Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

ACCEPTED MANUSCRIPT • OPEN ACCESS

Detecting beats in the photoplethysmogram: benchmarking open-source algorithms

To cite this article before publication: Peter H Charlton *et al* 2022 *Physiol. Meas.* in press https://doi.org/10.1088/1361-6579/ac826d

Detecting beats in the photoplethysmogram: benchmarking open-source algorithms

Peter H. Charlton, Kevin Kotzen, Elisa Mejía-Mejía, Philip J. Aston, Karthik Budidha, Jonathan Mant, Callum Pettit, Joachim Behar, *Member, IEEE*, Panicos A. Kyriacou, *Senior Member, IEEE*

Abstract

The photoplethysmogram (PPG) signal is widely used in pulse oximeters and smartwatches. A fundamental step in analysing the PPG is the detection of heartbeats. Several PPG beat detection algorithms have been proposed, although it is not clear which performs best. *Objective:* This study aimed to: (i) develop a framework with which to design and test PPG beat detectors; (ii) assess the performance of PPG beat detectors in different use cases; and (iii) investigate how their performance is affected by patient demographics and physiology. *Approach:* Fifteen beat detectors were assessed against electrocardiogram-derived heartbeats using data from eight datasets. Performance was assessed using the F_1 score, which combines sensitivity and positive predictive value. *Main results:* Eight beat detectors performed well in the absence of movement with F_1 scores of $\geq$ 90% on hospital data and wearable data collected at rest. Their performance was poorer during exercise with F_1 scores of 55-91%; poorer in neonates than adults with F_1 scores of 84-96% in neonates compared to 98-99% in adults; and poorer in atrial fibrillation (AF) with F_1 scores of 92-97% in AF compared to 99-100% in normal sinus rhythm. *Significance:* Two PPG beat detectors denoted 'MSPTD' and 'qppg' performed best, with complementary performance characteristics. This evidence can be used to inform the choice of PPG beat detector algorithm. The algorithms, datasets, and assessment framework are freely available.

Index Terms

atrial fibrillation, beat detection, electrocardiogram, heartbeat, photoplethysmography, pulse wave

Submitted for review on July 7, 2022. *(Corresponding author: Peter H. Charlton.)*

P. H. Charlton is with is with the Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK; and the Research Centre for Biomedical Engineering, City, University of London, London, EC1V 0HB, UK. (e-mail: pc657@medschl.cam.ac.uk).

K. Kotzen and J. Behar are with the Faculty of Biomedical Engineering, Technion-IIT, Israel. (e-mail: kkotzen@campus.technion.ac.il; jbehar@technion.ac.il).

E. Mejía-Mejía, K. Budidha, and P. A. Kyriacou are with the Research Centre for Biomedical Engineering, City, University of London, London, EC1V 0HB, UK. (e-mail: Karthik.Budidha@city.ac.uk; Elisa.Mejia-Mejia@city.ac.uk; p.kyriacou@city.ac.uk).

P. J. Aston and C. Pettit are with the Department of Mathematics, University of Surrey, Guildford, Surrey GU2 7XH, UK. (e-mail: p.aston@surrey.ac.uk; cp00809@surrey.ac.uk).

J. Mant is with the Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK. (e-mail: jm677@medschl.cam.ac.uk).

I. Introduction

THE photoplethysmogram (PPG) signal is acquired by a range of clinical and consumer devices, from pulse oximeters to smartwatches [1], [2]. It exhibits a pulse wave for each heartbeat, caused by the ejection of blood from the heart into the circulation. A wealth of physiological information can be deduced from the timing and shape of PPG pulse waves [3]. Consequently, a fundamental step in analysing the PPG is to detect individual pulse waves, corresponding to individual heartbeats. Indeed, several beat detection algorithms have been developed for the PPG, although it is not yet known how their performance compares.

It is important to assess the performance of beat detectors in different use cases where PPG signals can have different morphologies and levels of artifact [3]. Specifically, pulse oximeters acquire PPG signals at the finger close to major arteries, often with little motion artifact. In contrast, smart wearables such as smartwatches and fitness bands acquire the PPG at the wrist further from major arteries, often in challenging conditions such as during exercise. Assessing the performance of beat detectors across different use cases would allow one to select the best beat detector for a particular use case, and to understand its expected performance.

It is also important to investigate the impact of patient demographics and physiology on performance. First, it is important to assess performance during arrhythmias, since the PPG is now being used to identify atrial fibrillation (AF) [4]. Second, performance should be compared between ethnicities, as the performance of pulse oximeters has been found to be related to ethnicity [5]. Third, it is important to assess whether performance differs in babies, who have higher heart rates (HRs) than adults [6]. Assessing the impact of patient demographics and physiology on performance could highlight areas for future algorithm development.

This study aimed to: (i) develop an assessment framework with which to design and test PPG beat detectors; (ii) assess the performance of several beat detectors in different use cases; and (iii) investigate how their performance is affected by patient demographics and physiology. Fifteen open-source beat detectors were assessed against reference beats from electrocardiogram (ECG) signals in eight freely available datasets. This study builds on previous work which assessed the performance of four beat detectors on a single dataset [7], whereas this study assessed fifteen beat detectors across five datasets.

II. Materials and Methods

Ethical approval was not required for this study as it used pre-existing, anonymised data.

A. Datasets

The datasets used in this study are summarised in Table I, and are now described. For each dataset, the table indicates the duration of recordings and the total number of beats used in the analysis (shown for the MPSTD beat detector).

1) Hospital monitoring: A total of six datasets were used to assess performance during hospital monitoring: the CapnoBase and BIDMC datasets (which contain high-quality data), and four novel datasets extracted from the MIMIC Database (which contain real-world data).

The CapnoBase and BIDMC datasets were originally designed for developing and assessing PPG signal processing algorithms. They contain high-quality ECG and PPG signals with little artifact. Therefore, the performance of beat detectors on these datasets represents the best possible performance that could be expected in hospital monitoring. CapnoBase [8] contains data from 42 paediatric and adult subjects undergoing elective surgery and anaesthesia. BIDMC [9] contains data from 53 adults receiving critical care on a Medical Intensive Care Unit (46 subjects), Coronary Care Unit (6), or Surgical Intensive Care Unit (1). The BIDMC dataset was originally derived from the MIMIC-II Database [14], [15].

In addition, four novel datasets were extracted from the MIMIC-III Database [15], [16] for this study. These are named the 'MIMIC PERform' Datasets, as they contain (P) PPG, (E) ECG and (R) Respiration signals. These datasets were designed to be representative of real-world critical care data: their signals contain motion artifact and some low-quality periods. The MIMIC PERform Training and Testing Datasets each contain data 10 minutes of data from 200 patients, consisting of 100 adults and 100 neonates. The MIMIC PERform Testing Dataset was used to compare performance between adults and neonates in this study. The MIMIC PERform AF Dataset contains 20 minutes of data from 19 patients in AF, and 16 patients in normal sinus rhythm (non-AF). It was used to compare performance between AF and normal sinus rhythm. Labels of AF were obtained from manual annotations by cardiologists [10], [11]. The MIMIC PERform Ethnicity Dataset contains 10 minutes of data from 100 Black and 100 White subjects. It was used to compare performance between Black and White subjects, in keeping with [5]. All MIMIC PERform Datasets were extracted from the MIMIC-III Waveform Database, except for the Ethnicity Dataset, which was extracted from the MIMIC-III Matched Waveform Database [17]. Data were extracted by searching for MIMIC records which met the following criteria: (i) contain the required signals (PPG, ECG, and for all except the AF Dataset, respiration); (ii) are of sufficient duration ($\geq$ 10 minutes in the case of the Training, Testing and Ethnicity Datasets, and $\geq$ 20 minutes in the case of the AF Dataset); and (iii) contain minimal flat line segments (indicating sensor disconnection or saturation). The MIMIC Perform Datasets are available in [18].

TABLE I

DATASETS USED TO ASSESS THE PERFORMANCE OF PPG BEAT DETECTORS

Dataset	Subjects	PPG equipment	Reference beats	Duration (mins): med (quartiles)	Total beats
		Hospital monitoring (high-quality data)			
CapnoBase	42 patients undergoing elective surgery and routine anaesthesia [8].	Pulse oximeter at 300 Hz (upsampled from 100 Hz during acquisition)	Manual annotations of ECG (300 Hz)	7.7 (7.0 - 7.8)	24,945
BIDMC	53 critically-ill adult patients, a subset of the MIMIC II dataset [9].	Bedside monitor at 125 Hz (mostly finger PPG recordings)	ECG-derived QRS detections (125 Hz)	7.4 (6.9 - 7.7)	32,484
		Hospital monitoring (real-world data)			
MIMIC PERform Training Dataset	200 critically-ill patients during routine clinical care (100 adults, 100 neonates).	Bedside monitor at 125 Hz (mostly finger PPG recordings)	ECG-derived QRS detections (125 Hz)	5.7 (3.6 - 7.8)	115,941
MIMIC PERform Testing Dataset	200 critically-ill patients during routine clinical care (100 adults, 100 neonates).	Bedside monitor at 125 Hz (mostly finger PPG recordings)	ECG-derived QRS detections (125 Hz)	All: 5.2 (3.4 - 7.9) Adults: 7.7 (5.1 - 8.7) Neonates: 4.0 (2.6 - 5.3)	All: 116,585 Adults: 57,013 Neonates: 59,572
MIMIC PERform AF Dataset	35 critically-ill adults during routine clinical care (19 in AF, 16 not in AF), using AF labels provided by cardiologists [10], [11].	Bedside monitor at 125 Hz (mostly finger PPG recordings)	ECG-derived QRS detections (125 Hz)	AF: 17.8 (15.2 - 19.6) non-AF: 18.6 (17.3 - 19.4)	AF: 29,592 non-AF: 22,477
MIMIC PERform Ethnicity Dataset	200 critically-ill adults during routine clinical care (100 of Black ethnicity, 100 of White).	Bedside monitor at 125 Hz (mostly finger PPG recordings)	ECG-derived QRS detections (125 Hz)	Black: 8.0 (5.6 - 9.3) White: 7.0 (3.4 - 8.8)	Black: 61,756 White: 51,230
		Wearable data during different emotions			
WESAD	15 subjects during a laboratory-based protocol designed to induce different emotions [12].	Wristband (Empatica E4) at 64 Hz (with corresponding tri-axial accelerometry signals).	ECG-derived QRS detections (700 Hz)	Baseline: 19.1 (18.9 - 19.3) Amusement: 5.8 (5.8 - 5.8) Meditation: 6.3 (6.1 - 6.3) Stress: 10.3 (10.1 - 10.8)	Baseline: 20,519 Amusement: 6,213 Meditation: 6,395 Stress: 15,282
		Wearable data during activities of daily living			
PPG-DaLiA	15 subjects during a protocol of activities of daily living [13].	Wristband (Empatica E4) at 64 Hz (with corresponding tri-axial accelerometry signals).	Manual annotations of ECG (700 Hz)	Sitting: 9.8 (9.7 - 10.0) Working: 19.9 (19.7 - 20.5) Cycling: 7.8 (6.7 - 8.2) Walking: 10.8 (9.5 - 11.5) Lunch break: 32.4 (28.7 - 37.2) Car driving: 15.0 (14.1 - 15.8) Stair climbing: 7.5 (6.8 - 7.7) Table soccer: 4.8 (4.5 - 5.2)	Sitting: 9,022 Working: 21,272 Cycling: 13,956 Walking: 15,062 Lunch break: 37,247 Car driving: 18,883 Stair climbing: 12,466 Table soccer: 6,625

2) Wearable data: Two datasets were used, containing wrist PPG signals acquired using a wearable Empatica E4 device. The WESAD dataset was acquired during a protocol designed to induce different emotions: baseline, meditation, amusement, and stress. It contains data from 15 subjects, including 3 females, with a median age (lower - upper quartiles) of 27 (26 - 28) years, and BMI of 23 (22 - 25) kgm^{-2}. The PPG-DaLiA dataset was acquired during a protocol of activities of daily living, including: sitting, working, cycling, and running. It contains data from 15 subjects, including 3 females aged 28 (24 - 36) years, with a BMI of 22 (21 - 23) kgm^{-2}, and skin types on the Fitzpatrick scale of: 2 (1 subject), 3 (11 subjects), and 4 (3 subjects).

B. PPG Beat Detection

First, any PPG signals sampled at over 100 Hz were resampled at this frequency to reduce the time for computational analysis. For signals sampled at multiples of 100 Hz, this was performed using downsampling, and for other signals it was

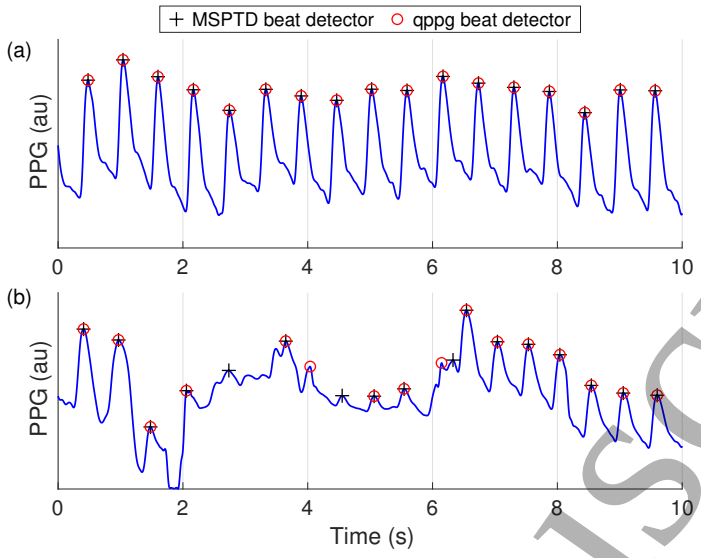

Fig. 1. **Detecting beats in the photoplethysmogram (PPG):** PPG pulse peaks detected by two beat detectors. (a) shows a high quality segment in which beats were accurately detected by both beat detectors; (b) includes a period of low quality between 1 and 7 s in which the two beat detectors disagreed. *au - arbitrary units*

performed using resampling with an antialiasing lowpass filter. Second, signals were band-pass filtered between 0.67 and 8.0 Hz to eliminate non-cardiac frequencies. Third, beats were detected using fifteen open-source PPG beat detectors in turn, as demonstrated for two beat detectors in Fig. 1. The beat detectors are described in Table II. Beat detection was performed on 20 s windows of PPG signal, overlapping by 5 s. Repeated beat detections due to overlapping windows were eliminated. This approach ensured that beat detectors were not penalised for missing beats at the start or end of a window. Fourth, windows were excluded if they contained a flat line lasting more than 0.2 s (typically caused by sensor disconnection or signal 'clipping'). The beat detectors are available in [19].

For consistency, each beat detector's annotations were used to obtain the corresponding middle-amplitude point of the systolic upslope on each detected PPG pulse wave [39], which was used for analysis. This point has been found to provide more accurate timings than peaks or onsets [39].

C. Reference ECG Beat Detection

The CapnoBase and PPG-DaLiA datasets contain manual beat annotations which were used as reference beats. In the remaining datasets reference beats were obtained from simultaneous ECG signals by: (i) detecting beats using two separate ECG beat detectors; (ii) identifying 'correct' beats as those which both beat detectors detected within 150 ms of each other; and (iii) excluding from the analysis any 20 s windows in which the two beat detectors did not agree. The two beat detectors were: the 'jqrs' ECG beat detector, which is based on the Pan and Tompkins method [40], [41] and the 'rpeakdetect' ECG beat detector [42].

D. Aligning PPG Beats with Reference ECG Beats

PPG and ECG signals were not necessarily precisely aligned, so the timings of PPG-derived beats and reference ECG-derived beats were aligned as follows. The time difference between each ECG-derived beat and its closest PPG-derived beat was calculated. Those ECG-derived beats for which the absolute time difference was $<$150 ms were determined to be correctly identified. This process was repeated when offsetting the beats by lags of -10 to 10 s, in increments of 20 ms. The lag which resulted in the highest proportion of correctly identified beats was accepted as the true lag and used to synchronise the timings of beats. Fig. 2(a) shows an example of this time-alignment.

E. Statistical Analysis

The ability of beat detectors to detect beats was assessed by comparing PPG-derived beats with reference beats. Reference beats were determined to be correctly identified if the closest PPG-derived beat was within $\pm$150 ms of a reference beat, as shown in Fig. 2(b). For each recording, the numbers of reference beats (n_{ref}), PPG-derived beats (n_{PPG}), and correctly identified beats (n_{correct}) were used to calculate the following:

$$\text{sensitivity (\%)}, \quad \text{Se} = \frac{n_{\text{correct}}}{n_{\text{ref}}} \times 100 \tag{1}$$

TABLE II
PPG BEAT DETECTORS

Beat Detector	Implementing Author	Original Author	Description
ABD: Automatic Beat Detection [20]	P. Charlton	M. Aboy *et al.*	The PPG is strongly filtered to retain frequencies around an initial heart rate estimate, differentiated, and peaks are detected above the 75^{th} percentile. Beats are identified as peaks in a weakly filtered PPG immediately following each peak identified in the differentiated signal.
AMPD: Automatic Multiscale Peak Detection [21]	P. Charlton	F. Scholkmann *et al.*	The PPG is detrended and segmented into 6s windows. A local maxima scalogram (LMS) is calculated: a matrix of random numbers, where the rows correspond to different scales (ranging from one sample to half the window duration), and the columns indicate PPG samples. The LMS values are set to zero when a PPG sample is higher than its neighbours at that particular scale. The LMS is truncated to only include scales smaller than the scale at which the most local maxima were identified. Beats are identified as samples which are deemed to be local maxima at all remaining scales.
ATM: Adaptive Threshold Method [22], [23]	D. Han	H. Shin *et al.*	The PPG is bandpass filtered between 0.5 and 20 Hz. Troughs are identified as local minima which are below an adaptive threshold. The adaptive threshold increases from the value of the previous trough, at a rate related to the PPG amplitude. Any troughs occuring within a period of 0.6 times the previous inter-beat-interval are excluded. The 'Vmin' implementation of this beat detector was used, as it performed slightly better than the 'Vmax' implementation in initial testing.
COppg: Percentile Peak Detector [24]	P. Charlton, C. Orphanidou, A. Darrell	C. Orphanidou *et al.*	In each 10 s PPG segment, beats are identified as peaks which are sufficiently close to (or above) the 90^{th} percentile of the PPG signal, using adaptive filtering.
ERMA: Event-Related Moving Averages [25]	E. Mejía-Mejía	M. Elgendi *et al.*	The PPG is bandpass filtered between 0.5 and 8 Hz, rectified to eliminate values below zero, and squared. Two moving averages are calculated: (i) $\mathrm{MA}_{\mathrm{peak}}$, a moving average of period 111 ms, emphasising systolic peaks; and (ii) $\mathrm{MA}_{\mathrm{beat}}$, a moving average of period 667 ms, emphasising individual beats. Beats are identified as maxima within periods lasting $\geq$ 111 ms where $\mathrm{MA}_{\mathrm{peak}} > \mathrm{MA}_{\mathrm{beat}} + \alpha$ (where α is a threshold).
HeartPy [26], [27]	P. Charlton	P. van Gent *et al.*	The PPG is squared and normalised. Peaks are detected as maxima above a moving average (of period 0.75s). This is repeated for moving averages of different amplitudes, producing a set of peaks for each amplitude. The set of peaks which produces a plausible HR and the lowest variability in inter-beat intervals (IBIs) is selected as the set of beats. Beats which result in outlying IBIs are eliminated.
IMS: Incremental Merge Segmentation [28]	M. Pimentel	W. Karlen *et al.*	Beats are detected at the end of continuous positive gradient segments (systolic upslopes) with an acceptable amplitude and duration, where the amplitude thresholds are adaptively calculated.
MSPTD: Multi-Scale Peak & Trough Detection [29]	S. Bishop	S. Bishop & A. Ercole	A modification of AMPD in which LMS matrices are calculated for both local maxima and local minima, so the algorithm detects both peaks and onsets. MSPTD also contains some optimisations to improve computational efficiency.
PDA: Peak Detection Algorithm [30]	E. Mejía-Mejía	E.J. Argüello Prada & R.D. Serna Maldonado	Systolic peaks are identified as peaks which follow an upslope (*i.e.* period of positive gradient) lasting $\geq$ 60% of the duration of the upslope leading to the previously detected systolic peak.
PWD: Pulse Wave Delineator [31]	B.N. Li	B.N. Li *et al.*	Pulse onsets and pulse peaks are identified from zero-crossing points in the first derivative of the PPG: onsets are identified as zero-crossing points before a maximal deflection, and peaks are identified as zero-crossing points immediately following maximal deflections.
Pulses: PPG Pulses Detector [32]	J. Lazaro, M. Llamedo Soria	J. Lazaro *et al.*	Peaks are identified in the differentiated PPG using an adaptive filter set to the amplitude of the previous peak, and decreases for a period after that peak at a rate dependent on previous inter-beat intervals. Beats are identified as maxima in the PPG within 300ms of each peak in the differentiated PPG.
qppg: Adapted Onset Detector [33]	W. Zong, G. Moody, Q. Li	W. Zong	Systolic upslopes are detected from a signal generated with a slope sum function, which sums the magnitudes of the PPG upslopes in the previous 0.17 s. Adaptive thresholding is used to identify systolic upslopes in this signal. The 'qppgfast' implementation of this beat detector was used, after testing showed it performed similarly to the original 'qppg' implementation.
SPAR: Symmetric Projection Attractor Reconstruction [34]	C. Pettit & P.J. Aston	C. Pettit *et al.*	The PPG is segmented into 20 s windows and time delay coordinates are used to represent it in 7-dimensional phase space with the time delay set to one seventh of the average inter-beat interval. The Symmetric Projection Attractor Reconstruction method is then used to construct an appropriate 2-dimensional projection of the phase space [35], [36]. Beats are identified as times at which the orbit crosses the x-axis. This implementation uses information from previous windows to inform beat detections in the current window.
SWT: Stationary Wavelet Transform [37]	D. Han	S. Vadrevu & M. Sabarimalai Manikandan	The PPG is decomposed using the Stationary Wavelet Transform. Multi-scale sum and products of selected detail subbands are calculated to emphasise systolic upslopes. An envelope is then extracted by: adaptive thresholding to reduce the influence of noise; calculating the Shannon entropy; and smoothing the result. Finally, beats are identified in the envelope using a Gaussian derivative filter.
WFD: Wavelet Foot Delineation [38]	E. Mejía-Mejía	N. Conn & D. Borkholder	The PPG is bandpass filtered between 0.5 and 8 Hz, and interpolated to 250 Hz. It is decomposed using a wavelet transform, retaining the fifth wavelet scale for analysis. This signal is rectified and squared to eliminate values below zero. Regions containing beats are identified as those where the signal exceeds a low-pass filtered version of the signal. The timing of the beat within each region is identified as the first zero-crossing of the third derivative, or failing that, the maximum in the second derivative.

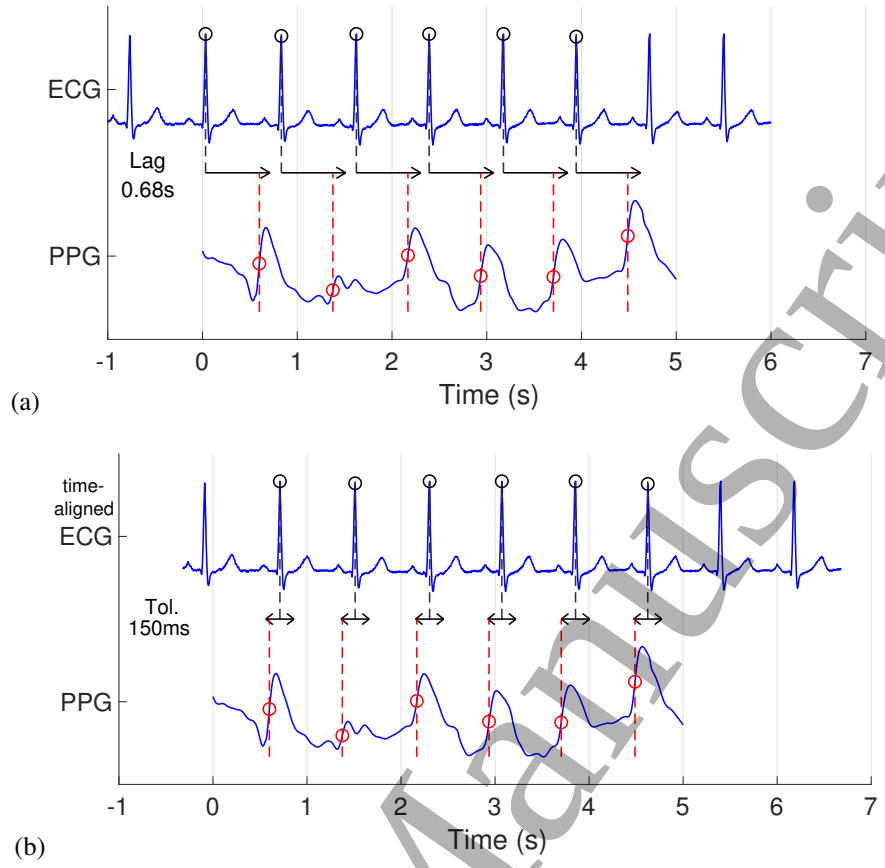

Fig. 2. **Comparing PPG-derived beats with reference beats:** (a) Time-alignment of electrocardiogram (ECG) and photoplethysmogram (PPG) signals. The time lag between ECG and PPG signals (0.6s in this case) was automatically identified from ECG and PPG beat timings. (b) Assessing the ability of a beat detector to detect beats in the PPG. Those beats detected in the PPG (red circles) which occured within ±150ms of time-aligned reference ECG beats were deemed to be correct.

$$\text{positive predictive value (\%)}, \quad \text{PPV} = \frac{n_{\text{correct}}}{n_{\text{PPG}}} \times 100 \tag{2}$$

$$F_1 \text{ Score (\%)}, \quad \text{F}_1 = \frac{2 \times \text{PPV} \times \text{Se}}{\text{PPV} + \text{Se}} \times 100 \tag{3}$$

Beat detectors were ranked according to the F_1 score, which is the harmonic mean of sensitivity and PPV.

The accuracy of PPG-derived heart rates (HRs) was assessed by comparing PPG-derived HRs to reference ECG-derived HRs. A HR (in beats per minute, bpm) was calculated at the time of each PPG-derived beat, from the number of PPG-derived beats in the preceding 8 s window (n_{beats}), as

$$HR = 60 \times \frac{n_{\text{beats}} - 1}{t(n_{\text{beats}}) - t(1)} \tag{4}$$

where t denotes the times of PPG-derived beats. Each HR signal was interpolated using sample-and-hold interpolation at 50 Hz. Performance was assessed as the mean absolute percentage error (MAPE) between time series. A median MAPE of $<10\%$ was deemed to be acceptable for HR monitoring. This was based on the acceptable limits of $\pm 10\%$ stated in the AAMI standard [43] and implemented using the MAPE statistic in [44], although we note that the true threshold of acceptability is likely to vary between applications [45].

Performance statistics are reported as median (25^{th} - 75^{th} percentiles). The Wilcoxon rank sum test was used to compare performances between groups, at a significance level of α = 0.05. A Holm-Sidak correction was made to correct for multiple comparisons.

III. Results

The main results are summarised in Table III. This table reports the performance of beat detectors (F_1 score) and their performance for HR monitoring (HR MAPE). Results are provided for the best-performing beat detectors (found to be MSPTD and qppg, as detailed in Section III-B), and all beat detectors (reported as the range in performance metrics from the worst to the best performance).

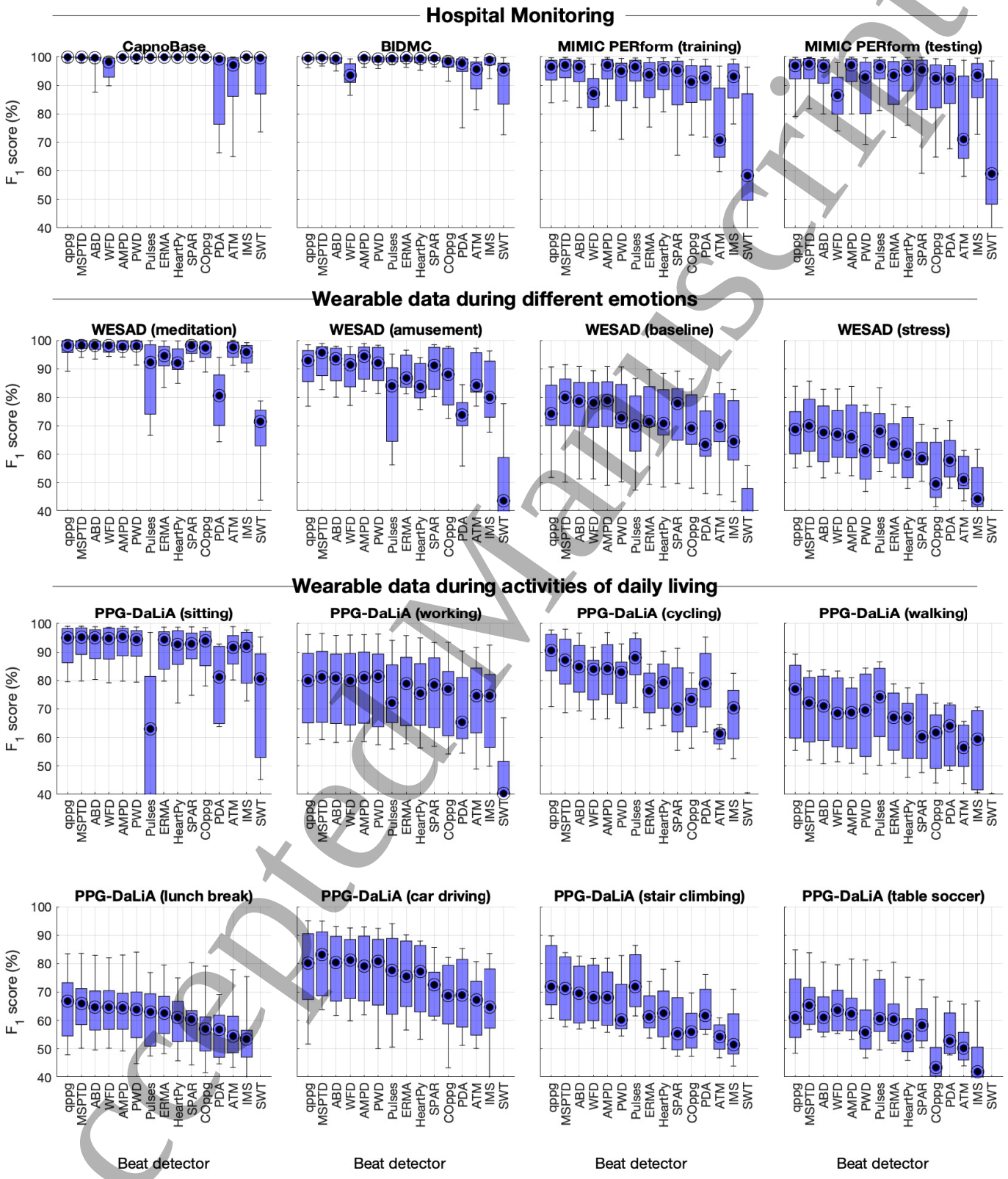

Fig. 3. Box plots showing the performance of beat detectors, expressed as the F_1 score. Each graph shows the results for each of the beat detectors on a particular dataset. Performance is shown as the median (circles), inter-quartile range (boxes), and 10^{th} and 90^{th} percentiles (whiskers) across subjects. *See Table II for definitions of beat detectors.*

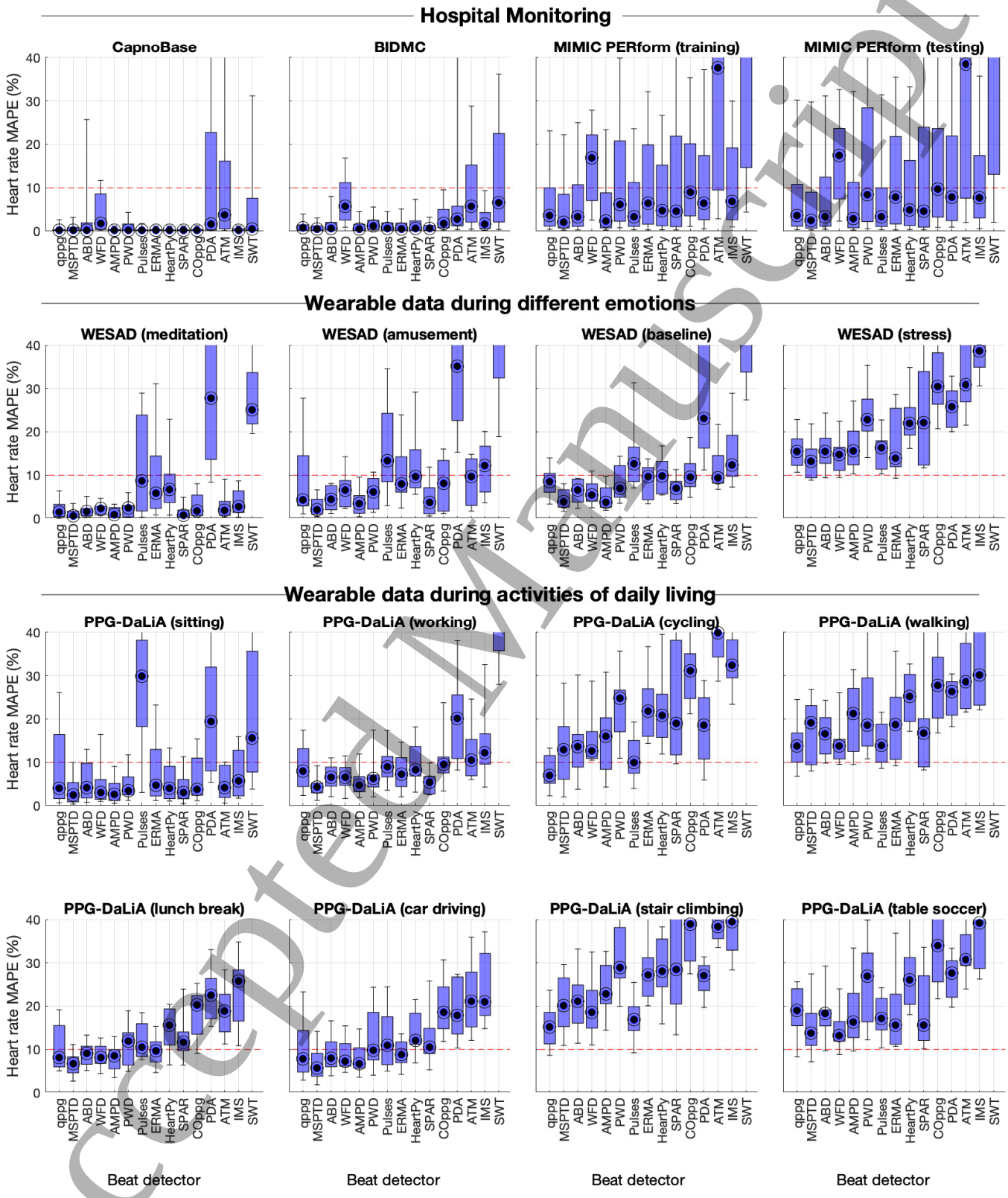

Fig. 4. Box plots showing the performance of beat detectors, expressed as the heart rate mean absolute percentage error (MAPE). Each graph shows the results for each of the beat detectors on a particular dataset. Performance is shown as the median (circles), inter-quartile range (boxes), and 10^{th} and 90^{th} percentiles (whiskers) across subjects. Dashed red lines indicate the acceptable performance of 10% MAPE. *See Table II for definitions of beat detectors.*

TABLE III
THE PERFORMANCE OF BEAT DETECTORS IN DIFFERENT USE CASES

Dataset	median F_1 score (%)			median HR MAPE (%)		
	MSPTD	qppg	All (min - max)	MSPTD	qppg	All (min - max)
Hospital Monitoring (high-quality data)						
CapnoBase	99.9	99.9	97.1 - 99.9	0.2	0.2	0.2 - 3.7
BIDMC	99.7	99.6	93.4 - 99.7	0.5	0.7	0.5 - 6.5
Hospital Monitoring (real-world data)						
MIMIC PERform Training Dataset	97.2	96.5	58.4 - 97.2	2.1	3.7	2.1 - 48.6
MIMIC PERform Testing Dataset	97.5	96.9	59.0 - 97.5	2.4	3.5	2.4 - 51.0
MIMIC PERform Testing Dataset (adults)	98.5	98.0	91.9 - 98.5	1.1	2.2	1.1 - 13.5
MIMIC PERform Testing Dataset (neonates)	95.9	95.2	50.7 - 95.9	4.9	5.5	4.8 - 59.7
MIMIC PERform AF Dataset (AF)	96.7	97.1	75.3 - 97.1	4.3	3.3	3.3 - 34.9
MIMIC PERform AF Dataset (non-AF)	99.7	99.6	91.3 - 99.7	0.4	0.6	0.4 - 6.9
MIMIC PERform Ethnicity Dataset (Black)	98.5	98.2	91.2 - 98.5	1.4	2.3	1.4 - 9.9
MIMIC PERform Ethnicity Dataset (White)	97.5	97.3	86.6 - 97.5	2.1	3.5	2.1 - 14.6
Wearable data during different emotions						
WESAD (meditation)	98.2	98.3	71.5 - 98.3	0.6	1.5	0.6 - 27.8
WESAD (amusement)	95.6	92.8	43.6 - 95.6	2.0	4.4	2.0 - 44.8
WESAD (baseline)	80.1	74.2	37.0 - 80.1	3.8	8.6	3.8 - 41.8
WESAD (stress)	70.1	68.7	17.9 - 70.1	13.2	15.5	13.2 - 67.7
Wearable data during activities of daily living						
PPG-DaLiA (sitting)	95.1	95.1	63.1 - 95.5	2.5	4.1	2.5 - 29.9
PPG-DaLiA (working)	81.2	80.0	40.3 - 81.4	4.3	8.0	4.3 - 48.6
PPG-DaLiA (cycling)	87.1	90.6	33.6 - 90.6	13.0	7.0	7.0 - 69.0
PPG-DaLiA (walking)	72.1	76.9	31.2 - 76.9	19.1	13.7	13.7 - 63.2
PPG-DaLiA (lunch break)	66.0	66.8	22.8 - 66.8	6.7	8.2	6.7 - 59.7
PPG-DaLiA (car driving)	83.1	80.2	30.5 - 83.1	5.7	7.8	5.7 - 61.0
PPG-DaLiA (stair climbing)	71.3	71.9	27.9 - 71.9	20.1	15.1	15.1 - 71.9
PPG-DaLiA (table soccer)	65.3	61.0	19.8 - 65.3	13.9	19.1	13.3 - 65.7

A. Performance of beat detectors in different use cases

The performance of beat detectors is presented in Fig. 3 using the F_1 score, and in Fig. 4 using the HR MAPE. Additional results are provided in Appendix I for sensitivity and PPV (Figs. A1 and A2 respectively). The key findings are as follows.

First, eight beat detectors performed very well across all datasets with low levels of movement: AMPD, MSPTD, qppg, PWD, ERMA, SPAR, ABD, and HeartPy. These had median F_1 scores of: $\geq$ 99% on the hospital monitoring datasets containing high-quality data (CapnoBase and BIDMC); $\geq$ 90% on the hospital monitoring datasets containing real-world data (MIMIC PERform Training and Testing Datasets); and $\geq$ 90% on the wearable datasets with low levels of movement (WESAD (meditation) and PPG-DaLiA (sitting)). The remainder of the Results will focus on these eight beat detectors. Fig. 5 (a) shows an example of (mostly) accurate beat detection during low levels of movement. Of note, the Pulses beat detector performed less well on the PPG-DaLiA (sitting) dataset because its assumed duration of the systolic upslope was no longer valid in these wrist signals acquired at rest.

Second, performance decreased during activities associated with more movement. The eight beat detectors which performed well on data with low levels of movement had median F_1 scores of 93-96% on PPG-DaLiA (sitting). This performance decreased to 70-91% on PPG-DaLiA (cycling), 60-77% on PPG-DaLiA (walking), and 55-72% on PPG-DaLiA (stair climbing). Performance was also poorer during stress, as shown by median F_1 scores of 59-70% on WESAD (stress) compared to 71-80% on WESAD (baseline). This was primarily due to beat detectors missing beats, rather than falsely detecting beats, as shown by the generally lower sensitivities than positive predictive values on PPG-DaLiA (walking) and WESAD (stress) datasets (see Appendix I, Figs. A1 and A2). Fig. 5 (b-d) show examples of beat detection during movement.

Third, the variability in performance between subjects was low during activities associated with low levels of movement, as shown by the relatively low inter-quartile ranges of F_1 scores (indicated by the heights of boxes) on WESAD (meditation) and PPG-DaLiA (sitting). However, performance varied much more between subjects in more challenging datasets, *e.g.* WESAD (stress) and PPG-DaLiA (walking).

B. Best-performing beat detectors

To identify the best-performing beat detectors, we focused on results from the MIMIC PERform Testing and PPG-DaLiA (working) datasets, since these are representative of real-world performance in critical care and daily life respectively. On the MIMIC PERform Testing Dataset, the top scoring beat detectors were MSPTD, AMPD, gppq, ABD, and Pulses (all with

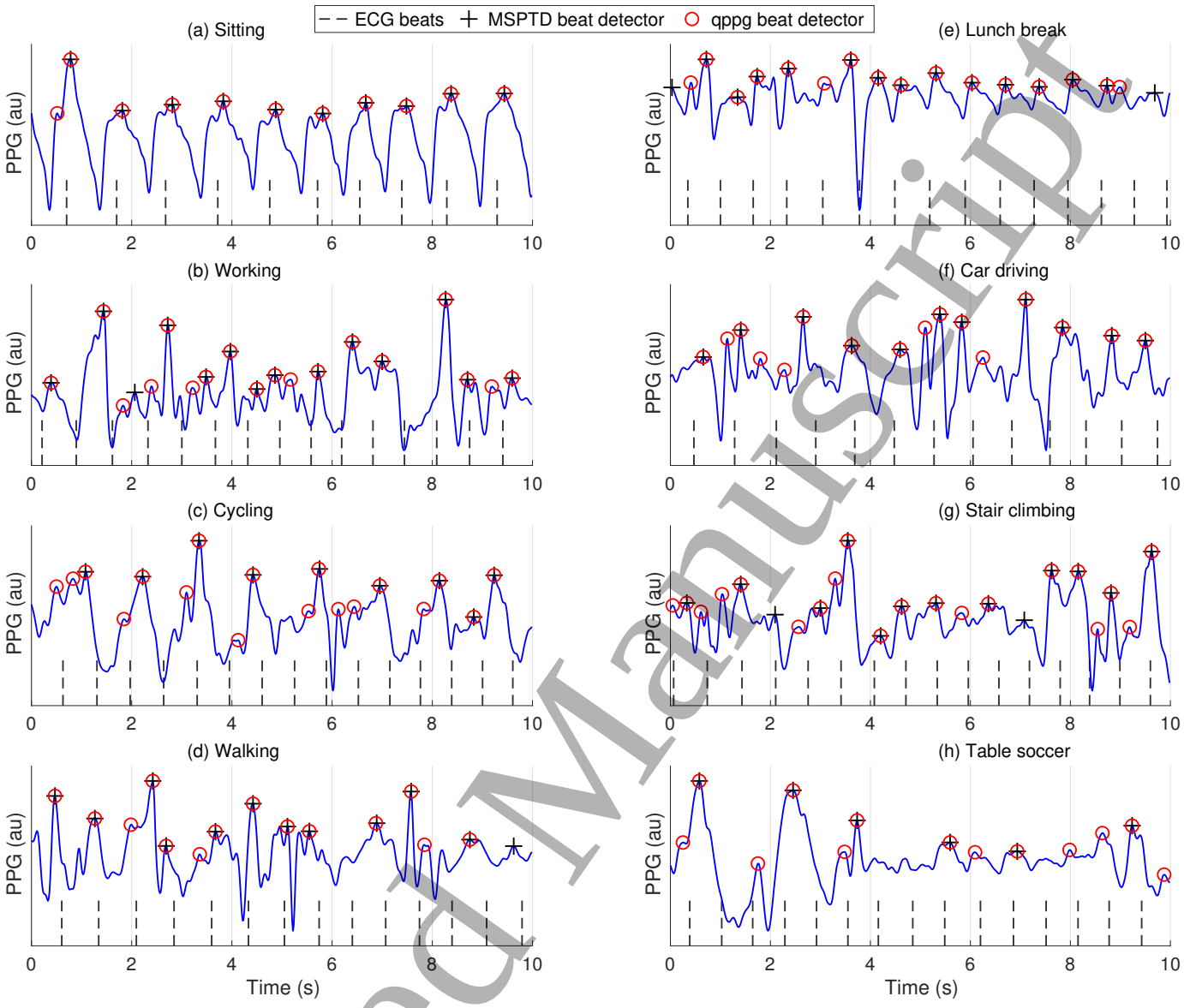

Fig. 5. **PPG beat detection during different activities:** PPG signals are shown for different activities of daily living from the PPG-DaLiA dataset. Beats detected by two PPG beat detectors are shown alongside reference ECG beats. *au - arbitrary units*

F_1 scores of 96.6-97.5 %, whereas the remainder scored $\leq$ 95.6 %). On PPG-DaLiA (working), the top scorers were PWD, MPSTD, AMPD, ABD, gppq, and WFD (all with F_1 scores of 80.0-81.4 %, whereas the remainder scored <79.0 %). In addition, MSPTD was the best performing beat detector on 5 out of the 12 WESAD and PPG-DaLiA datasets, and qppg was the best performing beat detector on 4 of these datasets. Therefore, we suggest that MSPTD and qppg performed best, although we note that this is subjective, and that some other beat detectors also performed well (notably ABD and AMPD).

The best-performing beat detectors have complementary performance characteristics: MSPTD tended to have a higher positive predictive value, whereas qppg tended to have higher sensitivity (see Appendix I, Figs. A1 and A2). Fig. 5 shows examples of this: qppg sometimes detected additional beats during noise (see Fig. 5(a) at 0.5 s), whereas MSPTD sometimes missed beats (see Fig. 5(c) at 7.5-9 s).

C. Acceptability for heart rate monitoring

The performance of beat detectors was deemed to be acceptable for HR monioring in some use cases but not others (see Fig. 4). All eight beat detectors which had been found to perform well on data with low levels of movement also had acceptable HR MAPEs of <10% on datasets associated with low and moderate levels of movement (the hospital monitoring datasets, and WESAD (meditation, amusement, baseline) and PPG-DaLiA (sitting, working)). At least some of these beat detectors did not perform acceptably on each of the remaining datasets. None of the eight beat detectors produced acceptable HR errors during stress (see WESAD (stress)). Five of the eight beat detectors (MSPTD, qppg, ABD, AMPD, and ERMA) also

produced acceptable errors during less intensive activities (PPG-DaLiA (lunch break), and PPG-DaLiA (car driving)). Only qppg performed acceptably on PPG-DaLiA (cycling). None of the beat detectors performed acceptably during more intensive exercise (PPG-DaLiA (walking), PPG-DaLiA (stair climbing), and PPG-DaLiA (table soccer)).

D. Association between performance and patient physiology and demographics

The associations between beat detector performance and the assessed factors are shown in Fig. 6.

The performance of beat detectors was poorer in AF (Fig. 6(a)). The eight beat detectors which performed well at rest achieved F_1 scores between 99.4-99.7% in sinus rhythm (non-AF), compared to 91.8-97.1% in AF. This was primarily because beat detectors missed beats during AF (see Appendix II, Figs. A3(a) and A4(a)), similarly to performance in movement. Performance was worse in AF subjects than non-AF subjects for all eight beat detectors at the 5% significance level, and four of these differences remained significant after accounting for multiple comparisons (0.2% significance level).

All eight beat detectors performed worse on neonates than adults, as shown in (Fig. 6(b)). Seven of these differences remained significant after accounting for multiple comparisons. These beat detectors missed beats, as shown by their lower sensitivities (see Appendix II, Fig. A3(b)). The lower performance in neonates may be because the neonatal PPG signals were of lower quality, as shown by them having lower SNRs (-10.9 (-12.2 - -8.8) dBc in neonates compared to -5.9 (-9.6 - -1.6) dBc in adults). In addition, some beat detectors may have been designed for use with adult data, who typically have HRs between 60 and 100 bpm, compared to neonatal HRs of 110-160 bpm [6].

Five beat detectors had lower F_1 scores on White subjects than Black subjects, as shown in (Fig. 6(c)), although none of these differences were significant after accounting for multiple comparisons.

E. Assessment framework

Table IV presents the proposed assessment framework. The MIMIC PERform datasets are recommended for developing and testing algorithms, and for comparing performance between adults and neonates. Out of the wearable datasets, WESAD is recommended for training and PPG-DaLiA for testing, as the latter allows performance to be assessed during several activities of daily living. The MIMIC PERform AF Dataset is recommended for assessing performance in AF, although it would benefit from inclusion of additional subjects in the future. The CapnoBase and BIDMC datasets were designated as 'preliminary design' datasets as all beat detectors achieved F_1 scores of $>$93% on these datasets, so it is unlikely they could be used to substantially improve beat detector design.

IV. Discussion

This study assessed the performance of several open-source PPG beat detectors across a range of datasets. Most beat detectors performed well on hospital data and at rest, but performed worse during movement, stress, and AF, and in neonates. The study provides a standardised framework with which to develop and test beat detectors.

The findings could inform PPG-based monitoring strategies and directions for algorithm development. The poorer performance of beat detectors during movement is reflected in current monitoring strategies. For instance, smartwatches which use the PPG to check for an irregular pulse often only do so whilst the subject is stationary [4] - a strategy which is supported by this study. Future work should investigate how best to use a simultaneous accelerometry signal to identify periods in which the subject is stationary and therefore beats can be accurately detected. The poorer performance in neonates and during AF indicates areas for development [23]. Future work could also assess performance in other situations which impact the pulse wave, such as during ectopic beats, hypoperfusion, and vascular disease. This study also provides motivation for strategies to improve beat detection and exclude unreliable data from analyses, such as motion artifact cancellation and signal quality assessment.

The beat detectors used in this study are indicative of the range of approaches proposed in the literature to detect beats in the PPG. As detailed in Table II, approaches included: (i) identifying peaks in the original PPG signal (HeartPy and COppg); (ii) identifying systolic upslopes using the original signal (IMS) or first derivative (qppg, ABD, PWD and Pulses); (iii) using the local maxima scalogram to identify peaks across several scales (MSPTD and AMPD); and (iv) representing the PPG in phase space (SPAR). The MSPTD and qppg beat detectors performed best in this study. MSPTD searches for peaks without using any prior knowledge of the characteristics of PPG pulse waves. In contrast, qppg searches for systolic upslopes based on their expected characteristics. In the future, different approaches could be combined to improve performance.

The algorithms, datasets, and assessment framework used in this study are all publicly available. This has several benefits. Firstly, it ensures that the study is reproducible. Secondly, it allows others to assess the performance of their own beat detection or quality assessment algorithms. Thirdly, the framework provides a basis with which to design (using the training datasets) and test such algorithms. Since the training datasets contain a variety of challenges, such as different use cases and populations, we expect that developers will benefit from using this framework for algorithm development. The framework cannot be considered to be exhaustive, and datasets recorded in additional settings, from further patient populations, could be added in the future. These resources and corresponding documentation are archived at [18], [19], whilst information on the most up to date version can be obtained at: https://github.com/peterhcharlton/ppg-beats.

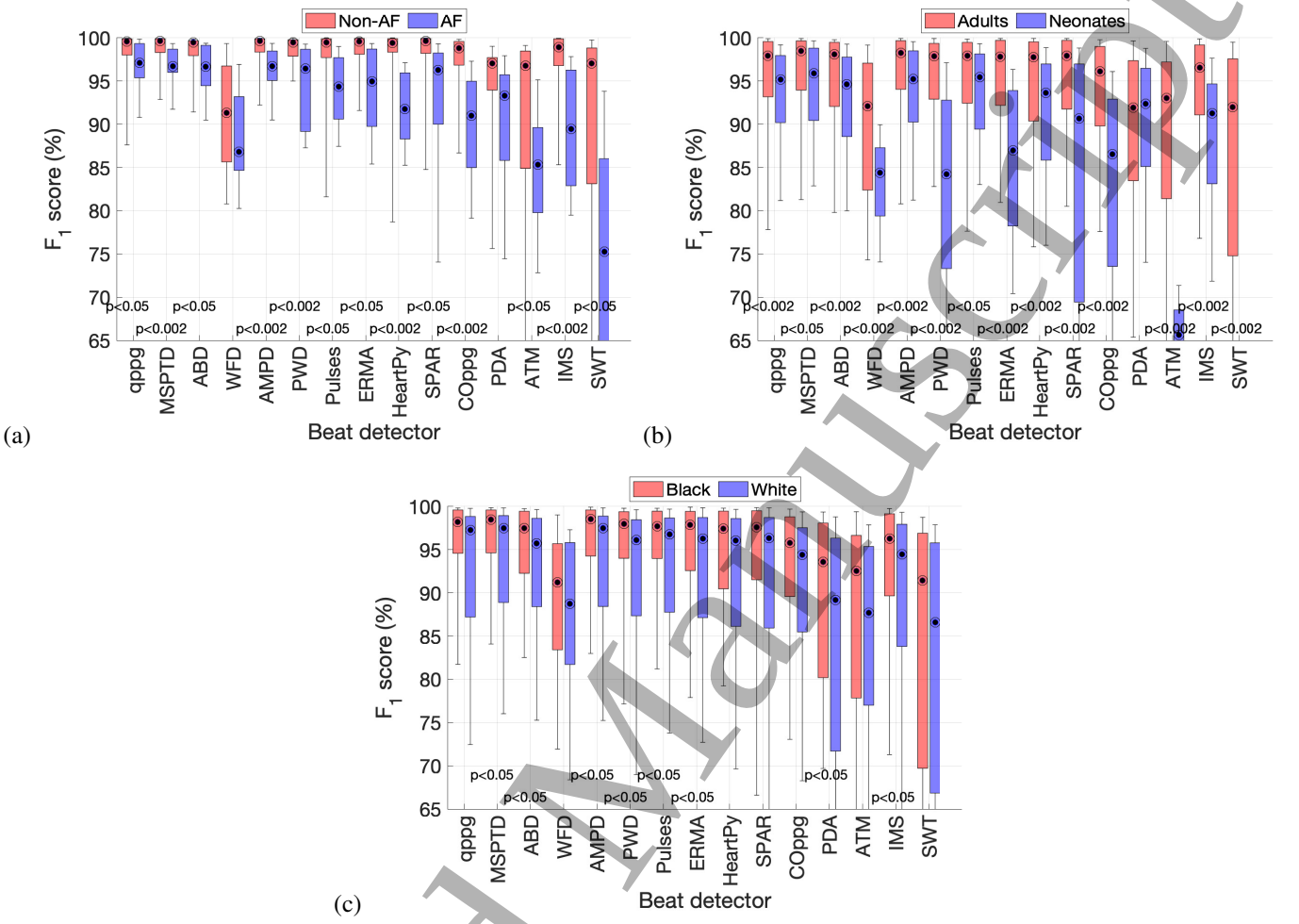

Fig. 6. Box plots showing the associations between beat detector performance and patient physiology and demographics. (a) comparison of subjects without and with atrial fibrillation (non-AF and AF); (b) adults compared to neonates; (c) Black compared to White subjects. Performance is shown as the median (circles), inter-quartile range (boxes), and $10^{\mathbf{th}}$ and $90^{\mathbf{th}}$ percentiles (whiskers) across subjects. *See Table II for definitions of beat detectors.*

TABLE IV
THE PROPOSED ASSESSMENT FRAMEWORK

Purpose	Dataset	Data access
	Algorithm Development	
Preliminary design	CapnoBase	Available in Matlab format after completing an agreement.
Preliminary design	BIDMC	Available in CSV, WaveForm DataBase, and Matlab format, under an ODbL licence.
Design with critical care data, and compare performance in adults and neonates	MIMIC PERform Training Dataset	Available in Matlab format, under an ODbL licence.
Design with wearable data	WESAD	Available in Python's PKL format, for non-commercial purposes.
Investigate impact of atrial fibrillation	MIMIC PERform AF Dataset	Available in Matlab format, under an ODbL licence.
	Algorithm Testing	
Testing with critical care data	MIMIC PERform Testing Dataset	Available in Matlab format, under an ODC-By licence.
Testing in activities of daily living	PPG-DaLiA	Available in Python's PKL format, for non-commercial purposes.

The key limitations are as follows. First, the study is limited to open-source beat detectors, rather than all those reported in the literature (see [3] for a description of additional beat detectors). Second, no attempt was made to improve the algorithms, but rather this study established the performance of existing algorithms. Third, some datasets were relatively small: WESAD and PPG-DaLiA contain data from 15 subjects, and the MIMIC PERform AF Dataset contains data from 35 patients. Fourth, the framework assumes that pulse arrival time (PAT) is constant within a subject's recording, which is reasonable for the short recordings in this study, but changes in PAT should be accounted for if using longer recordings [7].

V. Conclusions

This study demonstrated the high performance of the MSPTD and qppg beat detectors across a range of use cases. Most beat detectors performed well in the absence of movement, whereas performance was poorer during stress, activities of daily living, in neonates, and during AF. The results inform key directions for future work: (i) improving performance in neonates and during AF; (ii) investigating whether motion artifact cancellation improves performance; and (iii) investigating whether algorithms to assess signal quality can distinguish between periods in which beats can or cannot be accurately detected. The algorithms, datasets, and assessment framework used in this study are all publicly available in [18], [19].

Acknowledgment

This work was supported by British Heart Foundation (BHF) grants [FS/20/20/34626] and [PG/15/104/31913], and an EPSRC Impact Acceleration Award to PHC. PHC acknowledges Jonah Spencer's valuable input on the importance of PPG beat detection in neonates.

Competing Interests

P. J. Aston has a patent (WO2015121679A1 "Delay coordinate analysis of periodic data"), which covers the foundations of the SPAR method used in this paper.

Appendix I
Performance of PPG beat detectors in different use cases

The performance of photoplethysmogram (PPG) beat detectors in different use cases was presented in Fig. 3 in the main text, using the F_1 score to describe performance. Additional results are shown in: Fig. A1, which shows the sensitivity of beat detectors; and Fig. A2, which shows their positive predictive value.

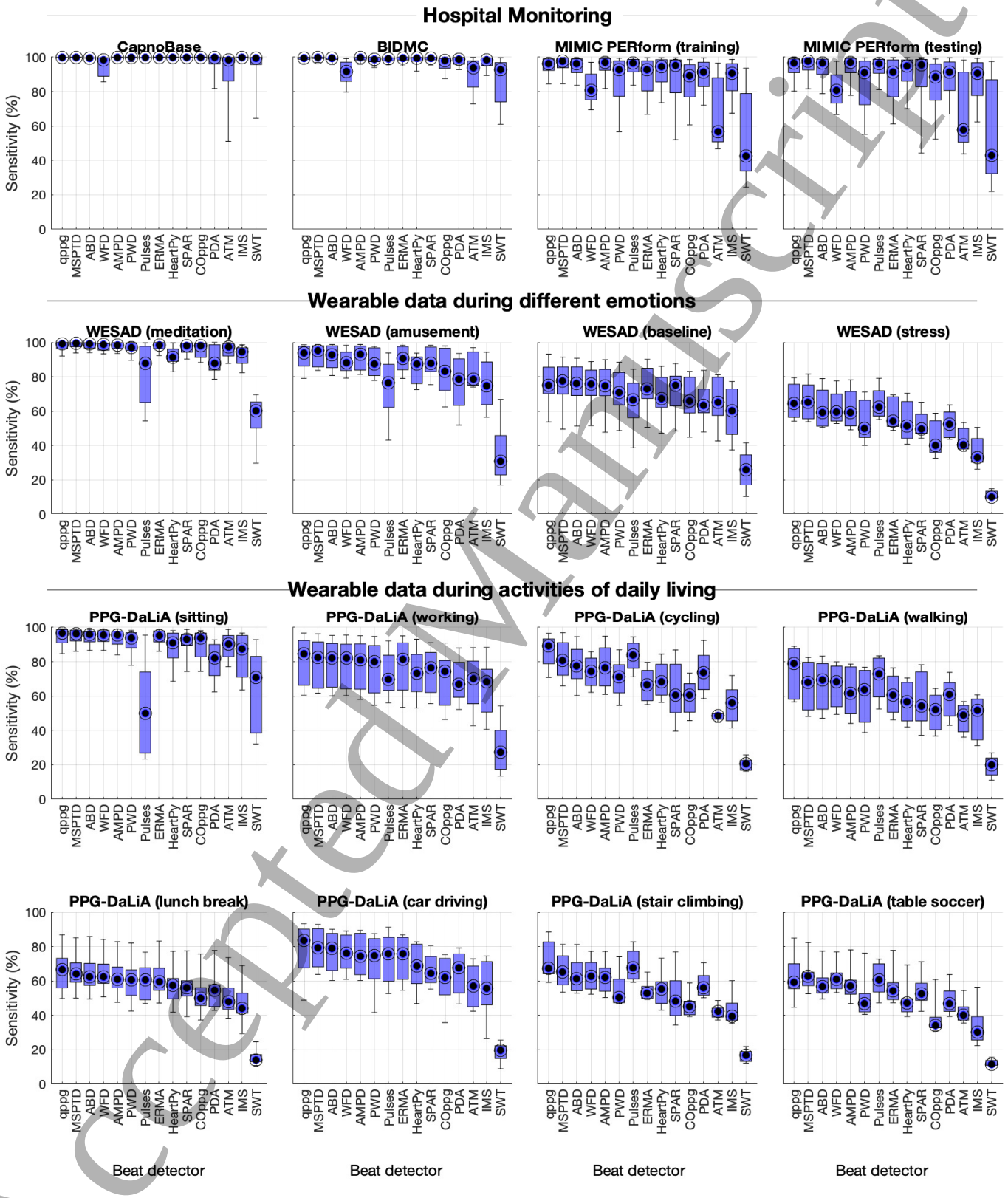

Fig. A1. Box plots showing the performance of beat detectors, expressed as the sensitivity. Each graph shows the results for each of the beat detectors on a particular dataset. Performance is shown as the median (circles), inter-quartile range (boxes), and 10^{th} and 90^{th} percentiles (whiskers) across subjects. *See Table 2in the main text for definitions of beat detectors.*

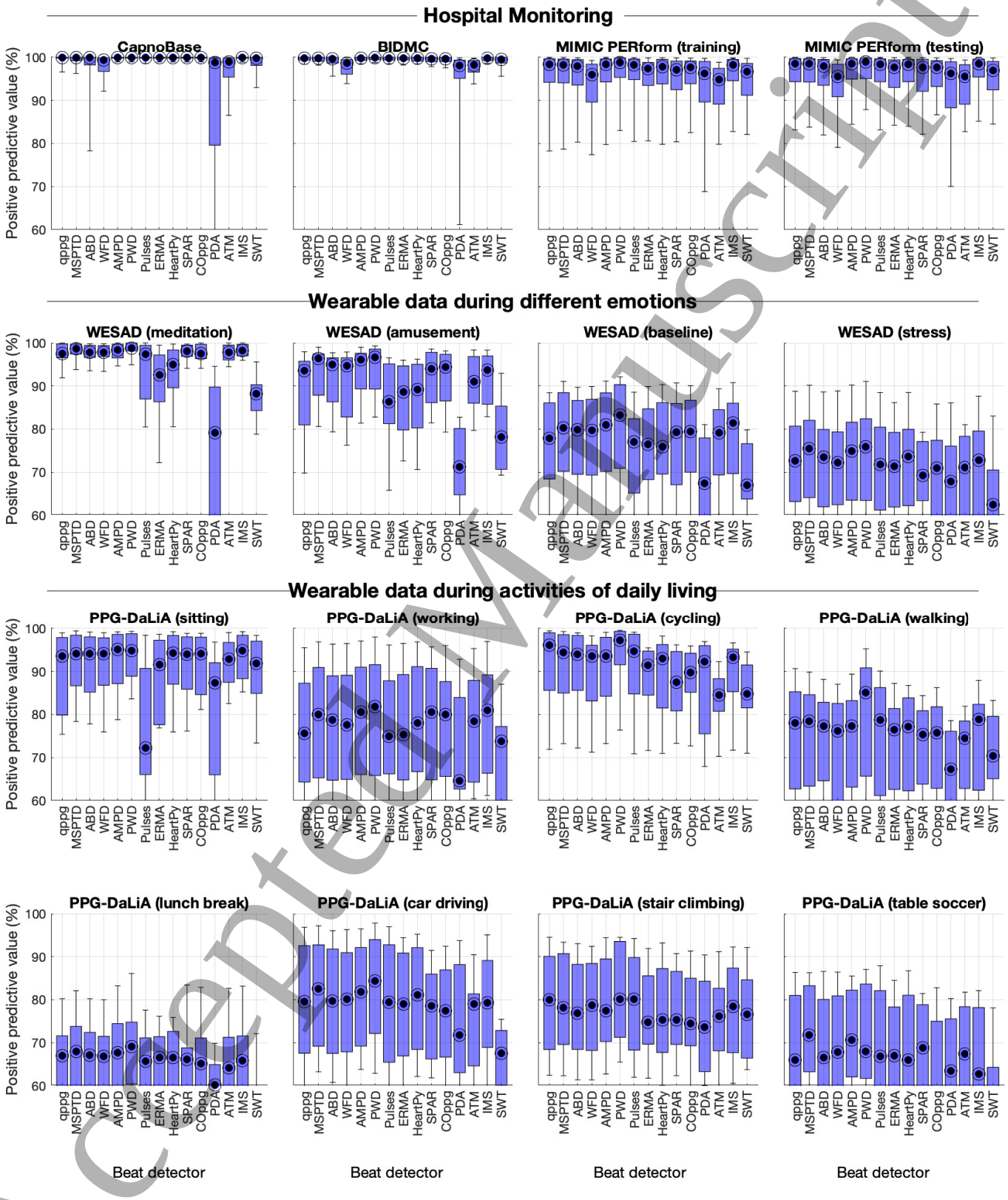

Fig. A2. Box plots showing the performance of beat detectors, expressed as the positive predictive value. Each graph shows the results for each of the beat detectors on a particular dataset. Performance is shown as the median (circles), inter-quartile range (boxes), and 10^{th} and 90^{th} percentiles (whiskers) across subjects. *See Table 2in the main text for definitions of beat detectors.*

APPENDIX II
ASSOCIATION BETWEEN PPG BEAT DETECTOR PERFORMANCE AND PATIENT DEMOGRAPHICS AND PHYSIOLOGY

Associations between PPG beat detector performance and patient demographics and physiology were presented in Fig. 5 in the main text, using the F_1 score to describe performance. Additional results are shown in: Fig. A3, which shows the sensitivity of beat detectors; and Fig. A4, which shows their positive predictive value.

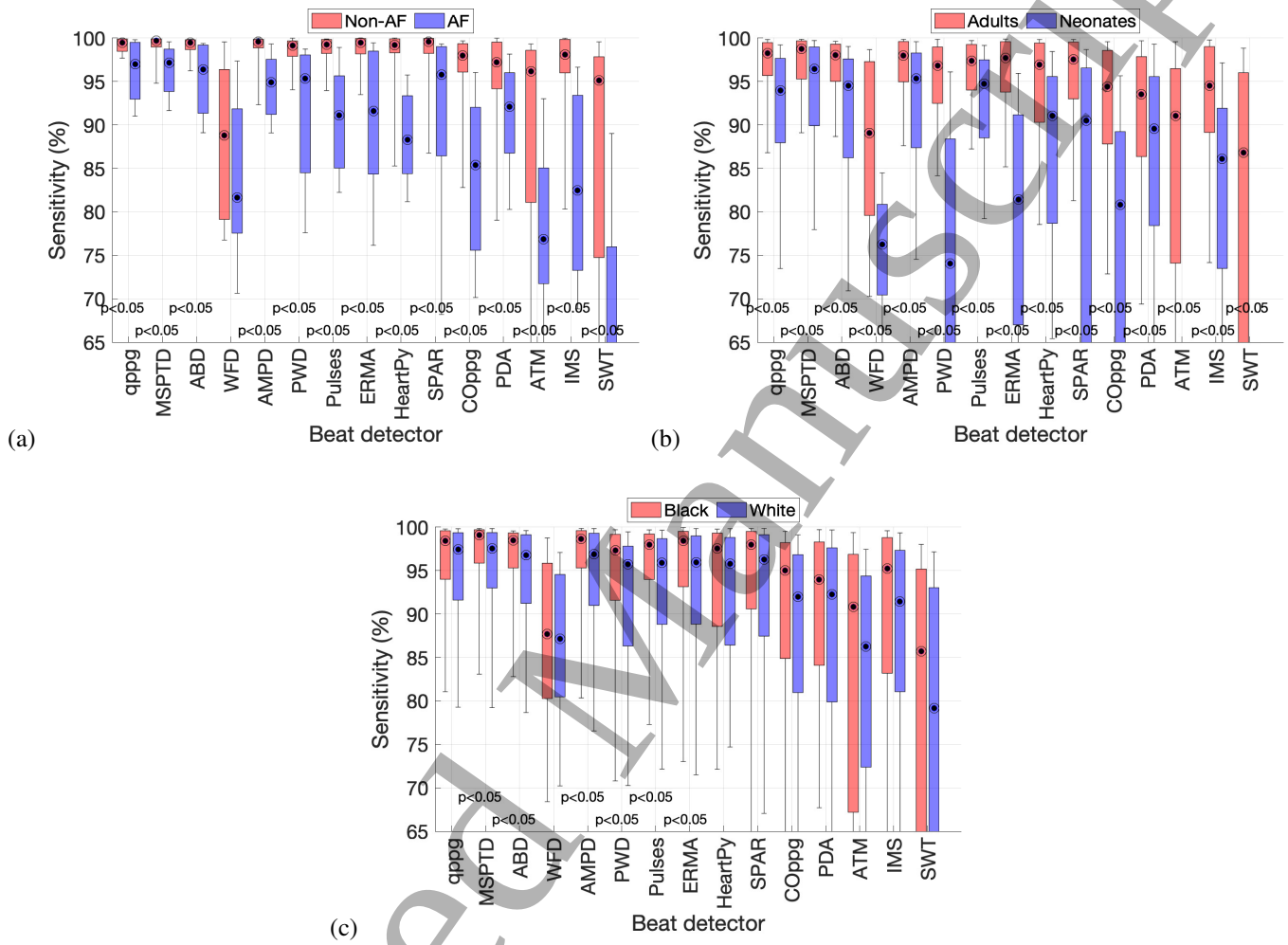

Fig. A3. Box plots showing the associations between beat detector performance and patient physiology and demographics, expressed as the sensitivity. Each graph shows the results for each of the beat detectors on a particular dataset. Performance is shown as the median (circles), inter-quartile range (boxes), and $10^{\mathbf{th}}$ and $90^{\mathbf{th}}$ percentiles (whiskers) across subjects. *See Table 2 in the main text for definitions of beat detectors.*

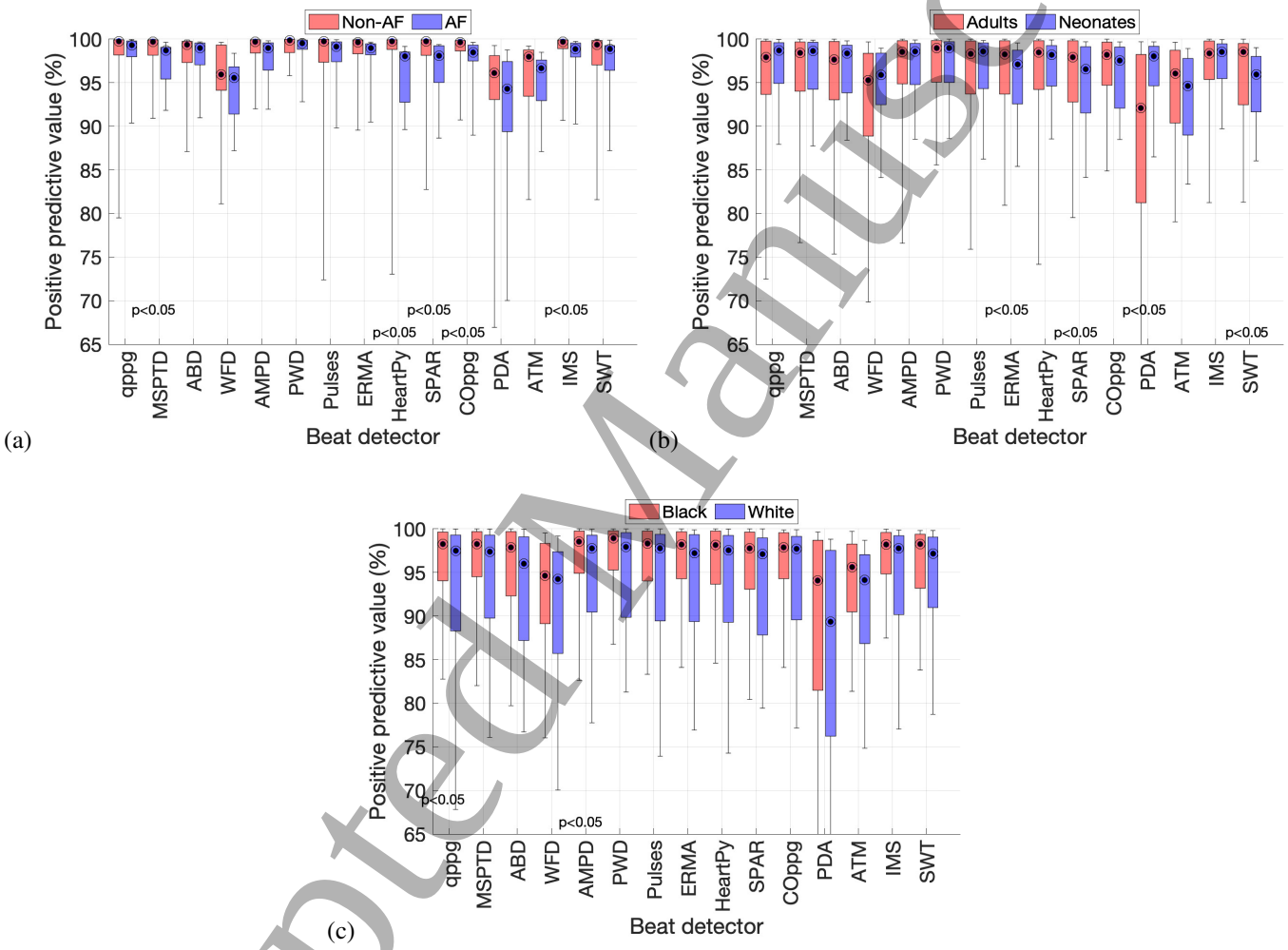

Fig. A4. Box plots showing the associations between beat detector performance and patient physiology and demographics, expressed as the positive predictive value. Each graph shows the results for each of the beat detectors on a particular dataset. Performance is shown as the median (circles), inter-quartile range (boxes), and $10^{\mathbf{th}}$ and $90^{\mathbf{th}}$ percentiles (whiskers) across subjects. *See Table 2 in the main text for definitions of beat detectors.*

References

[1] P. H. Charlton and V. Marozas, "Wearable photoplethysmography devices," in *Photoplethysmography*, 1st ed., P. Kyriacou and J. Allen, Eds. Elsevier, 2022, ch. 12, pp. 401–439. https://doi.org/10.1016/B978-0-12-823374-0.00011-6

[2] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological Measurement*, vol. 28, no. 3, pp. R1–R39, 2007. https://doi.org/10.1088/0967-3334/28/3/R01

[3] P. H. Charlton *et al.*, "Wearable Photoplethysmography for Cardiovascular Monitoring," *Proceedings of the IEEE*, vol. 110, no. 3, pp. 355–381, 2022. https://doi.org/10.1109/JPROC.2022.3149785

[4] M. V. Perez *et al.*, "Large-scale assessment of a smartwatch to identify atrial fibrillation," *New England Journal of Medicine*, vol. 381, no. 20, pp. 1909–1917, 2019. https://doi.org/10.1056/NEJMoa1901183

[5] M. W. Sjoding *et al.*, "Racial Bias in Pulse Oximetry Measurement," *New England Journal of Medicine*, vol. 383, no. 25, pp. 2477–2478, 2020. https://doi.org/10.1056/nejmc2029240

[6] S. Fleming *et al.*, "Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies," *Lancet*, vol. 377, no. 9770, pp. 1011–8, 2011. https://doi.org/10.1016/S0140-6736(10)62226-X

[7] K. Kotzen *et al.*, "Benchmarking Photoplethysmography Peak Detection Algorithms Using the Electrocardiogram Signal as a Reference," in *Proc CinC*. IEEE, 2021, pp. 1–4. https://doi.org/10.23919/CinC53138.2021.9662889

[8] W. Karlen *et al.*, "Multiparameter respiratory rate estimation from the photoplethysmogram," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 7, pp. 1946–53, 2013. https://doi.org/10.1109/TBME.2013.2246160

[9] M. A. F. Pimentel *et al.*, "Toward a robust estimation of respiratory rate from pulse oximeters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 8, pp. 1914–1923, 2017. https://doi.org/10.1109/TBME.2016.2613124

[10] S. K. Bashar *et al.*, "Noise detection in electrocardiogram signals for intensive care unit patients," *IEEE Access*, vol. 7, pp. 88 357–88 368, 2019. https://doi.org/10.1109/ACCESS.2019.2926199

[11] S. K. Bashar, "Atrial Fibrillation annotations of electrocardiogram from MIMIC III matched subset," 2020. https://doi.org/10.6084/m9.figshare.12149091.v1

[12] P. Schmidt *et al.*, "Introducing WeSAD, a multimodal dataset for wearable stress and affect detection," in *Proc. ICMI*, 2018, pp. 400–408. https://doi.org/10.1145/3242969.3242985

[13] A. Reiss *et al.*, "Deep PPG: large-scale heart rate estimation with convolutional neural networks," *Sensors*, vol. 19, no. 14, p. 3079, 2019. https://doi.org/10.3390/s19143079

[14] M. Saeed *et al.*, "Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database," *Critical Care Medicine*, vol. 39, no. 5, pp. 952–60, 2011. https://doi.org/10.1097/CCM.0b013e31820a92c6

[15] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. E215–20, 2000. https://doi.org/10.1161/01.CIR.101.23.e215

[16] A. E. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," *Scientific Data*, vol. 3, p. 160035, 2016. https://doi.org/10.1038/sdata.2016.35

[17] B. Moody *et al.*, "MIMIC-III Waveform Database Matched Subset (version 1.0)," 2020. https://doi.org/10.13026/c2294b

[18] P. H. Charlton, "MIMIC PERform Datasets," 2022. https://doi.org/10.5281/zenodo.6807403

[19] P. H. Charlton, "ppg-beats: algorithms to detect heartbeats in photoplethysmogram (PPG) signals," 2022. https://doi.org/10.5281/zenodo.6037646

[20] M. Aboy *et al.*, "An automatic beat detection algorithm for pressure signals," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 10, pp. 1662–1670, 2005. https://doi.org/10.1109/TBME.2005.855725

[21] F. Scholkmann, J. Boss, and M. Wolf, "An efficient algorithm for automatic peak detection in noisy periodic and quasi-periodic signals," *Algorithms*, vol. 5, no. 4, pp. 588–603, 2012. https://doi.org/10.3390/a5040588

[22] H. S. Shin, C. Lee, and M. Lee, "Adaptive threshold method for the peak detection of photoplethysmographic waveform." *Computers in biology and medicine*, vol. 39, no. 12, pp. 1145–52, 2009. https://doi.org/10.1016/j.compbiomed.2009.10.006

[23] D. Han *et al.*, "A Real-Time PPG Peak Detection Method for Accurate Determination of Heart Rate during Sinus Rhythm and Cardiac Arrhythmia," *Biosensors*, vol. 12, no. 2, p. 82, 2022. https://doi.org/10.3390/bios12020082

[24] C. Orphanidou *et al.*, "Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 832–8, 2015. https://doi.org/10.1109/JBHI.2014.2338351

[25] M. Elgendi *et al.*, "Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions," *PLoS ONE*, vol. 8, no. 10, pp. 1–11, 2013. https://doi.org/10.1371/journal.pone.0076585

[26] P. van Gent *et al.*, "Analysing noisy driver physiology real-time using off-the-shelf sensors: Heart rate analysis software from the taking the fast lane project," *Journal of Open Research Software*, vol. 7, no. 1, 2019. https://doi.org/10.5334/jors.241

[27] P. van Gent *et al.*, "HeartPy: A novel heart rate algorithm for the analysis of noisy signals," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 66, pp. 368–378, 2019. https://doi.org/10.1016/j.trf.2019.09.015

[28] W. Karlen, J. M. Ansermino, and G. Dumont, "Adaptive pulse segmentation and artifact detection in photoplethysmography for mobile applications," in *Proc. IEEE EMBS*. IEEE, 2012, pp. 3131–4. https://doi.org/10.1109/EMBC.2012.6346628

[29] S. M. Bishop and A. Ercole, "Multi-scale peak and trough detection optimised for periodic and quasi-periodic neuroscience data," in *Intracranial Pressure and Neuromonitoring XVI. Acta Neurochirurgica Supplement*, T. Heldt, Ed. Springer, 2018, vol. 126, pp. 189–195. https://doi.org/10.1007/978-3-319-65798-1_39

[30] E. J. Argüello Prada and R. D. Serna Maldonado, "A novel and low-complexity peak detection algorithm for heart rate estimation from low-amplitude photoplethysmographic (PPG) signals," *Journal of Medical Engineering and Technology*, vol. 42, no. 8, pp. 569–577, 2018. https://doi.org/10.1080/03091902.2019.1572237

[31] B. N. Li, M. C. Dong, and M. I. Vai, "On an automatic delineator for arterial blood pressure waveforms," *Biomedical Signal Processing and Control*, vol. 5, no. 1, pp. 76–81, 2010. https://doi.org/10.1016/j.bspc.2009.06.002

[32] J. Lázaro *et al.*, "Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children," *IEEE Journal of Biomedical and Health Informatics*, vol. 18, no. 1, pp. 240–246, 2014. https://doi.org/10.1109/JBHI.2013.2267096

[33] A. N. Vest *et al.*, "An open source benchmarked toolbox for cardiovascular waveform and interval analysis," *Physiological Measurement*, vol. 39, no. 10, 2018. https://doi.org/10.1088/1361-6579/aae021

[34] C. Pettit and P. Aston, "Photoplethysmogram (PPG) Beat Detection Using Symmetric Projection Attractor Reconstruction," *[in preparation]*.

[35] P. J. Aston *et al.*, "Beyond HRV: attractor reconstruction using the entire cardiovascular waveform data for novel feature extraction," *Physiological Measurement*, vol. 39, no. 2, p. 024001, 2018. https://doi.org/10.1088/1361-6579/aaa93d

[36] J. V. Lyle and P. J. Aston, "Symmetric projection attractor reconstruction: Embedding in higher dimensions," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 31, no. 11, p. 113135, nov 2021. https://doi.org/10.1063/5.0064450

[37] S. Vadrevu and M. Sabarimalai Manikandan, "A robust pulse onset and peak detection method for automated PPG signal analysis system," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 3, pp. 807–817, 2019. https://doi.org/10.1109/TIM.2018.2857878

[38] N. J. Conn and D. A. Borkholder, "Wavelet based photoplethysmogram foot delineation for heart rate variability applications," in *IEEE Signal Processing in Medicine and Biology Symposium*. IEEE, 2013. https://doi.org/10.1109/SPMB.2013.6736782

[39] E. Peralta *et al.*, "Optimal fiducial points for pulse rate variability analysis from forehead and finger photoplethysmographic signals," *Physiological Measurement*, vol. 40, no. 2, 2019. https://doi.org/10.1088/1361-6579/ab009b

[40] J. Behar *et al.*, "A comparison of single channel fetal ecg extraction methods," *Annals of Biomedical Engineering*, vol. 42, no. 6, pp. 1340–1353, 2014. https://doi.org/10.1007/s10439-014-0993-9

[41] A. E. Johnson *et al.*, "R-peak estimation using multimodal lead switching," in *Proc CinC*, vol. 41, 2014, pp. 281–284. https://ieeexplore.ieee.org/document/7043034

[42] G. Clifford, "rpeakdetect.m," Available at: http://www.mit.edu/∼gari/CODE/ECGtools/ecgBag/rpeakdetect.m. http://www.mit.edu/$\sim$gari/CODE/ECGtools/ecgBag/rpeakdetect.m

[43] ANSI/AAMI, "Cardiac monitors, heart rate meters, and alarms," Arlington, 2002.

[44] Consumer Technology Association, "Physical Activity Monitoring for Heart Rate, ANSI/CTA 2065," 2018. https://shop.cta.tech/collections/standards/products/physical-activity-monitoring-for-heart-rate

[45] J. M. Mühlen *et al.*, "Recommendations for determining the validity of consumer wearable heart rate devices: expert statement and checklist of the INTERLIVE Network," *British Journal of Sports Medicine*, vol. 55, no. 14, pp. 767–779, 2021. https://doi.org/10.1136/bjsports-2020-103148