics (Model 2: age, race, and gender), medical history (Model 3: COPD, hypertension, PVD, creatinine), and 4) physical exam (Model 4: BMI, SBP, DBP, HR, SPO₂). Additionally, in Model 5, we looked at the combined predictive performance of all clinical variables from Models 2, 3, and 4. Finally, we looked at the combined performance of the wearable-based features and the clinical covariates (Model 6) to evaluate the maximum accuracy when assessing all data that may be available at the bedside.

Algorithm Development: We designed six linear support vector machines (SVMs) corresponding to each of the six models using MATLAB [10] Statistics and Machine Learning Toolbox™. This was chosen because of the small sample size [11]. Further feature selection was performed using Bayesian Optimization technique for global optimization of model hyper-parameters [12].

Statistical Methods: For all continuous variables, we report medians (25-percentile, 75-percentile) and utilize a two-sided Wilcoxon ranksum test when comparing two samples. For binary features we report percentages and utilize a two-sided Chi-square test to assess differences in proportions between two samples. All classification results are based on 10-fold bootstrapping cross-validation; for each of the 10 iterations [13], we selected a random 70% of the data for training, and 30% for testing. We then report the mean (of the 10 iterations) area under receiver operating curves (AUROC) for the main outcome. In addition, we report the specificity, accuracy, and positive predictive value (PPV) at a fixed 90% sensitivity level. This was chosen with the likely clinical application aimed at ruling out HF, rather than ruling it in. We combined all the testing set predictions (probabilities of HF) across all ten folds to calculate a single pooled AUC [14].

A sensitivity analysis was also performed to examine for possible differences in model performance amongst various subgroups. For this purpose, a single model was developed on the entire sample to develop a score based on the sum of the products of the beta coefficient and predictors. The score was then evaluated based on its ability to detect heart failure with reduced ejection fraction vs. preserved ejection fraction. Furthermore, it was tested amongst those with and without AF.

Results

Baseline Characteristics:

We evaluated 97 subjects for this study, including 54 HF patients (56%). Also, 29 (30%) were admitted for acutely decompensated HF; the other reasons for admission were

heterogenous, including infection, pulmonary disease, and ischemic heart disease. The median ([25-percentile, 75-percentile]) age of the cohort was 60 [51 68] years; 65% were men, and 76% were African American. Approximately 48% of the subject had a history of AF. The HF patients had a higher prevalence of AF (25.6% vs. 66.7%; $p<0.01$) and COPD (7.0% vs. 37.0%; $p<0.01$). Moreover, both creatinine level (1.0 [0.8 1.2] vs. 1.32 [1.0 2.2]; $p<0.01$) and heart rate (70 [63 84] vs. 84 [71 94]; $p<0.01$) were elevated in the HF patients (see Table 1 for the baseline characteristics of the patient population). When comparing HFrEF and HFpEF, no significant differences were found between groups.

Table 2 lists the features in order of importance, and table 3 provides a summary of the performance of all six models. Notably, the wearable-based features alone (Model 1) achieved an AUC of 0.80, with an overall accuracy of 74%. The final features included in this model, after variable selection, included the Purity-SQI, SampEn, RR-rSTD, ACC-STD (4 features total). In comparison, all clinical variables combined (Model 5) yielded a testing AUC of 0.81 with an overall accuracy of 0.75. After variable selection, our combined model (Model 6) included 14 variables, including Purity-SQI, SampEn, RR-rSTD, ACC-STD, age, race, history of AF, chronic obstructive pulmonary disease, hypertension, peripheral vascular disease, serum creatinine, systolic blood pressure, diastolic blood pressure, and heart rate. This model achieved a testing AUC of 0.87, and when choosing a cut point with a sensitivity of 90%, the specificity was 72%, and accuracy was 82% (see Figure 3).

A sensitivity analysis was then performed to evaluate the model performance amongst different groups of heart failure. Heart failure with reduced ejection fraction (HFrEF) was present in 33% of subjects, while heart failure with preserved ejection fraction (HFpEF) was present in 23% of subjects. The AUC for the model in HFrEF (vs. no HF) was 0.92, and for HFpEF (vs. no HF) was 0.85. When the difference between AUCs was evaluated (using SAS 9.4), no significant difference was found, with $p=0.42$. Acutely decompensated HF (AUC 0.85) was discriminated from controls to a greater degree than chronic stable HF (AUC 0.77), although this difference was not statistically significant ($p=0.29$). In addition, a trend for higher AUC was found in those without AF (0.79) vs. AF (0.66), although the difference was not statistically significant ($p=0.44$).

Discussion

To our knowledge, this is the first study that has attempted to classify individuals with HF from physiologic data recorded from a wristband monitor. Our findings, which include an accuracy of 74% from the wristband alone, suggest that such a device may be useful in aiding in the diagnosis of HF at the bedside, especially when combined with socio-demographics, clinical history, and physical examination: when combined with wristband data, accuracy increased to 82%. These findings are of particular interest when considering the work-up of dyspnea in an outpatient or emergency room setting, when limited time and resources are available for a full echocardiogram. In many cases, a chart review (or medical history) and wristband wearable may be sufficient to utilize the algorithm and risk stratify for HF. The actual use of this technology in diagnosing and treating HF prospectively, however, is unknown.

Our algorithm relies upon well-known characteristics of HF; namely, because of decreased autonomic flexibility and increased sympathovagal balance, HRV and complexity are generally suppressed in HF, and therefore the predictive value of the metrics from the wristband are physiologically sound. Addition of clinical variables enriches the algorithm by providing more context, and can be assessed at the bedside via clinical examination. A potential future clinical application therefore is a bedside tool consisting of a wristband monitor and tablet in which the clinical features can be entered into the model.

Wristband technologies with PPG sensing have been increasingly shown to offer important physiologic data. This includes approximations of cardiac output and arterial stiffness, for example [15, 16]. As such, we consider these findings an important proof of concept that may be further enhanced in the future with additional PPG-based features that reflect oxygen saturation and cardiac output. Other portable technologies to risk stratify for HF and predict readmission may serve as an important predicate when evaluating this technology in various clinical scenarios. This includes CardioMEMS [17], a wireless pulmonary artery pressure monitor; ReDS [18], a system for non-invasively measuring pulmonary congestion; smartphone applications [19], which correlate with HF severity by counting physical activity; impedance devices, which measure chest congestion, and others [20]. The technology described here is more portable and low burden than most other technologies discussed, although it remains to be seen how the accuracy may compare.

Our study is subject to limitations. The small sample size did not allow us to derive separate models for subcategories of HF based on ejection fraction or ischemic heart disease; furthermore, the role of age, sex, and race could not be evaluated. Conditions that mimic HF such as anemia and depression were not fully evaluated when considering model accuracy. AF was heavily oversampled in the cohort, and therefore the results should be evaluated with caution in populations with lower AF prevalence, given the association of AF with HF. The clinical context for the participants was heterogeneous, ranging from HF to infection; this may limit generalizability to particular groups. Data on precise volume status was not available; however, based on chart review we were able to at least ascertain which participants were admitted for acutely decompensated HF. When applying the model to a clinical setting in which a 90% sensitivity is required (because of the costs of missing a potentially life-threatening diagnosis), the algorithm from the wristband alone yields only a 50% specificity; as such, it should only be applied to high pre-test probability settings and combined with clinical data to minimize the number of false positives. Alternatively, it may be considered as a screening tool in home settings or other screening programs.

In conclusion, we present an important proof-of-concept study that classifies chronic HF based on a limited number of features using PPG and accelerometry from a multisensor wristband. The results indicate that such a test may be a useful bedside diagnostic aid, although further testing is warranted to further develop and improve the performance of the algorithm in larger cohorts, as well as the study its potential clinical role in a prospective manner.

Acknowledgements:

Dr. Shah was funded by the NIH/NHLBI K23 HL127251, and NIH UL1 TR000454 (KL2 TR00045 scholarship). Dr. Nemati was funded by the NIH/NIEHS K01 ES025445. The Simband was provided on loan at no charge by Samsung electronics. The content is solely the responsibility of the authors and does not necessarily represent the official views of Samsung or the funding agencies.

References

1. Braunwald E, The war against heart failure: the Lancet lecture. *Lancet*, 2015. 385(9970): p. 812–24. [PubMed: 25467564]
2. Butman SM, et al. , Bedside cardiovascular examination in patients with severe chronic heart failure: importance of rest or inducible jugular venous distension. *J Am Coll Cardiol*, 1993. 22(4): p. 968–74. [PubMed: 8409071]
3. Kamal A, et al. , Skin photoplethysmography—a review. *Computer methods and programs in biomedicine*, 1989. 28(4): p. 257–269. [PubMed: 2649304]
4. Saul JP, et al. , Assessment of autonomic regulation in chronic congestive heart failure by heart rate spectral analysis. *The American journal of cardiology*, 1988. 61(15): p. 1292–1299. [PubMed: 3376889]
5. Shashikumar SP, et al. A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology. in *Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on*. 2017. IEEE.
6. Nemati S, Malhotra A, and Clifford GD, Data Fusion for Improved Respiration Rate Estimation. *EURASIP J Adv Signal Process*, 2010. 2010: p. 926305. [PubMed: 20806056]
7. Hjorth B, EEG analysis based on time domain properties. *Electroencephalogr Clin Neurophysiol*, 1970. 29(3): p. 306–10. [PubMed: 4195653]
8. Li Q and Clifford G, Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiological measurement*, 2012. 33(9): p. 1491. [PubMed: 22902950]
9. Floras JS, Sympathetic Nervous System Activation in Human Heart Failure. *Journal of the American College of Cardiology*, 2009. 54(5): p. 375. [PubMed: 19628111]
10. MATLAB and Statistics Toolbox 2017, The MathWorks, Inc.: Natick, Massachusetts, United States.
11. Cherkassky V and Ma Y, Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, 2004. 17(1): p. 113–126. [PubMed: 14690712]
12. Ghassemi M, et al. Global optimization approaches for parameter tuning in biomedical signal processing: A focus on multi-scale entropy. in *Computing in Cardiology Conference (CinC)*, 2014. 2014. IEEE.
13. Harrell FE Jr., Lee KL, and Mark DB, Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*, 1996. 15(4): p. 361–87. [PubMed: 8668867]
14. Airola A, et al. A comparison of AUC estimators in small-sample studies. in *Machine Learning in Systems Biology*. 2009.
15. Wu H-T, et al. , Arterial stiffness using radial arterial waveforms measured at the wrist as an indicator of diabetic control in the elderly. *IEEE Transactions on Biomedical engineering*, 2011. 58(2): p. 243–252. [PubMed: 20923726]
16. Wang L, et al. Noninvasive cardiac output estimation using a novel photoplethysmogram index. in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE 2009*. IEEE.
17. Givertz MM, et al. , Pulmonary Artery Pressure-Guided Management of Patients With Heart Failure and Reduced Ejection Fraction. *Journal of the American College of Cardiology*, 2017. 70(15): p. 1875–1886. [PubMed: 28982501]
18. Amir O, et al., Validation of remote dielectric sensing (ReDS™) technology for quantification of lung fluid status: Comparison to high resolution chest computed tomography in patients with and without acute heart failure. 2016. 221: p. 841–846.

19. Werhahn SM, et al. , Designing meaningful outcome parameters using mobile technology: a new mobile application for telemonitoring of patients with heart failure. *ESC Heart Failure*, 2019.
20. Walsh JA, Topol EJ, and Steinhubl SR, Novel Wireless Devices for Cardiac Monitoring. *Circulation*, 2014. 130(7): p. 573–581. [PubMed: 25114186]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

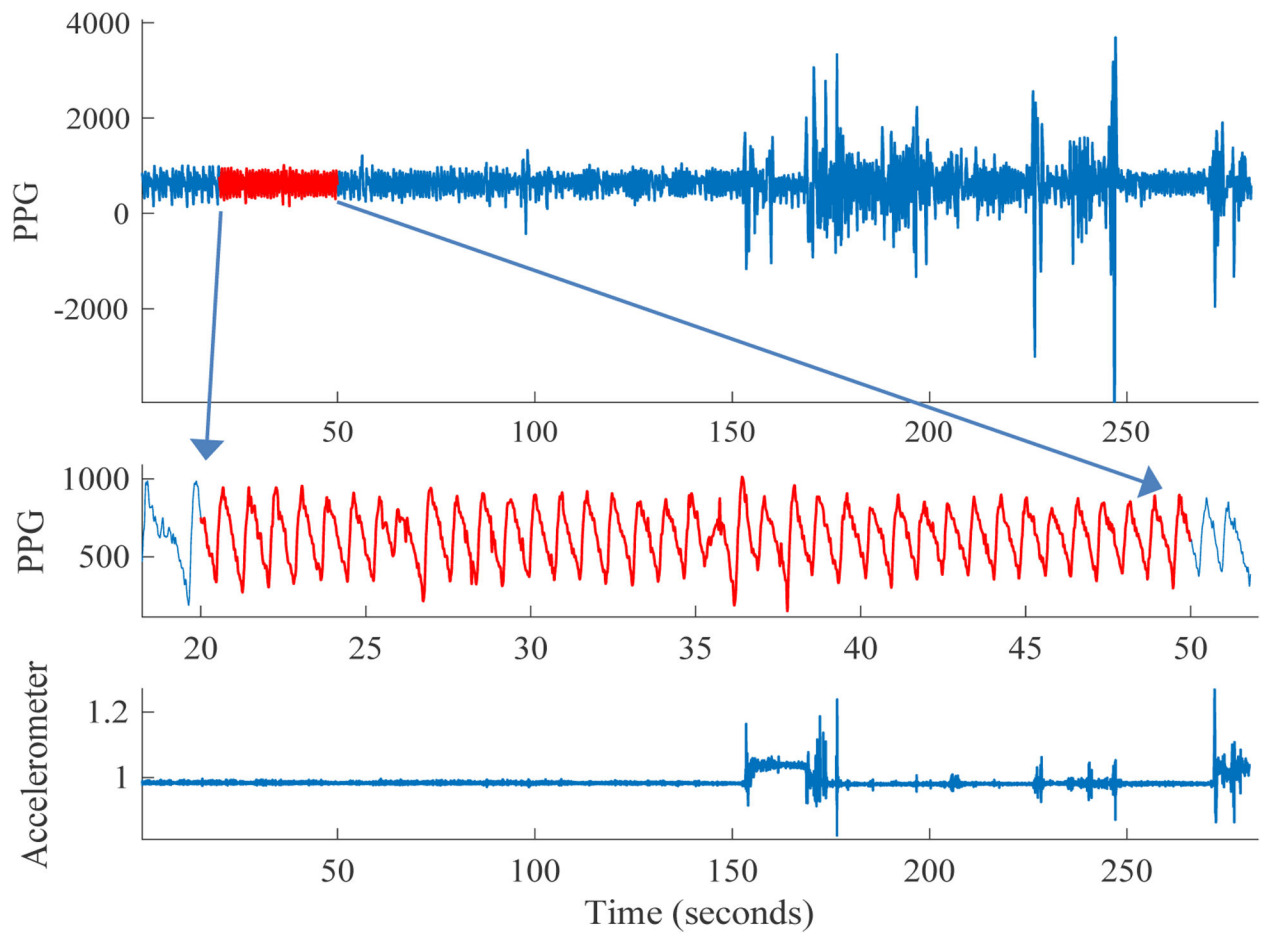


Figure 1.

Sample five minutes of PPG recording (top). We perform PPG and accelerometer-based signal quality assessment to select high signal quality (SQI) 30 second segments of the recording (middle). Movement, as reflected in the accelerometer (ACC) sensor magnitude (bottom) is often associated with degradation in PPG signal quality. Our ACC-based SQI metrics enables us to identify such high activity data segments.

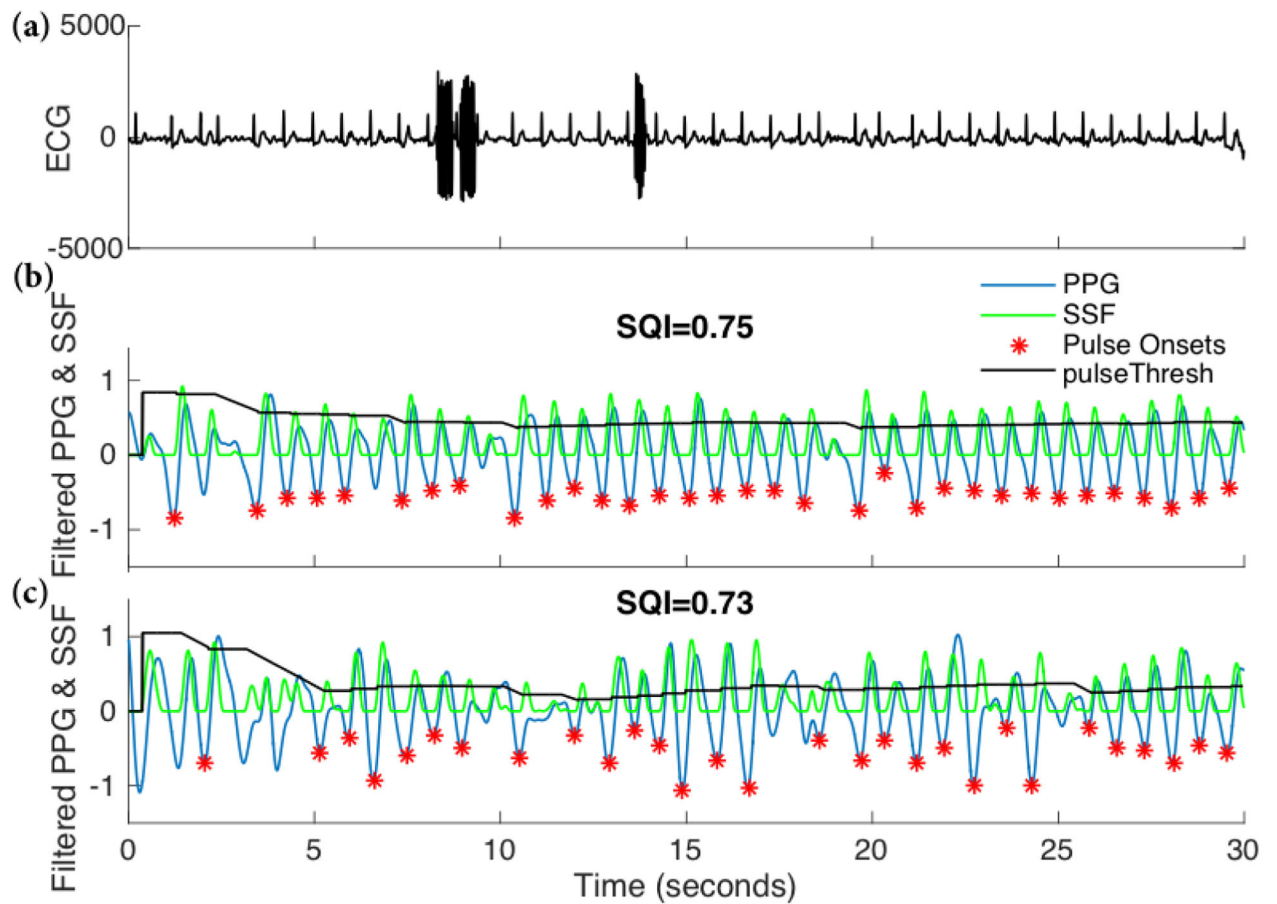


Figure 2.

Examples of recorded PPG signals (panels b and c) and the corresponding ECG recording (panel a). The PPG signal in the middle panel has a higher signal quality as quantified by the Purity-SQI (or simply SQI). The slope sum function (SSF) based pulse onset detection algorithm enables identification of each pulse (*) when the SSF peak is above the pulse threshold (pulseThresh).

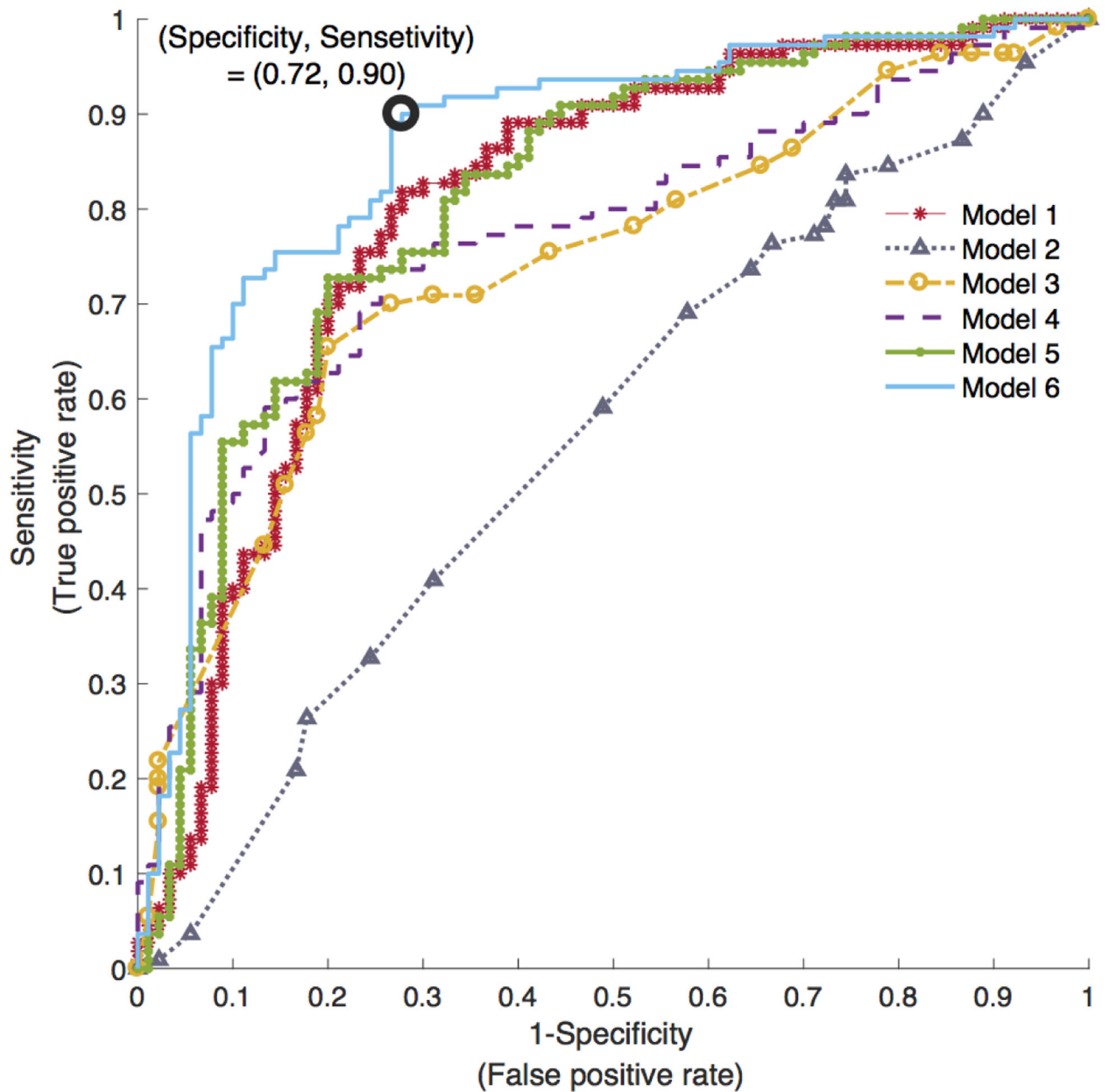


Figure 3.

Area under the receiver-operating characteristic (ROC) curve of all six HF classification models. Model 1: wearable features; Model 2: socio-demographics features; Model 3: patient's medical history features; Model 4: physical exams features; Model 5, combined clinical features from Models 2, 3, and 4; and Model 6: combined wearable and clinical features.

Table 1.

Baseline characteristics

	Non-HF	HF	p
Sample size	43	54	
Age	58 [51 68]	60 [52 68]	0.80
Race (White)	18.6%	20.4%	0.83
Race (African American)	79%	74%	0.56
Race (other)	2.4%	5.6%	0.4
Male	55.8%	72.2%	0.09
AF	25.6%	66.7%	<0.01
COPD	7.0%	37.0%	<0.01
Hypertension	70.0%	74.1%	0.64
PVD	2.3%	5.6%	0.42
Creatinine	1.0 [0.8 1.2]	1.32 [1.0 2.2]	<0.01
BMI	27.7 [24.4 33.2]	31.3 [24.6 38.7]	0.18
SBP	128 [110 136]	117 [108 127]	0.06
DBP	72 [66 80]	73 [66 82]	0.91
HR	70 [63 84]	84 [71 94]	<0.01
SpO ₂	98 [95 100]	98 [95 99]	0.27

For continuous variables, mean values are presented, followed by 95% confidence intervals in brackets.

Abbreviations: AF=atrial fibrillation, COPD=chronic obstructive pulmonary disease, PVD=peripheral vascular disease, BMI=body mass index, SBP=systolic blood pressure, DBP=diastolic blood pressure, HR=heart rate, SpO₂=blood oxygen saturation level

Table 2.

List of Features in Order of Importance

1. Photoplethysmogram Signal Quality Index (Purity-SQI)
2. Sample Entropy
3. Standard deviation of beat-to-beat intervals
4. Standard deviation of accelerometer amplitude
5. Heart rate
6. Systolic blood pressure
7. Creatinine
8. History of hypertension
9. Chronic Obstructive Pulmonary Disease
10. Peripheral Vascular Disease

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Performance summary of all six models

Model	AUC ROC Test (Train)	Specificity* Test (Train)	Accuracy* Test (Train)	PPV* Test (Train)
1	0.80 (0.80)	0.53 (0.50)	0.74 (0.72)	0.70 (0.69)
2	0.56 (0.59)	0.11 (0.18)	0.55 (0.58)	0.55 (0.58)
3	0.74 (0.70)	0.21 (0.22)	0.62 (0.60)	0.59 (0.59)
4	0.77 (0.76)	0.27 (0.27)	0.62 (0.62)	0.60 (0.61)
5	0.81 (0.88)	0.57 (0.71)	0.75 (0.82)	0.72 (0.80)
6	0.87 (0.90)	0.72 (0.81)	0.82 (0.86)	0.80 (0.86)

* Model sensitivity was fixed at 0.90

Model 1: Purity-SQI, SampEn, RR-rSTD, ACC-STD from wristband

Model 2: Age, Race, and Gender

Model 3: chronic obstructive pulmonary disease, hypertension, peripheral vascular disease, and creatinine,

Model 4: body mass index, systolic blood pressure, diastolic blood pressure, heart rate, and oxygen saturation

Model 5: Models 2, 3, and 4 combined

Model 6: Model 1 + 5 combined