

# An Open-Source Biosignal-Specific Physiological Feature Extraction Tool for Machine Learning and Pattern Recognition

Mohsen Nabian<sup>1</sup>, Yu Yin<sup>1</sup>, Zulqarnain Khan<sup>1</sup>, Jolie Wormwood<sup>2</sup>, Karen S. Quigley<sup>2</sup>, Lisa F. Barrett<sup>2</sup>, and Sarah Ostadabbas<sup>1\*</sup>

**Abstract**—Electrocardiogram (ECG), Electrodermal Activity (EDA), Electromyogram (EMG), continuous Blood Pressure (BP) and Impedance Cardiography (ICG) are among the physiological signals widely used in various biomedical applications including health tracking, sleep quality assessment, early disease detection/diagnosis, telemedicine and human affective state recognition. This paper presents the development of a biosignal-specific tool for processing and feature extraction of these physiological signals according to state-of-the-art studies reported in the scientific literature and feedback received from field experts. This tool is intended to assist researchers in affective computing, machine learning, and pattern recognition to extract the physiological features from these biosignals automatically and reliably. In this paper, we provide algorithms for signal-specific quality checking, filtering, and segmentation as well as the extraction of features that have been shown to be highly relevant to category discrimination in biomedical and affective computing applications. This tool is an open-source software written in MATLAB and a graphical user interface (GUI) is also provided for the convenience of the users. The GUI is compatible with MathWorks Classification Learner app for further classification purposes such as model training, cross-validation scheme farming, and classification result computation.

**Index Terms**—Biosignal Processing, Electrocardiogram (ECG), Electrodermal Activity (EDA), Electromyography (EMG), Impedance Cardiography (ICG), Blood Pressure (BP), Affective Computing, Machine Learning, Pattern Recognition, Health Informatics, Feature Extraction, Quality Checking, Dimensionality Reduction.

## 1 INTRODUCTION

Measures of physiological signals (biosignals), including the Electrocardiogram (ECG), Impedance Cardiogram (ICG), Electromyogram (EMG) and measures of Electrodermal Activity (EDA) and continuous noninvasive arterial pressure (continuous BP) can provide rich information about physiological functioning of the body and are thought to provide insight on diverse phenomena, such as sleep quality and human affective state. These measures are thus widely used as input data for machine learning and pattern recognition models focused on preventive care, diagnostics, telemedicine and for guiding therapy [1]–[8]. For example, a support-vector-machine model (SVM) was trained on ECG signal features, such as Inter-beat Interval (IBI), to predict Ischaemic heart disease [2], and an SVM model was trained on EDA signal features to recognize the ease of engagement in children during social interactions [3]. A model trained on EMG signal features was used to aid the diagnosis of neuromuscular disorders [4], and a model trained on BP signal features was used to aid in the diagnosis of Diabetes [5]. Valvular heart diseases have been predicted by models using ICG signal features [6] and physiological responding under stress has been predicted by multi-modal data comprising ECG, EMG, EDA, and ICG signal features [7]. These are just a few examples of the

rapidly expanding research utilizing biosignals in the area of medical informatics [9].

In each study, a large amount of physiological data are usually collected. However, to effectively reduce the dimensionality of these data, researchers often extract specific features from each biosignal to utilize as input data for machine learning or pattern recognition models. However, to extract these features, a significant amount of knowledge is required since distinct methods for noise and artifact removal, segmentation, and characteristic point detection are required for each signal modality [10]. Due to the rising interest and usability of these biosignals in medical and bioengineering research, several commercial software products and open-source toolboxes have been developed to ease the processing and feature extraction requirements for these signals. For example, several biosignal specific processing algorithms have been published for ECG [11], [12], [12], [13], EMG [14], [15], EDA [16]–[18], continuous blood pressure (BP) [19], [20] and ICG [21], [22]. However, a large portion of these software tools are not designed for a machine learning processing pipeline, and therefore are manually intensive. For example, many demand visual inspections because they are not sufficiently robust to morphological variations of the biosignals and are therefore unreliable without manually-intensive visual inspection. Also, in the majority of these tools, basic quality checking of these biosignals is not automated. Moreover, although ICG signal features, such as cardiac output (CO), pre-ejection period (PEP), and stroke volume (SV) are associated with human psychosocial states, such as threat and challenge responses to stress [23]–[27],

• <sup>1</sup>Augmented Cognition Lab (ACLab), Electrical and Computer Engineering Department, Northeastern University, MA, USA. \*Corresponding Author's email: Ostadabbas@ece.neu.edu.

• <sup>2</sup>Department of Psychology, Northeastern University, MA, USA.

This work is partially supported by funding from MathWorks.

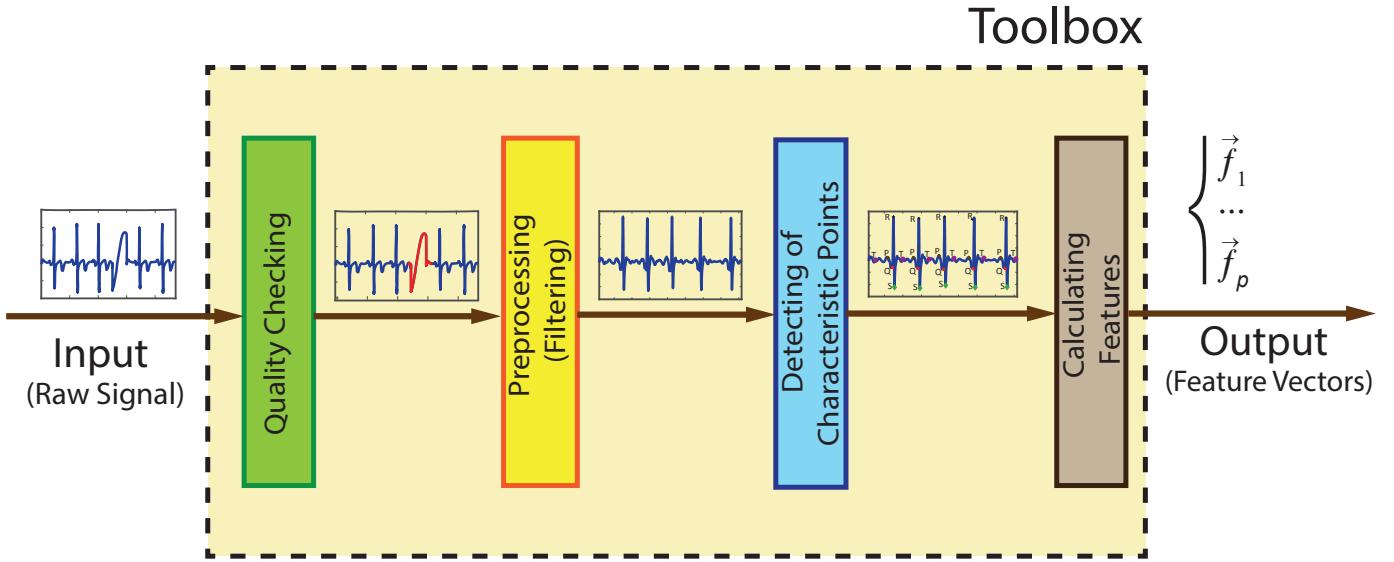


Fig. 1. Toolbox processing pipeline. 1) The Input (left) of the toolbox is a raw segment of either ECG, EDA, EMG, BP or ICG Biosignal modality. The Toolbox Output (right) is the signal-specific feature vector. Inside the toolbox comprises of the following consecutive processing steps : A) Basic quality checking of raw signals. Bad quality portions of the input signal are detected and marked as red for further elimination or replacement. B) Preprocessing (filtering) of raw signals for noise or artifact removal. C) Characteristic point detection on the filtered signal based on the signal-specific robust algorithms. D) Calculating features as functions of characteristics points detected on the signal.

very few software/toolboxes have ICG processing tools along with other biosignals [28].

The present work provides an open-source biosignal-specific tool for psychologists, neuroscientists and researchers in affective computing, machine learning and pattern recognition to process ECG, EMG, EDA, ICG and continuous BP biosignals all in one easy-to-use, MATLAB based toolbox that extracts relevant features for each signal automatically and reliably. The biosignal processing pipeline employed in our toolbox includes applying signal-specific algorithms for (1) checking the quality of collected physiological data, (2) the pre-processing of biosignal data (e.g., noise filtering), (3) the segmentation of continuous biosignal data, (4) the detection of characteristic points on a biosignal waveform, and (5) feature extraction. The signal-specific algorithms employed in our toolbox are developed according to state-of-the-art studies reported in the scientific literature [11]–[22], [29] and are based on feedback received from field experts.

The present paper also further extends the knowledge of readers and researchers about the basic physiological implications of these biosignal modalities, signal-specific characteristic points, the informative signal-specific features commonly used in medical informatics as well as some of the state-of-the-art signal-specific algorithms in literature that are developed for these purposes (processing and feature extraction). Based upon the provided algorithms in this work, researchers are also able to further extend it to extract other biosignal-specific features of their own interest for their intended applications.

It is noted that this paper is a complete version of the previously published work [30], and the toolbox developed in this work has been widely used in some affective computing studies for processing and feature extraction of biosignals [7]. Moreover, a python version of this toolbox is currently

in development and will be available soon. The developed toolbox is accessible via the MathWorks File Exchange site [31]. The toolbox was developed such that the outputted feature matrix could be directly used by machine learning algorithms for further analysis, or even directly passed as input into the Classification Learner App provided by MATLAB.

## 2 MATERIALS AND METHODS

We developed a series of signal-specific algorithms to extract physiologically-related features from raw biosignals of ECG, EDA, EMG, ICG and continuous BP. To this end, for each biosignal, several layers of algorithms are applied in consecutive order. The input-output as well as inner toolbox layers are schematically illustrated in Figure 1. Although, the algorithms are signal-specific, the general approaches and processes are similar across all five signal modalities: First, we perform basic automatic quality checking. Chunks of the raw signal that are purely noise or are of low quality due to improper attachment of sensors or large movements of subjects are detected and marked in red, identifying them for further replacement or removal from the original signal. Signals then undergo a filtering process designed to remove high frequency noise (e.g., 60Hz electrical noise) and smaller artifacts. Next, signal-specific algorithms are used to detect characteristic points of each signal's waveform, and finally, relevant physiological features are calculated and outputted as a feature matrix for further machine learning and pattern recognition analysis. The list of characteristic points as well as features for each biosignal are given in Table 1. Since this paper is primarily focused on the biosignal-specific feature extraction algorithms, we first explained the filtering, characteristic point detection, and feature extraction algorithms for each bio-signal, and then we elaborated on

the algorithms used for the quality checking and briefly introduced the graphical user interface (GUI).

## 2.1 Datasets Description

In this work, a dataset from the Interdisciplinary Affective Science Laboratory in the Psychology department at Northeastern University was used for algorithm testing and verification. The data was recorded as part of a larger research project funded by the Army Research Institute for the Behavioral and Social Sciences (W5J9CQ-12-C-0049 to L.F.B.). Data from 100 participants (40 males and 60 females) was selected from the larger dataset of 260 participants. All of the physiological measures included in our toolbox, including ECG, EMG, EDA, ICG and continuous BP, were recorded from each subject while ..... All biosignals were recorded using BioLab v.3.0.13 (Mindware Technologies; Gahanna, OH), and were acquired on a BioNex 8-Slot Chassis with a sampling rate of 1000Hz.

We included samples of this dataset in the toolbox directory for tutorial purposes. For further algorithm testing and verification, we also employed datasets from the PhysioBank ATM website [32] as well as some datasets shared by the MIT Affective Computing group [33].

## 2.2 Biosignal #1: Electrocardiogram (ECG)

ECG is the main diagnostic approach for detecting cardiovascular diseases [34] and contains rich information relevant to human health, sleep quality, and emotional states [11]. The ECG signal comprises a consistent, repeating complex waveform created by each heartbeat, which contains several critical sub-waves including a P wave, a QRS complex, and a T wave (as shown in Fig. 2). By automatically detecting these characteristics points of the ECG signal for each heartbeat, we can extract relevant physiological features. We first used filtering to remove noise and artifacts. Then, influenced by algorithms proposed by Pan-Tompkins [12] for QRS detection, a robust algorithm was developed and used to detect the characteristic points of the ECG waveform (P, Q, R, S, T points). Finally, several physiological features were then calculated as functions of the detected signal characteristic points. Detailed explanations for each step, are provided in the following sections.

### 2.2.1 ECG Preprocessing

Corruption of the ECG signal may be caused by power line interference or baseline wandering (BW), which may be caused by respiration or large movements of the subjects or the instruments [13], [35], [36]. Therefore, preprocessing of these signals prior to applying any detection algorithm is necessary. With a thorough analysis of the signal power spectral density (PSD) of measured ECG signals in our dataset, we found that a band-pass filter, such as an Elliptic, Gaussian or Butterworth filter, provided satisfactory performance in eliminating baseline and high-frequency noise without significantly decreasing the amplitude of the QRS complexes. Although we used a band-pass Elliptic filter by default, users are able to choose other types of filters in the developed GUI, including Gaussian and Butterworth bandpass filters.

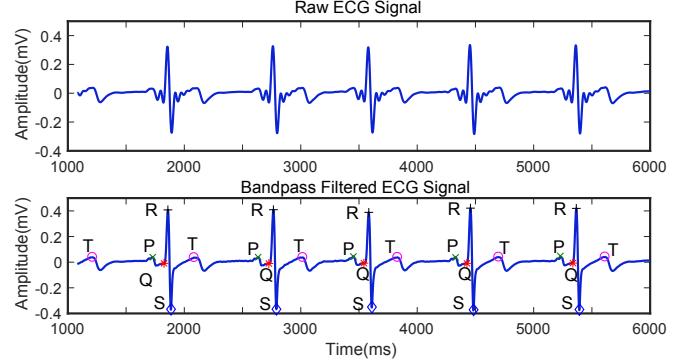


Fig. 2. Raw ECG signal (top); bandpass filtered ECG signal (bottom) with the characteristic points (P,Q,R,S and T) detected on the signal.

### 2.2.2 ECG Characteristic Points Detection

Inspired by the Pan-Tompkins algorithm [12], we proposed an algorithm for robust QRS detection on the ECG signal. The steps are as follows:

**Step i:** Given the fact that the human heart-rate does not exceed 5 beats per second [37], [38], two adjacent QRS complexes cannot appear more closely than 200 ms [12]. Therefore, a sliding window with the size of 400 ms and step size of 1 sampling rate iteratively scans the entire signal. At each iteration, we find the global maximum in the window signal and, providing this global maximum took place exactly in the middle of the window, we mark that global maximum as a potential R point. By scanning the entire input ECG signal, we have selected the peaks which are likely to be the R peaks. The  $i$ th peak is denoted by  $R_i$ .

**Step ii :** We eliminate the R peaks detected in *step i* which are lower in amplitude than the Amplitude Threshold (Amp-Thr) which is initially set to 1/3 times the maximum amplitude of the first 2 seconds of the signal. In order to adapt to the changes in the ECG over time, upon passing sufficient R peaks from the beginning of the signal, Amp-Thr is updated to 0.75 times the mean of the amplitude of the last eight determined R peaks.

**Step iii :** We scan through all the detected adjacent R peak to R peak time intervals (R-R intervals). If any R-R interval is greater than a threshold named RR-Thr, it implies that an R peak is missing. Thus, we assign a new R peak as the global maximum amplitude of the signal within the range of  $R_i+200$  and  $R_{i+1}-200$  msec [12]. According to [12], the rate of change for adjacent RR intervals will not exceed 166%. Thus, RR-Thr is initially set to the previous R-R interval multiplied by 1.66. Upon detecting enough R-R intervals, however, the RR-Thr is updated to reflect the average of the last eight R-R intervals multiplied by 1.66.

**Step iv:** Detection of Q, S, P, and T points were then based on the location of the detected R peaks.  $Q_i$  is identified as the first closest local minimum before  $R_i$  and within the interval of  $R_i-70$  msec and  $R_i$ . Likewise,  $S_i$  is identified as the first closest local minimum after  $R_i$  and within the interval of  $R_i$  and  $R_i+70$  msec. The peak of the P wave,  $P_i$ , is identified as the global maximum point found in the interval of  $R_i-120$  msec to  $R_i$ , and the peak of the T wave,  $T_i$ , is identified as the global maximum point found in the interval of  $R_i$  to  $R_i+300$  msec. The detected locations of P, Q, R, S, and T on a few sample heartbeats

TABLE 1

List of biosignals that analyzed in this paper, their extracted characteristic points as well as calculated morphological features.

Signal	Characteristics	Features
ECG	P, Q, R, S, T	RR interval, SD, SDSD, RMSSD, NN50, pNN50, EDR, QR to QS ratio and RS to QS ratio
EDA	SCRs	SCR duration mean, SCR amplitude mean, SCR rise-time mean, Signal mean, Number of detected SCRs
EMG	-	MAV, Zero crossing count, Slope Sign Change, Waveform Length, Log Detector, SD, RMS
ICG	Q-point, B-point, X-point, (dZ/dt)max	PEP, LVET, SV, CO, TPR
BP	Foot, Systole, Notch, Dicrotic Peak	Diastolic pressure, Systolic pressure, Dicrotic Pressure, Notch pressure, Mean arterial pressure (MAP)

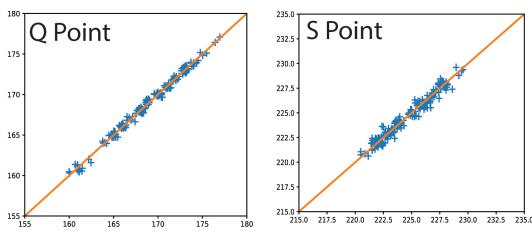


Fig. 3. (left) Validation of Q-point detection, visual detection (x-axis) vs. algorithm detection (y-axis), (right) Validation of S-point detection, visual detection (x-axis) vs. algorithm detection (y-axis).

are shown in Fig. 2. We investigated the accuracy of our Q and S point detection algorithm by comparing the results of our automatic detection algorithm with visual inspections on 1000 randomly selected heartbeat ECG cycles across 10 different subjects (Fig. 3). Results showed that almost all the Q and S points were accurately detected. Slight differences seen in Fig. 3 were likely caused by the lower resolution in the visual readings for the manual inspection.

### 2.2.3 ECG Feature Extraction

The following morphological features were extracted based on the QRS detection:

**IBIM** is the mean of all Inter-beat Intervals (IBIs) in a segment, where each IBI is the time (msec) between adjacent RR intervals [14].

**IBISD** is the standard deviation of all IBIs in the input signal segment [14].

**SDSD** is defined as the standard deviation of the differences between adjacent IBIs in a signal segment [14]. SDSD may be found by the following equation:

$$SDSD = 1/N \sqrt{\sum_{i=2}^N (RR_i - RR_{i-1})^2} \quad (1)$$

where  $RR_i$  is the  $i$ th RR interval,  $\mu = (\sum_{i=2}^N (RR_i - RR_{i-1})) / (N - 1)$  and  $N$  = number of RR intervals.

**RMSSD** is the square root of the mean of the sum of the squares of the differences between adjacent IBIs in a segment [15].

$$RMSSD = \sqrt{\frac{1}{N} \sum_{i=2}^N (RR_i - RR_{i-1})^2} \quad (2)$$

**NN50** is defined as the number of pairs of adjacent IBIs where the first IBI exceeds the second IBI by more than 50ms [14].

**pNN50** is defined as the number of pairs of adjacent IBIs where the second IBI exceeds the first IBI by more than 50ms [14].

**ECG Derived Respiratory (EDR) mean and standard deviation** [14], [39], [40] are related to the respiratory efforts. EDR is calculated based on the area of each normal QRS complex measured over a fixed window width  $w$ , which is determined by the interval from the PQ junction to the J-point (junction point) [41]. We estimated  $w$  as two times the sum of the QR interval and the RS interval ( $w = 2 \times (QR + RS)$ ). The mean and standard deviation of EDR are output as two features.

**QR to QS ratio and RS to QS ratio** are defined as the ratio of the QR interval to the QS interval, and the ratio of the RS interval to the QS interval, respectively, for each R peak.

### 2.3 Biosignal #2: Electrodermal Activity (EDA)

EDA reflects momentary or long term changes in the skin's electrical conductivity due to various internal and external stimuli [42]. Changes in EDA are primarily caused by changes in the activity of the eccrine sweat glands, which are exclusively innervated by the sympathetic nervous system (SNS). Because of this relationship between EDA activity and SNS activity, EDA has been widely studied in the emotion detection research as an indirect measure of arousal [43]. The EDA signal is composed of two activities, tonic and phasic. The slowly varying base signal is the tonic part, also called the skin conductance level (SCL). The faster-changing part is called phasic activity or skin conductance responses (SCRs). SCRs are related to more acute exterior stimuli or non-specific activation [16].

### 2.3.1 EDA Preprocessing

To eliminate noise carried by the EDA signal, a filtering step is required prior to implementing an SCR detection algorithm. The filtering options provided for EDA in our GUI are the same as those described above for the ECG signal. In some studies, a Gaussian filter with a low-pass frequency of 1Hz has been recommended and used for the filtering the EDA signal [17].

### 2.3.2 EDA Characteristic Points Detection

SCRs on the EDA signal can be detected by performing differentiation and subsequent convolution with a 20-point Bartlett window [18], where a Bartlett window is a triangle function represented as:

$$w(n) = \begin{cases} \frac{2n}{N}, & 0 \leq n \leq N/2 \\ 2 - \frac{2n}{N}, & N/2 < n \leq N \end{cases} \quad (3)$$

where  $N$  is the window size (=20) and  $n$  is the  $n - th$  signal sample in the window.

The occurrence of the SCR is detected by finding two consecutive zero-crossings, from negative to positive and positive to negative of the bartlett differentiated EDA signal. We considered negative to positive as the beginning and positive to negative as the end of each SCR. The amplitude of the SCR is obtained by finding the maximum value between these two zero-crossings. Detected SCRs with amplitudes smaller than 10 percent of the maximum SCR amplitudes already detected on the differentiated signal are excluded [18]. Fig. 4 demonstrates the detection of the SCRs. On the top plot, the detected SCR's beginnings, ends, and peaks are shown on the differentiated signal. In the bottom plot, the corresponding locations of the previously detected SCRs are provided on the original low-pass filtered signal.

### 2.3.3 EDA Feature Extraction

After detecting the SCRs on the signal, the following features can be extracted:

**SCR duration mean** is the average of the detected SCRs' durations within a given input segment. Duration of an SCR is defined as the time (in ms) from the beginning to the end of an SCR.

**SCR amplitude mean** is the average of the amplitude of the detected SCRs within a given input segment.

**SCR rise-time mean** is the average of the SCR rise-times within a given input segment, which is defined as the time (in ms) between the beginning and the peak of an SCR.

**Mean Skin Conductance (MSC)** is the average of the low-pass filtered signal within a given input segment.

**Number of detected SCRs** is the number of the detected SCRs in the input segment.

**Tonic SCL** is the average value of the EDA signal within a given input segment, with data from the start to end of each SCR excluded.

If no SCRs are detected throughout the input segment, all the features except MSC and Tonic SCL will be output as Nan values.

### 2.4 Biosignal #3: Electromyography (EMG)

EMG measures changes in the electrical activity of motor neurons which directly innervate skeletal muscles, and is used to infer muscle activity [44]. EMG is sometimes used as a diagnostic procedure to determine the health of muscles and the nerve cells that control them (motor neurons) [45]. EMG signals are also widely used in emotion classification studies as a means of examining facial expressions [46].

#### 2.4.1 EMG Preprocessing

Filtering is a required step for reducing baseline and artifact noise in the EMG signal. The proper choice of EMG filtering depends on the specific muscles that are being measured. For example, different filtering options are used for facial muscles vs. larger muscles in the arms/legs. Moreover, the speed / duration of the anticipated response is also important to be considered. For example, different filters should be considered for studying a blink response vs. a response involving sustained muscle activation. Thus, various filtering options including the type of filter as well as cutoff frequencies are available to be chosen by the users. A band-pass Elliptic filter with the cutoff frequency of 10-300Hz [47] was provided as the default option.

#### 2.4.2 EMG Feature Extraction

We selected the feature options for the EMG based on the features introduced in [29].

**MAV** is the mean absolute value of the EMG signal in a time window analysis with  $N$  samples.  $x_k$  is the  $k$ th sample in this analysis window [29].

$$MAV = \frac{1}{N} \sum_{k=1}^N |x_k| \quad (4)$$

**Zero crossing count** is the number of times the EMG signal crosses zero within a time window analysis. It is a simple measure associated with the frequency of the signal. To avoid signal crossing counts due to low-level noise, a threshold  $\epsilon$  is included in the counting ( $\epsilon = 0.015V$ ) [29]. The zero crossing count is increased by one if:

$$|x_k - x_{k+1}| \geq \epsilon$$

and

$$(x_k > 0 \text{ and } x_{k+1} < 0) \text{ or } (x_k < 0 \text{ and } x_{k+1} > 0).$$

**Slope Sign Change** is related to the signal frequency and is defined as the number of times that the slope of the EMG waveform changes sign within an analysis window. A count threshold  $\epsilon$  is used to reduce noise-induced counts ( $\epsilon = 0.015V$ ) [29]. The slope sign count increases by one if

$$|x_k - x_{k+1}| \geq \epsilon \text{ or } |x_k - x_{k-1}| \geq \epsilon,$$

and

$$(x_k > x_{k-1} \text{ and } x_k > x_{k+1}) \text{ or } (x_k < x_{k-1} \text{ and } x_k < x_{k+1}).$$

**Waveform Length** provides a measure of the complexity of the signal. It is defined as the cumulative length of the EMG signal within the analysis window [29].

$$waveLen = \sum_{k=1}^N |\Delta x_k| \quad (5)$$

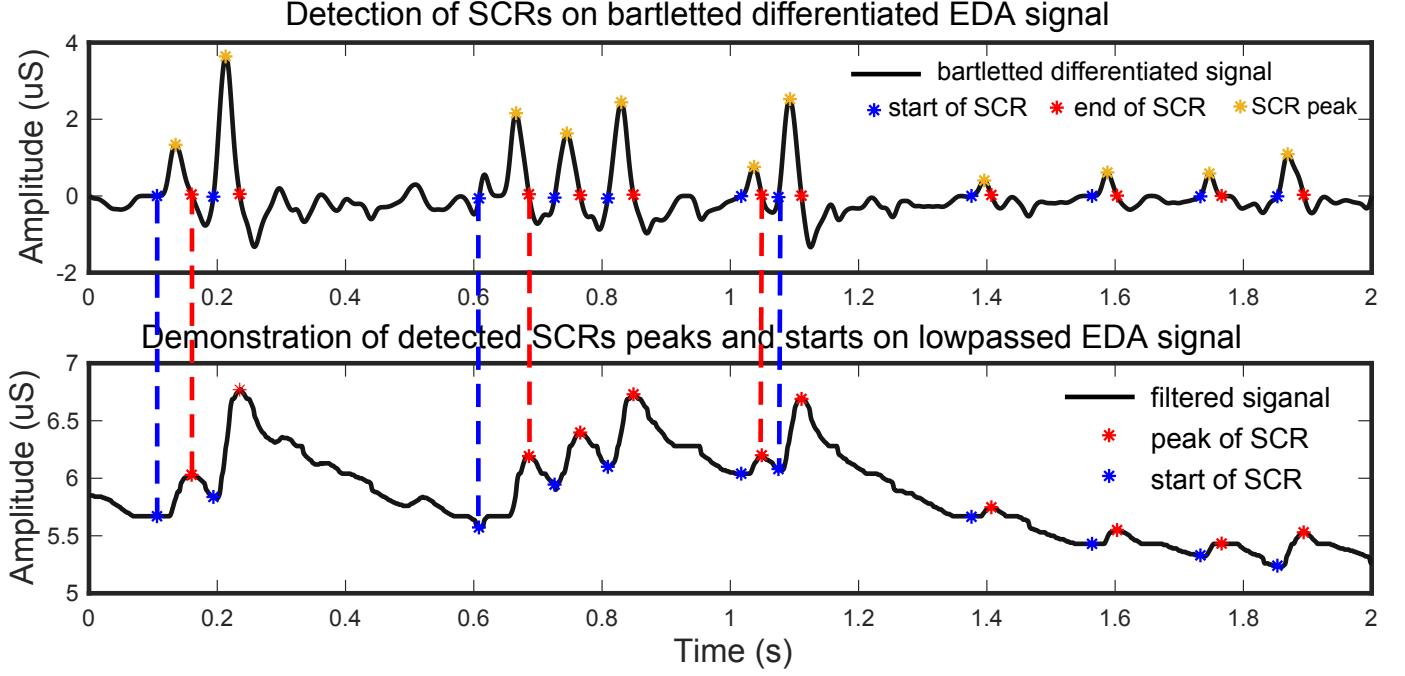


Fig. 4. Detection of SCRs on bartlett differentiated EDA signal (top) and demonstration of detected SCRs' peaks and starts on low-passed EDA signal (bottom).

where  $\Delta x_k = x_k - x_{k-1}$ .

**Log Detector** is an estimate of the exerted muscle force. The nonlinear detector is characterized as  $\log(|x_k|)$  and the log-Detect feature is defined as:

$$\text{logDetect} = e^{\frac{1}{N} \sum_{k=1}^N \log |x_k|} \quad (6)$$

**SD** is standard deviation of the amplitudes of the input EMG signal (after filtering).

**RMS** is the root mean square of the amplitudes of the input EMG signal (after filtering).

**Peak Amplitude** is the amplitude of the maximum of the input EMG signal.

**Peak Latency** is the time in which the maximum of the input EMG signal takes place.

## 2.5 Biosignal #4: Continuous Blood Pressure (BP)

Blood pressure (BP) in the circulatory system is often measured for monitoring of the cardiovascular system since it is related to the complex, combined effects of the force and rate of the heartbeat as well as the diameter and elasticity of the main arterial walls [48]–[50]. While some researchers rely on intermittent readings of systolic and diastolic blood pressure, recent technological advances are making it more and more common for researchers to rely instead on continuous noninvasive measures of BP. As such, automatic, robust and accurate extraction of beat-to-beat physiological features such as mean arterial pressure (MAP), systolic pressure, and diastolic pressure from continuous BP signals is crucial. The feature extraction process for continuous BP signals includes preprocessing (filtering), characteristic point detection, and feature calculation.

### 2.5.1 BP Preprocessing

The low-pass filter frequency for continuous BP should be at least 10 times higher than the heart rate (in Hz) and less than half of the sampling rate [51]. The typical heart-rate for humans varies between 40 bpm and 100 bpm [37], [38]. Therefore, a low-pass filter with a cutoff frequency of 40Hz is often used for artifact removal in BP signals [51]. However, in our software, other filter options, similar to those provided for ECG (e.g., XXXX), are also available for BP.

### 2.5.2 BP Characteristic Points Detection

An algorithm implemented by A. Laurin [19] for the feature extraction of arterial blood pressure (BP) is used in this work. This algorithm, inspired by the derivatives and thresholds described in Pan-Tompkins [12] and criteria described by Sun et al. [20], primarily detects the characteristic points of the BP signal. These points including Systolic, Diastolic (foot), Dicrotic Notch and Dicrotic Peak points are depicted in Fig. 5. The foot index is defined as the maximum point in the second derivative of the BP time-series in each interval. The Systolic point is defined as the maximum of the waveform following the foot index, relative to a window of radius 1/8 sec around itself [19]. The Dicrotic Notch is defined as the minimum of the subtraction of the signal and the straight line going from systole to diastole. Finally, the Dicrotic Peak is defined as the minimum of the second derivative of the time-series following the Dicrotic Notch, relative to a window of radius RR/5 sec around itself where RR is the median heartbeat interval computed from the foot points.

### 2.5.3 BP Feature Extraction

Based on the extraction of these characteristic peaks, the following five physiological features are extracted [19].

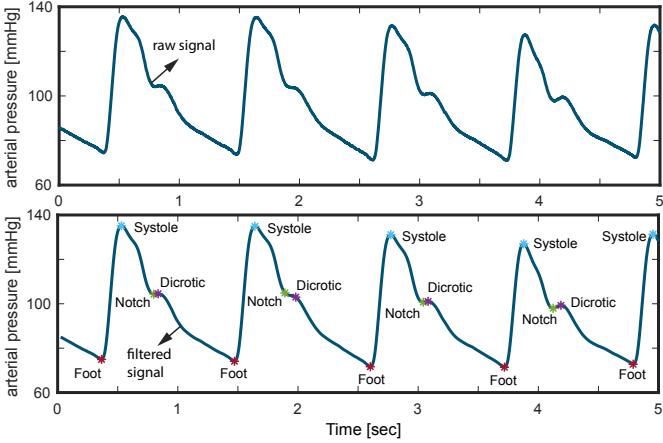


Fig. 5. Raw continuous blood pressure (BP) signal (Top), Corresponding low-pass filtered continuous BP signal (bottom) with the characteristic points (Foot,Systolic,Dicrotic Notch and Dicrotic Peak) detected on the signal.

**Mean Diastolic Pressure** (foot pressure) which is the mean of the all the signal values (pressures) at the Foot indexes.

**Mean Systolic Pressure** is the mean of the all the signal values (pressures) at the Systolic indexes.

**Mean Dicrotic Peak pressure** is the mean of the all the signal values (pressures) at the Dicrotic Peak indexes.

**Mean Dicrotic Notch pressure** is the mean of the all the signal values (pressures) at the Dicrotic Notch indexes.

**Mean arterial pressure (MAP)** represents mean signal value (in pressure) for the entire input BP signal.

**Total Peripheral Resistance (TPR)** is another important BP feature related to the resistance of the systemic circulation of the blood [52]. Because the calculation of TPR also requires features extracted from the ICG signal, the description of TPR is provided in the ICG section below.

## 2.6 Biosignal #5: Impedance Cardiography (ICG)

ICG signals include a measure of basal thoracic impedance ( $Z_0$ ) and its first derivative ( $dZ/dt$ ), though feature extraction typically relies mostly on the  $dZ/dt$  signal. ICG is a non-invasive method to extract important information (features) regarding cardiac activity. These features include: (1) left ventricular ejection time (LVET), (2) stroke volume (SV), (3) cardiac output (CO), (4) pre-ejection period (PEP), (5) total peripheral resistance (TPR) [52], and (6) mean  $Z_0$  [21]. LVET is the time period in which blood flows across the aortic valve, and PEP is the time period between the electrical activity signaling the start of ventricular contraction and the onset of blood being ejected into the aorta [53]. SV is defined as the blood volume pumped by the heart per heart beat, and CO is the blood volume pumped by the heart per minute [54]. Finally, TPR represents the resistance of the systemic circulation that carries oxygenated blood from the heart to the body, and returns deoxygenated blood to the heart. These five features can be calculated from the detected characteristic points on the ICG as well as some feature from ECG, synchronous transthoracic basal impedance ( $Z_0$ ) and blood pressure (BP) signals. The methods used to extract these features are based upon the discussions provided in published articles [21], [55].

### 2.6.1 ICG Preprocessing

The preprocessing of the ICG signal includes filtering, segmentation, and ensemble averaging. Each step, is explained in the following sections.

**Noise and artifact removal** is a necessary step to clean the ICG signal. A second order elliptic band-pass filter with the low cut-off frequency of 0.75Hz and high cut-off frequency of 40Hz is recommended in available Biosignal software, such as Biolab [56]. However, our work allows the users to adjust the low and high cut-off frequencies as well as to choose the filter type.

**Segmentation** of the ICG signal into beat-to-beat intervals is required for the ensemble averaging step. We considered a beat-to-beat interval to start at  $T_R - 250$  msec and to end at  $T_R + 500$  msec where  $T_R$  is the time in msec when an R peak is located on the corresponding ECG signal [21], [56].

**Ensemble averaging** of multiple cardiac cycles is performed to eliminate the stochastically distributed noise as well as respiratory influences and movement artifacts on the ICG signal [21], [55], [57]. The number of cycles is adjustable by the users and it is set to 8 beats by default.

### 2.6.2 ICG Characteristic Points Detection

In order to extract features from the ICG signal, the following characteristic points have to be identified on the ICG ( $dZ/dt$ ) and ECG signals:

(1) Q-point on the corresponding ECG signal, (2) B-point, (3)  $(dZ/dt)_{max}$ , and (4) X-point. In the following section, methods and algorithms to detect these points are discussed.

**Q-point** time on the ICG signal is identical to the Q-point time on the ECG signal (Fig. 6). Therefore, to extract the Q-point on ICG, the corresponding synchronous ECG signal is required.

**B-point** automatic identification comes with some difficulties due to variations in the morphology of the ICG signal [21]. There have been several algorithmic definitions proposed to approximate the location of the B-point [21], [22]. Figure Fig. 6 presents three examples showing the variation of the morphology of the ICG waveform. The general guideline is that the B-point indicates the onset of the rapid upstroke toward  $(dZ/dt)_{max}$  [21]. An algorithm known as *second derivative classification* [22], [58] is used to find the B-point, in which we first calculate the 2nd time derivatives of the  $dZ/dt$  signal and then we find the times when the maximum peak of this 2nd derivative of  $dZ/dt$  occurs within a small time period( 80 msec) before the time that  $(dZ/dt)_{max}$  occurs. This algorithm robustly detects the B-point in all three of the different morphology samples illustrated in Fig. 6. To evaluate the performance of the algorithm, we compared B-points detected by the algorithm to B-points detected by visual inspection by trained scorers for 150 cardiac cycles from 20 different subjects [7]. Presented results in Fig. 7 (left) show that the automatic detection exactly matched visual detection more than 90% of the time. The remaining 10% disagreement was the result of the existence of outlier signals as well as small, sharp notches (B-points) on the signal that were not visible to the naked eye and therefore not picked by the human inspectors.

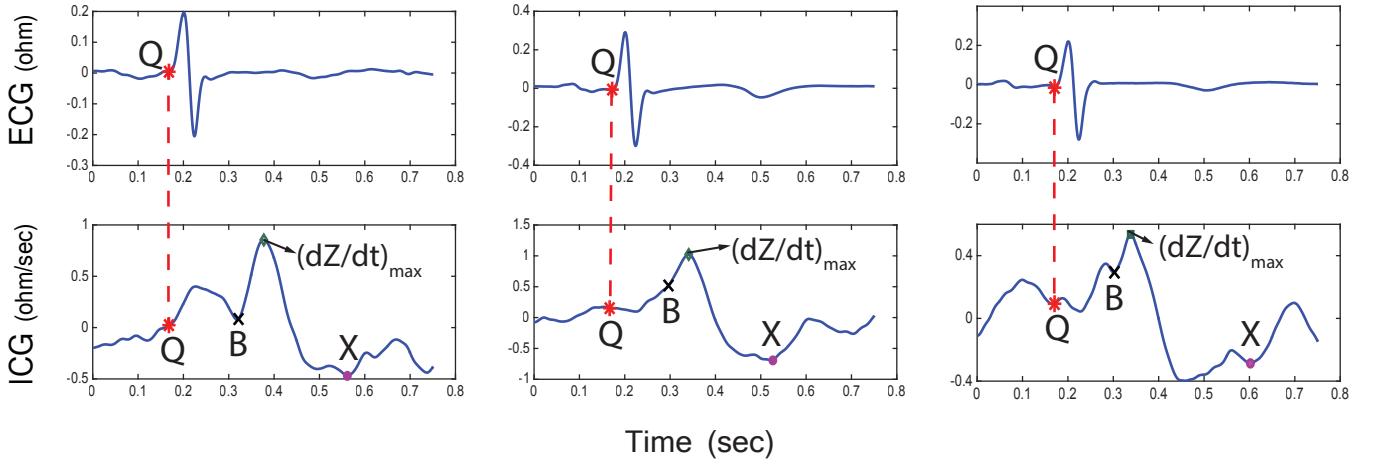


Fig. 6. Three cases of ICG signal(bottom) with the corresponding ECG signal (top) indicating variations in the morphology of ICG signal with the characteristic points identified by the algorithm. Left, the B- and X- points are local minimum and global minimum respectively; Middle, the B-point is a notch; Right, the X-point is just a local minimum and NOT a global minimum.

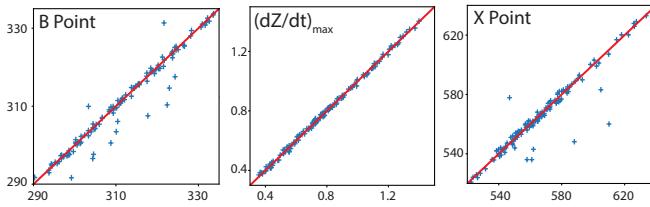


Fig. 7. Validation of (left) B-point location, (middle)  $dZ/dt_{max}$  value, (right) X-point location, detected by visual inspection (x-axis) vs. algorithm (y-axis) for 150 random cardiac cycles.

The Root-mean-square deviation (RMSD) [59] between the manual and algorithm detection of the B-point was 5.38 msec.

$(dZ/dt)_{max}$  is identical to the global maximum value of the ICG signal in one cardiac cycle. Full agreement exists between manual visual detection and the developed detection algorithm for  $(dZ/dt)_{max}$  (Fig. 7(b))

**X-point** represents the closing of the aortic valve to prevent the blood from the aorta streaming back into the left ventricle [21]. By using the 2nd derivatives of  $dZ/dt$ , we developed an X-point detection algorithm which is robust to morphological variations in the  $dZ/dt$  signals. The X-point is often defined by the point where the minimum value of  $dZ/dt$  in one cardiac cycle occurs. However, due to variations in the morphology of the  $dZ/dt$  waveform, this does not always hold true. For instance, in Fig. 6 the selected X-point is not the global minimum of the ICG signal in the right-most case. In such cases, the X-point is not a minimum point but a notch in the signal. To capture these variations, the 2nd derivative of the  $dZ/dt$  can be used: both minimum points and notches appear as peak values in the 2nd time derivative of the  $dZ/dt$  signal. Therefore, the algorithm proposed in this work is as follows: We search for all the peak indexes for the 2nd time derivative of the  $dZ/dt$  signal in a period of  $[T_B + 230 \text{ msec}, T_B + 400 \text{ msec}]$  and we choose the index that corresponds to the lowest value in the original  $dZ/dt$  signal. If no peak is found in the 2nd time derivative of the  $dZ/dt$  signal in this interval, we repeat the same process within a more relaxed period of

$[T_B + 200 \text{ msec}, T_B + 400 \text{ msec}]$ . These threshold values (200, 230 and 400 msec) are chosen based on the review of many  $dZ/dt$  signals, as well as recommendations made by field experts. Fig. 7(right) shows the result of the validation of the algorithm for X-point detection in the same sample dataset described for validating the B-point automatic detection algorithm. Root-mean-square deviation (RMSD) between manual- and algorithm-based detection of the X-point was 10.59 msec.

### 2.6.3 ICG Feature Extraction

After detecting the characteristic points explained above, we can compute physiological features of interest including pre-ejection period (PEP), left ventricular ejection time (LVET), stroke volume (SV), cardiac output (CO), and total peripheral resistance (TPR) using the following formulas [21]:

$$PEP = T_B - T_Q$$

$$LVET = T_X - T_B$$

$$SV = \rho \times (L/Z_0)^2 \times LVET \times (dZ/dt)_{max}$$

$$CO = SV \times heart rate / 1000$$

$$TPR = (BP/CO) \times 80$$

where  $T_B$ ,  $T_Q$  and  $T_X$  are the times (in sec) when B-, Q- and X-points occur, respectively.  $\rho$  is the resistivity of blood (in ohm-cm) and is usually set to 135 ohm-cm [21].  $L$  (cm) is the distance between the recording electrodes used for obtaining ICG and BP is the mean value of blood pressure over the same segment. Finally, the mean value of  $Z_0$  is also reported as a feature, since researchers may be interested in having an estimate of basal thoracic impedance during tasks or baselines.

## 3 QUALITY ASSESSMENT OF BIOSIGNALS

Signal recordings during improper sensor attachment or large movements not only carry no useful information, they also may result in malfunctioning of decision making biomedical systems (e.g., emergency alarms) and/or corrupt machine learning models that are trained on these signals.

Therefore, it is necessary to implement basic data quality assessments ahead of biosignal processing such that these large-scale quality issues can be detected and eliminated (or properly adjudicated).

While basic data quality assessments have traditionally been done by manual visual inspection, the quantity of data being collected in many modern psychophysiology experiments makes this approach unfeasible. Thanks to technological advancements in physiological sensor technology, particularly mobile sensors, it is not uncommon for research studies to collect massive amounts of physiological signal data (on the order of tens of GBs to multiple TBs) [7], [60]. Hence, manual quality checking should be replaced by automated methods. There have been numerous biosignal-specific algorithms developed for the automatic quality checking of biosignals [61]–[63], [63]–[66]. Influenced by these algorithms found in the literature, here we developed algorithms for basic quality checking of ECG, BP, ICG and EDA signals. An algorithm inspired by Moody [64] is used for the quality checking of ECG, ICG and BP, while an algorithm proposed by Klencher [65] is used for the quality checking of EDA.

### 3.1 ECG, ICG and BP Quality Checking

A rule-based algorithm proposed by Moody [64] was primarily developed for the quality checking of ECG. This algorithm has achieved among the highest accuracy scores in the Physionet/CinC Challenge [63], [66]. The algorithm inspects chunks of the signals based on the following three criteria:

1. If the signal is flat, that could be a sign that the sensors are not attached. Mathematically speaking, if the standard deviation of a chunk of a signal is very close to zero, that chunk is a flat signal.

2. The minimum, maximum, and the maximum-minimum difference values of the signal have to be within an acceptable range. Otherwise, too small or too large values may be the result of improper attachment of the sensors or large noise spikes that are severely drifting the baseline [64].

3. The ECG signal is mostly fairly quiet except for a short time period which contains large and sudden changes. Therefore, the frequency of large changes should be within a certain range to ensure the quality of the signal. In other words, given a signal chunk (with the size of larger than one heartbeat), the differences between the maximum and minimum of its small sub-chunks as a random variable will make a highly positively skewed distribution. We measure the skewness using the formula provided by Groeneveld & Meeden [67]. This rule, detects chunks of signal that are not similar to the specific shapes of the clean signal.

If a chunk of the signal is failed by any of these criteria, that chunk will be marked as low quality. Since ECG, ICG and continuous BP contain similar frequencies and patterns related to heart activity, by tuning the threshold parameters used for rules 1-3 described above for ECG, we applied similar quality checking algorithms for the other two biosignals (BP and ICG). In order to evaluate the performance of our basic quality checking algorithms for ECG, ICG, and

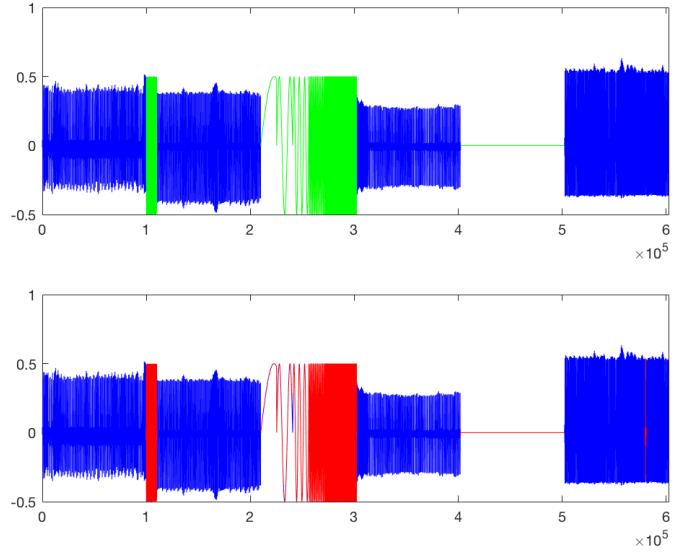


Fig. 8. Quality checking of ECG signal. Top : A 700-second raw signal composed of 4 different clean ecg datasets (blue) as well as three irrelevant synthetically generated signals (green). Bottom: low quality signal portions detected by red color

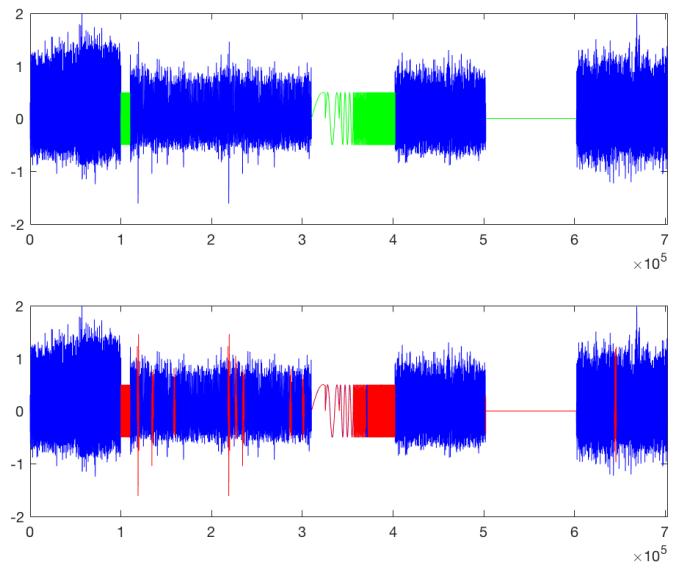


Fig. 9. Quality checking of ICG signal: Top : A 600-second raw signal composed of 4 different clean icg datasets (blue) as well as three irrelevant synthetically generated signals (green). Bottom: low quality signal portions detected by red color

BP, we examined the algorithm on signals comprising a combination of clean signals and synthetically generated corrupted signals (Fig. ??). Accordingly, we created input signals containing four segments of clean biosignal from four different datasets [7], [32], [33] as well as several segments with synthetically generated noise and artifacts between the segments of clean signals. As shown in Fig. ?? (bottom), our proposed algorithms robustly detected all the artifact segments (red color) while correctly passing the clean signals (blue color).

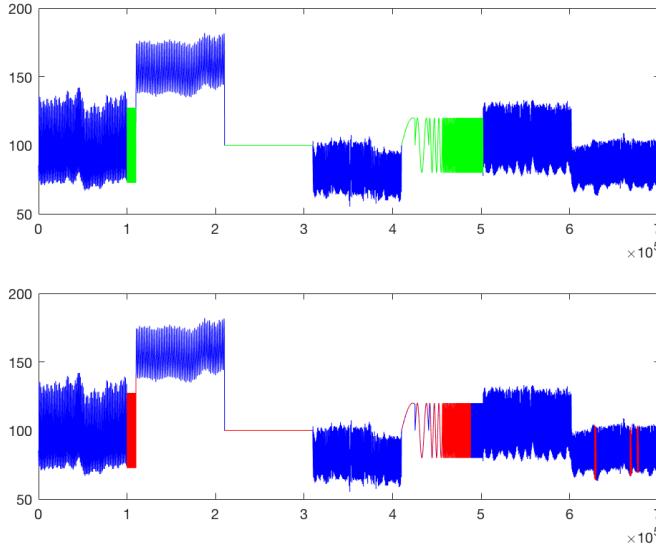


Fig. 10. Quality checking of BP signal. Top : A 700-second raw signal composed of 4 different clean bp datasets (blue) as well as three irrelevant synthetically generated signals (green). Bottom: low quality signal portions detected by red color

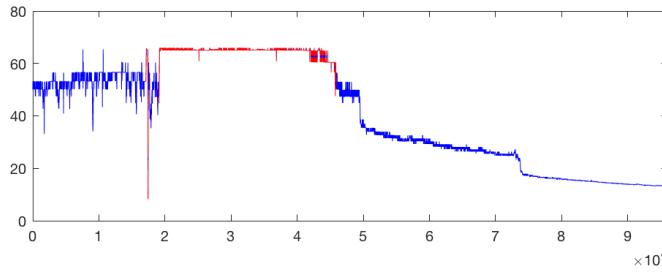


Fig. 11. Quality checking of EDA signal.top: A raw signal composed of 4 different clean EDA datasets as well as three irrelevant synthetically generated signals. bottom: low quality signal portions detected by red color

### 3.2 EDA Quality Checking

According to [65], [68], three simple rules for determining invalid data segments in the EDA signal are used:

1. The minimum or maximum values of the EDA signal lie outside a certain acceptable range.
2. Changes in the EDA signal happen too quickly to reflect naturally occurring changes in sweat gland activity.
3. The EDA signal portions within 5 sec of invalid portions from rules 1 and 2 are also invalid.

Fig. 11 illustrates the a sample output of the EDA quality checking algorithm on real EDA data with portions of low quality signal marked as red. The algorithm has accurately detected both out-of-range portions of the signal as well as portions of the signal where the signal is changing too quickly. The hyper-parameters for rules 1 and 2 should be selected by the user to capture the specifics of their equipment and amplification configurations for EDA data collection.

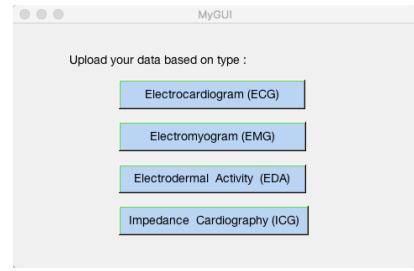


Fig. 12. First page of GI II to select physiological signal type

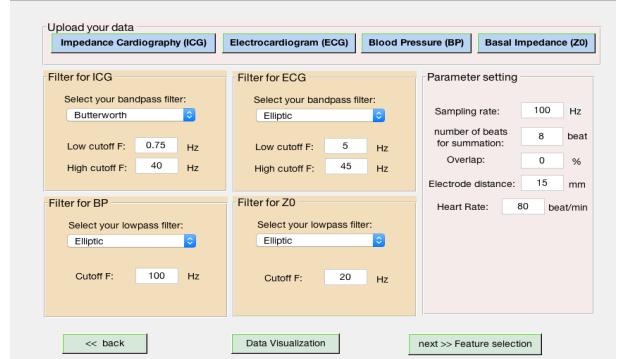


Fig. 13. Example of second page of GUI specialized for preprocessing of ICG signal.

## 4 GRAPHICAL USER INTERFACE (GUI)

Our biosignal-specific processing tool produces a feature matrix extracted from each biosignal which could be further used as input feature matrices for a variety of machine learning and pattern recognition algorithms, such as decision trees, support vector machines (SVM), and k-nearest neighbors. They could also be passed to the Classification Learner app in MatLAB designed for machine learning analysis [69].

### 4.1 Signal-Specific Preprocessing Page

On this page of the GUI (see Fig. 12), the user selects the type of biosignal which is intended to be processed. There are ECG, EMG, EDA, continues BP and ICG options. After selecting a signal type, a second page appears which requests the following inputs be uploaded or entered: (1) the raw signals, (2) filter characteristics, (3) specific parameters which may be required for some signal modalities (ex. sampling rate, heart beat and electrode distance for ICG). After importing the raw signal, the user can also visualize the time series signal at any stage in the preprocessing pipeline.

For ICG feature extraction, synchronous ECG, continuous BP,  $dz/dt$ , and  $Z_0$  signals all need to be uploaded (Fig. 13) as they are required for the characteristic point detection and feature extraction from the ICG signal.

On the filter panel (Fig. 13), users are able to choose the filter type (Elliptic, Butterworth, or Gaussian) as well as low and high cutoff frequencies.

Fig. 13 illustrates the GUI page for the preprocessing of ICG.

### 4.2 Feature Extraction Page

On the feature extraction page, users are able to select the features of interest among all the computed feature vectors

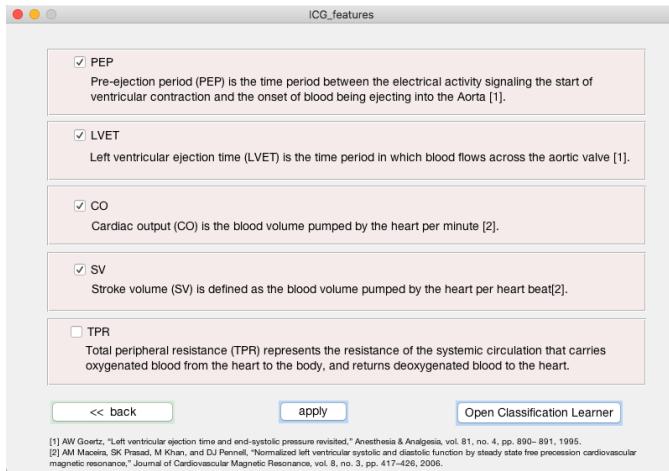


Fig. 14. Example of third page of GUI specialized for feature extraction from ICG signal.

for the analyzed signal. One can also continue directly to the classification stage by clicking the *Classification Learner App* button which opens up the MathWorks learner app. This app will be able to use the created feature matrix and perform multiple classification tasks. The feature extraction page for the ICG signal is shown in Fig. 14.

## 5 CONCLUSION

Here we present the development of a biosignal-specific processing and feature extraction software toolbox for ECG, EDA, EMG, continuous BP and ICG biosignal modalities which utilizes state-of-the-art algorithms provided in the scientific literature or based on feedback received from field experts. This open-source tool assists researchers in machine learning, affective computing and psychology by automatically and reliably extracting signal-specific physiologically-relevant feature vectors which can be further used as input feature vectors for a wide variety of machine learning and pattern recognition models.

## REFERENCES

- [1] Mathias Baumert, Alberto Porta, and Andrzej Cichocki, "Biomedical signal processing: From a conceptual framework to clinical applications [scanning the issue]," *Proceedings of the IEEE*, vol. 104, no. 2, pp. 220–222, 2016.
- [2] Matjaž Kukar, Igor Kononenko, Ciril Grošelj, Katarina Kralj, and Jure Fettich, "Analysing and improving the diagnosis of ischaemic heart disease with machine learning," *Artificial intelligence in medicine*, vol. 16, no. 1, pp. 25–50, 1999.
- [3] Javier Hernandez, Ivan Riobo, Agata Rozga, Gregory D Abowd, and Rosalind W Picard, "Using electrodermal activity to recognize ease of engagement in children during social interactions," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2014, pp. 307–317.
- [4] Abdulhamit Subasi, "Classification of emg signals using pso optimized svm for diagnosis of neuromuscular disorders," *Computers in biology and medicine*, vol. 43, no. 5, pp. 576–586, 2013.
- [5] Nahla Barakat, Andrew P Bradley, and Mohamed Nabil H Barakat, "Intelligible support vector machines for diagnosis of diabetes mellitus," *IEEE transactions on information technology in biomedicine*, vol. 14, no. 4, pp. 1114–1120, 2010.
- [6] Souhir Chabchoub, Sofienne Mansouri, and Ridha Ben Salah, "Detection of valvular heart diseases using impedance cardiography icg," *Biocybernetics and Biomedical Engineering*, 2017.
- [7] Aya Khalaf, Mohsen Nabian, Miaolin Fan, Yu Yin, Jolie Wormwood, Erika Siegel, Karen Quigley, Lisa Feldman Barrett, Murat Akcakaya, Chun-An Chou, et al., "Analysis of multimodal physiological signals within and across individuals to predict psychological threat vs. challenge," 2017.
- [8] Pranav Rajpurkar, Awni Y Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, 2017.
- [9] Deborah Lewis, Betty L Chang, and Charles P Friedman, "Consumer health informatics," in *Consumer Health Informatics*, pp. 1–7. Springer, 2005.
- [10] Carmen Vidaurre, Tilmann H Sander, and Alois Schlögl, "Biosig: the free and open source software library for biomedical signal processing," *Computational intelligence and neuroscience*, vol. 2011, 2011.
- [11] Vaibhav D Awandekar Priya Rani Rameshwari S Mane, A N Cheeran, "Cardiac arrhythmia detection by ecg feature extraction," *International Journal of Engineering Research and Applications (IJERA)*, vol. 3, no. 2, pp. 327–332, 2013.
- [12] Jiapu Pan and Willis J Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, , no. 3, pp. 230–236, 1985.
- [13] Mahesh S. Chavan, R. A. Agarwala, and M. D. Uplane, "Digital elliptic filter application for noise reduction in ecg signal," *4th WSEAS International Conference on ELECTRONICS, CONTROL and SIGNAL PROCESSING*, pp. 58–63, 2005.
- [14] Majdi Bsoul, Hlaing Minn, Mehrdad Nourani, Gopal Gupta, and Lakshman Tamil, "Real-time sleep quality assessment using single-lead ecg and multi-stage svm classifier," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 1178–1181.
- [15] Philip De Chazal, Conor Heneghan, Elaine Sheridan, Richard Reilly, Philip Nolan, and Mark O'Malley, "Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea," *Biomedical Engineering, IEEE Transactions on*, vol. 50, no. 6, pp. 686–696, 2003.
- [16] HFS Gamboa and ALN Fred, "An electrodermal activity psychophysiological model," 2005.
- [17] Alberto Greco, Gaetano Valenza, and Enzo Pasquale Scilingo, *Advances in Electrodermal Activity Processing with Applications for Mental Health: From Heuristic Methods to Convex Optimization*, Springer, 2016.
- [18] Kyung Hwan Kim, SW Bang, and SR Kim, "Emotion recognition system using short-term monitoring of physiological signals," *Medical and biological engineering and computing*, vol. 42, no. 3, pp. 419–427, 2004.
- [19] Alexandre Laurin, "BPannotate," <https://www.mathworks.com/matlabcentral/fileexchange/60172-bp-annotate>, 2017, [Online; accessed 02-05-2018].
- [20] JX Sun, AT Reisner, and RG Mark, "A signal abnormality index for arterial blood pressure waveforms," in *Computers in Cardiology, 2006*. IEEE, 2006, pp. 13–16.
- [21] Michael T Allen, Jochen Fahrenberg, Robert M Kelsey, William R Lovallo, Lorenz JP Doornen, et al., "Methodological guidelines for impedance cardiography," *Psychophysiology*, vol. 27, no. 1, pp. 1–23, 1990.
- [22] Liang-Yu Shyu, Yuh-Shii Lin, Chun-Peng Liu, and Wei-Chih Hu, "The detection of impedance cardiogram characteristic points using wavelet transform," *Computers in biology and medicine*, vol. 34, no. 2, pp. 165–175, 2004.
- [23] Joe Tomaka, Jim Blascovich, Robert M Kelsey, and Christopher L Leitten, "Subjective, physiological, and behavioral effects of threat and challenge appraisal," *Journal of Personality and Social Psychology*, vol. 65, no. 2, pp. 248, 1993.
- [24] Joe Tomaka, Jim Blascovich, Jeffrey Kibler, and John M Ernst, "Cognitive and physiological antecedents of threat and challenge appraisal," *Journal of personality and social psychology*, vol. 73, no. 1, pp. 63, 1997.
- [25] Wendy Berry Mendes, Brenda Major, Shannon McCoy, and Jim Blascovich, "How attributional ambiguity shapes physiological and emotional responses to social rejection and acceptance," *Journal of personality and social psychology*, vol. 94, no. 2, pp. 278, 2008.
- [26] Jeremy P Jamieson, Matthew K Nock, and Wendy Berry Mendes, "Mind over matter: reappraising arousal improves cardiovascular

- and cognitive responses to stress.,” *Journal of Experimental Psychology: General*, vol. 141, no. 3, pp. 417, 2012.
- [27] Karen S Quigley, Lisa Feldman Barrett, and Suzanne Weinstein, “Cardiovascular patterns associated with threat and challenge appraisals: A within-subjects analysis,” *Psychophysiology*, vol. 39, no. 3, pp. 292–302, 2002.
- [28] Alois Schlögl and Clemens Brunner, “Biosig: a free and open source software library for bci research,” *Computer*, vol. 41, no. 10, 2008.
- [29] Dennis Tkach, He Huang, and Todd A Kuiken, “Study of stability of time-domain features for electromyographic pattern recognition,” *Journal of neuroengineering and rehabilitation*, vol. 7, no. 1, pp. 1, 2010.
- [30] Mohsen Nabian, Athena Nouhi, Yu Yin, and Sarah Ostadabbas, “A biosignal-specific processing tool for machine learning and pattern recognition,” *Healthcare Innovations and Point of Care Technologies (HI-POCT)*, 2017 IEEE, 2017.
- [31] “A Biosignal-Specific Processing Tool,” <http://www.northeastern.edu/ostadabbas/codes/>, Accessed: 2017.
- [32] PhysioBank ATM, ” <https://physionet.org/cgi-bin/atm/ATM>, 2016.
- [33] MIT Effective Computing Group, ” <http://affect.media.mit.edu>, 2016.
- [34] Dirk De Bacquer, Guy De Backer, Marcel Kornitzer, and Henry Blackburn, “Prognostic value of ecg findings for total, cardiovascular disease, and coronary heart disease death in men and women,” *Heart*, vol. 80, no. 6, pp. 570–577, 1998.
- [35] Hassan Ghasemzadeh, Sarah Ostadabbas, Eric Guenterberg, and Alexandros Pantelopoulos, “Wireless medical-embedded systems: A review of signal-processing techniques for classification,” *IEEE Sensors Journal*, vol. 13, no. 2, pp. 423–437, 2013.
- [36] Manuel Blanco-Velasco, Binwei Weng, and Kenneth E Barner, “Ecg signal denoising and baseline wander correction based on the empirical mode decomposition,” *Computers in biology and medicine*, vol. 38, no. 1, pp. 1–13, 2008.
- [37] Amer I Aladin, Seamus P Whelton, Mouaz H Al-Mallah, Michael J Blaha, Steven J Keteyian, Stephen P Juraschek, Jonathan Rubin, Clinton A Brawner, and Erin D Michos, “Relation of resting heart rate to risk for all-cause mortality by gender after considering exercise capacity (the henry ford exercise testing project),” *American Journal of Cardiology*, vol. 114, no. 11, pp. 1701–1706, 2014.
- [38] Matthew Lombard and Theresa Ditton, “At the heart of it all: The concept of presence,” *Journal of Computer-Mediated Communication*, vol. 3, no. 2, pp. 0–0, 1997.
- [39] G Dorfman Furman, Z Shinar, A Baharav, and S Akselrod, “Electrocardiogram derived respiration during sleep,” in *Computers in Cardiology*, 2005. IEEE, 2005, pp. 351–354.
- [40] George B Moody, Roger G Mark, Andrea Zoccola, and Sara Mantero, “Derivation of respiratory signals from multi-lead ecgs,” *Computers in cardiology*, vol. 12, no. 1985, pp. 113–116, 1985.
- [41] Raymond J Gibbons, Gary J Balady, J Timothy Bricker, Bernard R Chaitman, Gerald F Fletcher, Victor F Froelicher, Daniel B Mark, Ben D McCallister, Aryan N Mooss, Michael G O'reilly, et al., “Acc/aha 2002 guideline update for exercise testing: summary article: a report of the american college of cardiology/american heart association task force on practice guidelines (committee to update the 1997 exercise testing guidelines),” *Journal of the American College of Cardiology*, vol. 40, no. 8, pp. 1531–1540, 2002.
- [42] Michael E Dawson, Anne M Schell, and Diane L Filion, “The electrodermal system,” *Handbook of psychophysiology*, vol. 2, pp. 200–223, 2007.
- [43] Maria S Perez-Rosero, Behnaz Rezaei, Murat Akcakaya, and Sarah Ostadabbas, “Decoding emotional experiences through physiological signal processing,” *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pp. 881–885, 2017.
- [44] Joey Benedek and R Hazlett, “Incorporating facial emg emotion measures as feedback in the software design process,” *Proc. Human Computer Interaction Consortium*, 2005.
- [45] M BI Reaz, MS Hussain, and Faisal Mohd-Yasin, “Techniques of emg signal analysis: detection, processing, classification and applications,” *Biological procedures online*, vol. 8, no. 1, pp. 11–35, 2006.
- [46] Paul Ekman, “Facial expression and emotion.,” *American psychologist*, vol. 48, no. 4, pp. 384, 1993.
- [47] CJ De Luca, LD Gilmore, M Kuznetsov, and SH. Roy, “Filtering the surface emg signal: Movement artifact and baseline noise contamination.,” *Journal of Biomechanics*, vol. 43, no. 8, pp. 1573–1579, 2010.
- [48] W Zong, T Heldt, GB Moody, and RG Mark, “An open-source algorithm to detect onset of arterial blood pressure pulses,” in *Computers in Cardiology*, 2003. IEEE, 2003, pp. 259–262.
- [49] Shang-Yi Chuang, Jia-Ju Liao, Chia-Ching Chou, Chia-Chi Chang, and Wai-Chi Fang, “An effective arterial blood pressure signal processing system based on eemd method,” in *VLSI Design, Automation and Test (VLSI-DAT), 2014 International Symposium on*. IEEE, 2014, pp. 1–4.
- [50] J-F Augusto, J-L Teboul, P Radermacher, and P Asfar, “Interpretation of blood pressure signal: physiological bases, clinical relevance, and objectives during shock states,” *Intensive care medicine*, vol. 37, no. 3, pp. 411–419, 2011.
- [51] Simon S Young, *Computerized data acquisition and analysis for the life sciences: a hands-on guide*, Cambridge University Press, 2001.
- [52] Benjamin Etstern and TH Li, “Hemodynamic changes during thiopental anesthesia in humans: cardiac output, stroke volume, total peripheral resistance, and intrathoracic blood volume,” *The Journal of clinical investigation*, vol. 34, no. 3, pp. 500–510, 1955.
- [53] Axel W Goertz, Tobias Mehl, Karl H Lindner, Michael G Rockemann, Uwe Schirmer, Bernhard Schwilk, and Michael Georgieff, “Effect of 7.2% hypertonic saline/6% hetastarch on left ventricular contractility in anesthetized humans,” *The Journal of the American Society of Anesthesiologists*, vol. 82, no. 6, pp. 1389–1395, 1995.
- [54] AM Maceira, SK Prasad, M Khan, and DJ Pennell, “Normalized left ventricular systolic and diastolic function by steady state free precession cardiovascular magnetic resonance,” *Journal of Cardiovascular Magnetic Resonance*, vol. 8, no. 3, pp. 417–426, 2006.
- [55] Robert M Kelsey and William Guethlein, “An evaluation of the ensemble averaged impedance cardiogram,” *Psychophysiology*, vol. 27, no. 1, pp. 24–33, 1990.
- [56] “Knowledge Base Impedance Cardiography,” <https://support.mindwaretech.com/article-categories/imp/>, Accessed: 2017.
- [57] Robert M Kelsey, Sarah Reiff, Stefan Wiens, Tamara R Schneider, Elizabeth S Mezzacappa, and William Guethlein, “The ensemble-averaged impedance cardiogram: An evaluation of scoring methods and interrater reliability,” *Psychophysiology*, vol. 35, no. 3, pp. 337–340, 1998.
- [58] Javier Rodríguez Árbol, Pandelis Perakakis, Alba Garrido, José Luis Mata, M Carmen Fernández-Santaella, and Jaime Vila, “Mathematical detection of aortic valve opening (b point) in impedance cardiography: A comparison of three popular algorithms,” *Psychophysiology*, vol. 54, no. 3, pp. 350–357, 2017.
- [59] Michel Petitjean, “On the root mean square quantitative chirality and quantitative symmetry measures,” *Journal of Mathematical Physics*, vol. 40, no. 9, pp. 4587–4595, 1999.
- [60] Jin Joo Lee, Brad Knox, Jolie Baumann, Cynthia Breazeal, and David DeSteno, “Computationally modeling interpersonal trust,” *Frontiers in psychology*, vol. 4, pp. 893, 2013.
- [61] Pankaj Malhotra, Lovekesh Vig, Gautam Shroff, and Puneet Agarwal, “Long short term memory networks for anomaly detection in time series,” in *Proceedings. Presses universitaires de Louvain*, 2015, p. 89.
- [62] Udit Satija, Barathram Ramkumar, and M Sabarimalai Manikandan, “An automated ecg signal quality assessment method for unsupervised diagnostic systems,” *Biocybernetics and Biomedical Engineering*, vol. 38, no. 1, pp. 54–70, 2018.
- [63] Eduardo Morgado, Felipe Alonso-Atienza, Ricardo Santiago-Mozos, Óscar Barquero-Pérez, Ikaro Silva, Javier Ramos, and Roger Mark, “Quality estimation of the electrocardiogram using cross-correlation among leads,” *Biomedical engineering online*, vol. 14, no. 1, pp. 59, 2015.
- [64] Benjamin E Moody, “Rule-based methods for ecg quality control,” in *Computing in Cardiology*, 2011. IEEE, 2011, pp. 361–363.
- [65] Ian Kleckner, “Electrodermal activity quality assessment,” <https://github.com/iankleckner/EDAQA>, 2016.
- [66] Ikaro Silva, George B Moody, and Leo Celi, “Improving the quality of ecgs collected using mobile phones: The physionet/computing in cardiology challenge 2011,” in *Computing in Cardiology*, 2011. IEEE, 2011, pp. 273–276.
- [67] Richard A Groeneveld and Glen Meeden, “Measuring skewness and kurtosis,” *The Statistician*, pp. 391–399, 1984.
- [68] Ian R Kleckner, Rebecca M Jones, Oliver Wilder-Smith, Jolie B Wormwood, Murat Akcakaya, Karen S Quigley, Catherine Lord, and Matthew S Goodwin, “Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data,” *IEEE Transactions on Biomedical Engineering*, 2017.

[69] "Mathworks Classification Learner," <https://www.mathworks.com/products/statistics/classification-learner.html>, Accessed: 2017.



**Mohsen Nabian** received his B.Sc. in Mechanical Engineering from Sharif University of Technology, Tehran, Iran in 2012 and received M.Sc. in Computer Science from Northeastern University, Boston, MA in 2016. He is also a Ph.D. candidate at Northeastern University and a researcher at the Augmented Cognition Lab (ACLab). His research interests include bio-signal processing and machine learning in healthcare.



**Lisa Feldman Barrett** received a bachelors degree in Psychology from the University of Toronto, and a PhD in Psychology from the University of Waterloo. She is a University Distinguished Professor of Psychology and Director of the Interdisciplinary Affective Science Laboratory in the Psychology Department at Northeastern University. She also holds appointments as a Research Scientist in the Department of Psychiatry at Massachusetts General Hospital (MGH), as a Neuroscientist in the Department of Radiology at MGH, and as a Lecturer in Psychiatry at Harvard Medical School. She received a National Institutes of Health Director's Pioneer Award for her groundbreaking research on emotion in the brain, and is an elected member of the Royal Society of Canada. Her research examines emotion from both psychological and neuroscience perspectives, developing a general framework for understanding how the brain creates the mind.



**Yu Yin** received B.Sc. degree in Electrical and Information Engineering Department, Wuhan University of Technology, China, in 2016. She is currently an M.Sc. student in the Electrical and Computer Engineering Department of Northeastern University. She is also a member of the Augmented Cognition Lab (ACLab). Her research interests include bio-signal processing, computer vision, and machine learning.



**Jolie B. Wormwood** received B.Sc. degrees in Mathematics and Psychology from Ithaca College. She received a M.Sc. degree and Ph.D. in Social Psychology from Northeastern University. She completed a postdoctoral research assistantship in the Interdisciplinary Affective Science Laboratory at Northeastern University, where she is currently an associate research scientist. She is interested in the role of the body in affect, decision making, and risk perception.



**Sarah Ostadabbas** received her B.Sc. in both Electrical and Biomedical Engineering from Amirkabir University of Technology, Tehran, Iran in 2005 and an M.Sc. degree in Control Engineering from Sharif University of Technology, Tehran, Iran in 2007. In 2014, she received her Ph.D. in Electrical Engineering from the University of Texas at Dallas. She is currently an Assistant Professor in the Electrical and Computer Engineering Department of Northeastern University, where she direct the Augmented Cognition Lab (ACLab) with the goal of enhancing human information-processing capabilities through the design of adaptive interfaces via physical, physiological, and cognitive state estimation. Her research focus is at the intersection of computer vision and machine learning with the multidisciplinary elements from behavioral sciences.



**Karen Quigley** received a Bachelors degree in Zoology, a Masters degree in Psychobiology, and a Ph.D. in Psychobiology from The Ohio State University before completing a postdoctoral fellowship at Columbia University. She is currently a Research Associate Professor and co-director of the Interdisciplinary Affective Science Laboratory in the Psychology Department at Northeastern University, as well as a Research Physiologist for the Center for Healthcare Organization & Implementation Research (CHOIR) at the Edith Nourse Rogers Memorial VA Hospital, and Corresponding Director of the CHOIR Post-doctoral Research Program. Her research focuses on how humans utilize information from the body in creating emotional and affective states, individual differences in emotional reactivity, and how emotional states influence behavior, cognitions, and health.