

# A Biosignal-Specific Processing Tool for Machine Learning and Pattern Recognition

Yu Yin\*, Mohsen Nabian\*, Athena Nouhi\*, and Sarah Ostadabbas

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

## I. INTRODUCTION

Physiological signals (biosignals) including Electrocardiogram (ECG), Impedance Cardiography (ICG), Electromyogram (EMG) and Electrodermal Activity (EDA) have rich information about physiological functioning of the body, sleep quality, and human affective state, thus are widely used as input data for machine learning and pattern recognition research that are focused on preventive care, diagnostics, and guiding therapy [1].

A significant amount of knowledge is required to properly process these biosignals since distinct methods for noise and artifact removal, segmentation and feature extraction are needed for each signal modality [2]. To this end, there have been many published works introducing biosignal specific processing algorithms for ECG [3], [4], [5], [4], EMG [6], [7], EDA [8], [9], [10] and ICG [11], [12]. Due to the huge rising interest and usability of these biosignals in medical and bio-engineering research, several commercial software along with open-source toolboxes are developed to ease the pre-processing and feature extraction of these signals. However, a large portion of these software are not suitable for a machine learning processing pipeline, since they are manually intensive and sometimes are unreliable and demanded visual inspection. Moreover, although it has been shown that ICG signal features such as cardiac output

(CO) and stroke volume (SV) are correlated with human psychosocial states [13], [14], [15], [16], [17], very few software/toolbox are found to have ICG processing tool along with other biosignals.

The present work aims at providing an open-source biosignal-specific tool for psychologist, neuroscientists and researchers in machine learning and pattern recognition to process ECG, EMG, EDA and ICG biosignals, all in one easy-to-use, MATLAB based toolbox and to extract relevant features for each signal automatically and reliably. The developed toolbox is accessible via MathWorks File Exchange site [18]. The preprocessing algorithms such as noise filtering as well as algorithms used for segmentation and feature extraction of each signal are based upon the state-of-the-art published scientific articles [4], [10], [19] and each are briefly described in this paper. The feature matrix outputted by our algorithm could be directly used by machine learning algorithms for further analysis. Besides, it could be simply passed to Classification Learner App provided by MATLAB.

## II. MATERIALS AND METHODS

### A. Dataset

All physiological measures presented in this paper including the ECG, EMG, EDA, ICG, and continuous blood pressure (BP), were recorded using BioLab v.3.0.13 (Mindware Technologies; Gahanna, OH), and were acquired on a BioNex 8- Slot Chassis (Model 50-3711-08) with sampling rate of 1000Hz. This dataset was recorded at Psychology Department of Northeastern University and we got access to the dataset via collaboration on several affective computing projects. Recorded data from 100 participants, including 40 males and 60 females, was used in this study, where 20% of them were employed for algorithm testing and verification purposes. Our open-source toolbox include sample physiological signals from 10 participants of the study.

### B. Biosignal #1: Electrocardiogram (ECG)

ECG is the main diagnostic approach for detecting cardiovascular diseases and contains many information such as human health situation, sleep quality and emotional states [3]. ECG cycle created by each heart beat contains several critical waves including P, QRS and T waves shown in Fig. 1. In our biosignal-specific processing tool, different types of filters could be chosen to remove ECG signal noise. Then a robust QRS detection algorithm is proposed based on Pan-Tompkins algorithm [4].

Y. Yin, M. Nabian, A. Nouhi, and S. Ostadabbas are with the Augmented Cognition Laboratory (ACLab), Electrical and Computer Engineering Department, Northeastern University, MA, USA (corresponding author's e-mail: ostadabbas@ece.neu.edu).

\*These authors contributed equally to this work. This work was supported by a funding from MathWorks.

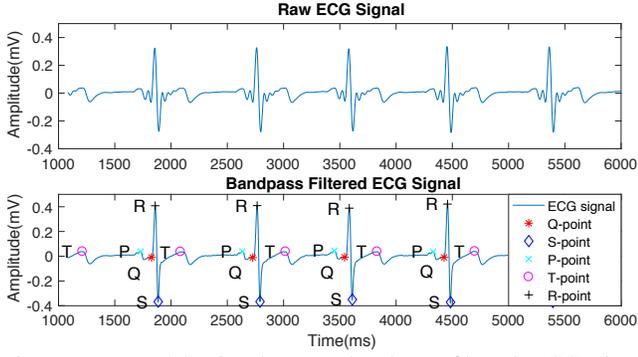


Fig. 1: Raw ECG signal (top); bandpass filtered ECG signal (bottom).

1) *ECG Preprocessing*: We have provided three filter type options for noise removal, including Elliptic, Gaussian, and Butterworth filters. With a thorough analysis of the signal power spectral density (PSD), we found out that a band-pass Elliptic filter has the best performance on ECG signal to remove baseline and high frequency noise. Hence we assigned this filter as the default option in the filter menu.

2) *ECG Feature Extraction*: Inspired by the Pan-Tompkins algorithm, we have modified the detection algorithm to reflect a more robust QRS detection. The detected locations of R, P, T, Q, and S in a few sample heartbeats are shown in Fig. 1.

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

**RR interval** is defined as the means of all detected RR intervals in a segment [6].

**SD** is the standard deviation of all detected RR intervals in a segment [6].

**SDD** is defined as the standard deviation of the differences between adjacent RR intervals in a segment [6].

**NN50** is defined as the number of pairs of adjacent RR intervals where the first RR interval exceeds the second RR interval by more than 50ms [6].

**pNN50** is defined as the number of pairs of adjacent RR intervals where the second RR interval exceeds the first RR interval by more than 50ms [6].

**EDR** which is calculated based on the area of each normal QRS complex measured over a fixed window width (which is determined by the interval from the PQ junction to the J-point of a normal QRS).  $w$  is the width of moving window. We estimated  $w$  with two times sum of QR interval and RS interval ( $w = 2 \times (QR + RS)$ ). The center of the window is aligned over each R peak to calculate the area under it. This feature is calculation of mean of EDR. Standard deviation of EDR is also computed as a feature.

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

### C. Biosignal #2: Electrodermal Activity (EDA)

EDA reflects momentary or long term changes in skin electrical conductivity due to various internal or external stimuli [20]. The EDA signal is composed of two activities,

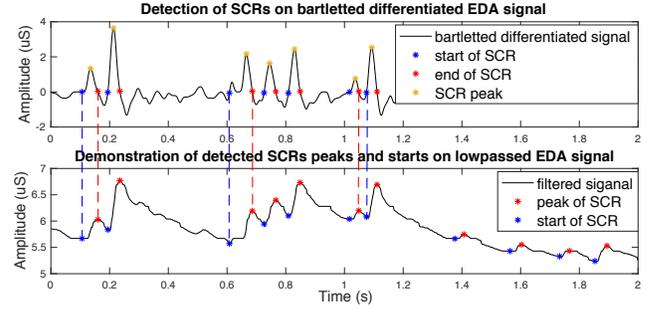


Fig. 2: Detection of SCRs on bartlett differentiated EDA signal (top) and demonstration of detected SCRs' peaks and starts on lowpassed EDA signal (bottom).

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 response (SCR), which is related to exterior stimuli or non-specific activation and is the bumps that appear in the signal [8]. Because of the relation of skin conductance variations to the sympathetic nervous system, EDA signal has been widely studied in emotion detection research [21]. Many important features for this purpose are extracted from SCRs. As a result, it is important to detect SCRs accurately.

1) *EDA Preprocessing*: To eliminate noises carried by the EDA signal, a filtering step required prior to detection algorithm. The filtering options provided here are the same as ECG signals. In some studies, a Gaussian low-pass filter have been recommended and used for the filtering process of EDA [9].

2) *EDA Feature Extraction*: SCRs was detected by performing differentiation and subsequent convolution with a 20-point Bartlett window [10], where Bartlett window is a triangle function represented as:

$$w(n) = \begin{cases} \frac{2n}{N}, & 0 \leq n \leq N/2 \\ 2 - \frac{2n}{N}, & N/2 < n \leq N \end{cases} \quad (1)$$

The occurrence of the SCR is detected by finding two consecutive zero-crossings, from negative to positive and positive to negative. 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 in that signal are excluded [10]. Fig. 2 demonstrates the detection of SCRs. On the top plot, the detected SCR's beginnings, ends, and peaks are shown on the differential signal. In the bottom plot, the corresponding locations of the previously detected SCRs are provided on the original low-pass filtered signal. After detecting the SCRs on the signal, the following features can be extracted:

**SCR duration mean** is the average of SRC durations. Duration of an SCR is defined as distance from location of the beginning to the end of an SCR.

**SCR amplitude mean** is the average of the amplitude of the detected SCRs.

**SCR rise-time mean** is the average of the SCR rise-times. Rise-time of an SCR is defined as time distance between the beginning to the peak of an SCR.

**Signal mean** is the average of the low-pass filtered signal.

**Number of detected SCRs** is the number of the detected SCRs in the EDA signal.

#### D. Biosignal #3: Electromyography (EMG)

EMG measures minute changes in the electrical activity of muscles, which reflects minute muscle movements [22]. EMG is a diagnostic procedure to determine the health of muscles and the nerve cells that control them (motor neurons) by recording and evaluating the electrical activity produced by skeletal muscles [23]. EMG signals are also widely used in the emotion classification studies as a means to create a taxonomy of facial expression [24].

1) *EMG Preprocessing*: Band-pass filters are used for the noise removal of EMG signals [25]. The type of filter options for EMG are the same as those provided for ECG. By the power spectral density inspection of the EMG signals, an Elliptic filter with the cutoff frequency of 10-300Hz was recommended for the EMG filter.

2) *EMG Feature Extraction*: We selected the feature options for EMG based on the features introduced in [19].

**MAV** is the mean absolute value of the EMG signal in a time window analysis with N samples.

**Zero crossing count** is the numbers of times the EMG signal crosses zero within a time window analysis.

**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.

**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 [19].

**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:

$$\log Detect = e^{\frac{1}{N} \sum_{k=1}^N \log |x_k|} \quad (2)$$

**SD** is standard deviation of the EMG low-passed signal using the filter selected in the EMG preprocessing page.

**RMS** is calculation of root mean square of the EMG low-passed signal using the filter selected in the EMG preprocessing page.

#### E. Biosignal #4: Impedance Cardiography (ICG)

ICG signal (also called  $dZ/dt$ ) measurement is a non-invasive method to extract important information (features) regarding the heart activity. These features include: (1) left ventricular ejection time (LVET), (2) stroke volume (SV), (3) cardiac output (CO), (4) pre-ejection period (PEP), and (5) total peripheral resistance (TPR). These five variables can be extracted from the shape characteristics of ICG signal combined with the corresponding ECG signal, transthoracic basal impedance ( $Z_0$ ) signals and blood pressure (BP) signal. The methods used to extract these features are based upon the discussions provided in [11], [26].

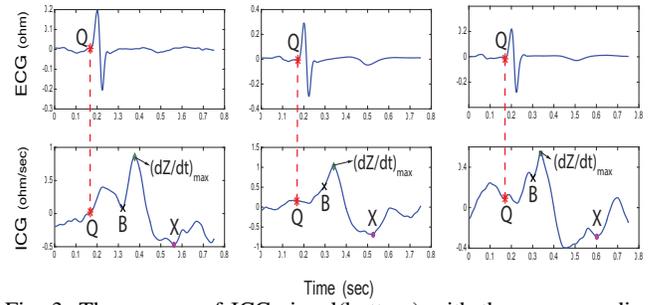


Fig. 3: Three cases of ICG signal(bottom) with the corresponding ECG signal (top) indicating variations in the morphology of ICG signal with the characteristic points identified by the algorithm. Left, the B- and X- points are local minimum and global minimum respectively; Middle, the B-point is a notch; Right, the X-point is just a local minimum and NOT a global minimum.

1) *ICG Preprocessing*: The preprocessing of ICG includes noise removal and ensemble averaging technique:

**Noise and artifact removal** To remove noise and artifact from the ICG signal, our developed software allows the users to adjust the low and high cut-off frequencies as well as the filter type. However, a second order elliptic band-pass filter with low cut-off frequency of 0.75Hz and high cut-off frequency of 40Hz is recommended in available Bio-signal software such as Biolab [27].

**Ensemble averaging** of multiple cardiac cycles is performed to remove the stochastically distributed noise as well as respiratory influences and movement artifacts on the ICG signal [11], [26], [28]. In this work, the number of cardiac cycles for ensemble averaging can be chosen as an input.

2) *ICG Feature Extraction*: In order to extract features from the ICG signal, the following characteristic points have to be identified on the ICG signal:

**Q-point** location on the ICG signal is identical to the location of Q-point on the ECG signal (Fig. 3).

**B-point** automatic identification comes with some difficulties due to the variations in the morphology of the ICG signal [11]. There has been several algorithmic definitions proposed to approximate the location of the B-point [11], [12]. Fig. 3 presents three examples showing the variation of morphology of the ICG waveform. The general guideline is that the B-point is on the onset of the rapid upstroke toward the  $(dZ/dt)_{max}$  [11]. An algorithm known as *second derivative classification* [12], [29] is used to find the B-point which robustly respond to all the three different morphology illustrated in Fig. 3. To evaluate the performance of the algorithm, B-point is detected both visually and by algorithm for 150 random cases with 20 cardiac cycles for ensemble averaging from 20 subjects Fig. 4 (left). Results show more than 90% exact detection. The remaining 10% comes from the outlier signals as well as invisible sharp notches.

$(dZ/dt)_{max}$  is identical to the global maximum value of the ICG signal in one cardiac cycle. Fig. 4(b) validates the algorithm performance for the detection of  $(dZ/dt)_{max}$ .

**X-point** represents the closing of the aortic valve to prevent the blood from the aorta streaming back into the left ventricle [11]. By using the 2nd derivatives of ICG, we present a X-point detection algorithm, which is robust to the

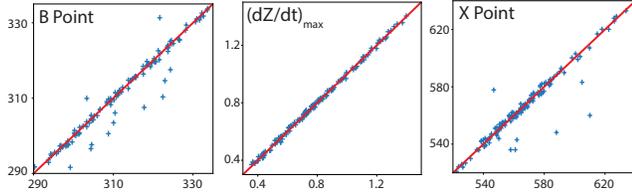


Fig. 4: Validation of (left) B-point location, (middle)  $dZ/dt_{max}$  value, (right) X-point location, detected by visual inspection (x-axis) vs. algorithm (y-axis) for 150 random cardiac cycles.

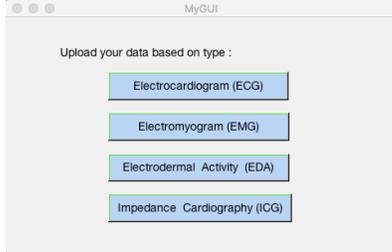


Fig. 5: First page of GUI to select physiological signal type

morphological variations in ICG signals. Fig. 4(right) shows the result of the validation of the algorithm for the X-point detection in the same experiment described for the B-point. Fig. 3 (right) demonstrate three ICG signals with different morphologies in which the characteristic points are correctly and robustly identified by our algorithm.

3) *ICG Feature Extraction*: After detecting the characteristic points explained above, we can compute the features with the following formula [11]:

$$PEP = T_B - T_Q$$

$$LVET = T_X - T_B$$

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

$$CO = SV \times \text{heartrate}/1000$$

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

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 (ohm-cm) and is usually set to 135 ohm-cm [11].  $L$  (cm) is the distance between the recording electrodes.

### III. GRAPHICAL USER INTERFACE (GUI)

Our biosignal-specific processing tool produces the feature matrix extracted from each biosignal which could be further used as input feature matrix for variety of machine learning and pattern recognition algorithms. They could also be passed to the Classification Learner app designed for machine learning analysis [30]. Using this app, several classification algorithms, such as decision trees, support vector machines (SVM), and k-nearest neighbors can be directed implemented.



Fig. 6: Example of second page of GUI specialized for preprocessing of ICG signal.

#### A. Signal-Specific Preprocessing Page

In this page (see Fig. 5), the user selects the type of the biosignal which is intended to be processed. There are ECG, EMG, EDA, and ICG options. By choosing each signal type, a second page appears in which the following inputs are being asked: (1) the raw signals, (2) filter characteristics, (3) special parameters which may be required for each signal (ex. heart beat, electrode distance). It is noted that the raw signal has to be in a matrix format with rows of data stack on top of each other. After importing raw signal, user can also visualize the time series signal at any stage in the preprocessing pipeline.

To increase the signal-to-noise ratio (SNR) of the uploaded signal, a filter panel is also provided in this page. On filter panel, you can choose the filter type as well as other parameters of the filter such as low and high cutoff frequencies. By pressing the *Apply* button or *Feature selection* button, the filter will be applied to all of the data. An example of preprocessing page for the ICG signal is shown in Fig. 6.

#### B. Feature Extraction Page

In feature extraction page, user can select the feature vectors of interest among all the computed feature vectors for the analyzed signal. By clicking *Apply* button, the feature matrix with the name of "Feature-Matrix" will be created in the current MATLAB workspace. One can further continue to the classification stage by clicking *Classification Learner App* button which opens up the MathWorks learner app and will be able to use the created feature matrix and perform multiple classification tasks.

### IV. CONCLUSION

In this work, we have created a biosignal-specific processing and feature extraction software in MATLAB based on the state-of-the-art algorithms provided in scientific literature for each type of biosignals. The software computes the corresponding feature matrix of biosignals, which can be further used as input feature matrix for variety of machine learning and pattern recognition algorithms. This open-source tool could be useful in facilitating research in machine learning, affective computing, and psychology.

## REFERENCES

- [1] M. Baumert, A. Porta, and A. 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] C. Vidaurre, T. H. Sander, and A. Schlögl, "Biosig: the free and open source software library for biomedical signal processing," *Computational intelligence and neuroscience*, vol. 2011, 2011.
- [3] V. D. A. P. R. 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.
- [4] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *Biomedical Engineering, IEEE Transactions on*, no. 3, pp. 230–236, 1985.
- [5] M. 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.
- [6] M. Bsoul, H. Minn, M. Nourani, G. Gupta, and L. 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.
- [7] P. De Chazal, C. Heneghan, E. Sheridan, R. Reilly, P. Nolan, and M. 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.
- [8] H. Gamboa and A. Fred, "An electrodermal activity psychophysiological model," 2005.
- [9] A. Greco, G. Valenza, and E. P. Scilingo, *Advances in Electrodermal Activity Processing with Applications for Mental Health: From Heuristic Methods to Convex Optimization*. Springer, 2016.
- [10] K. H. Kim, S. Bang, and S. 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.
- [11] M. T. Allen, J. Fahrenberg, R. M. Kelsey, W. R. Lovallo, L. J. Dornen *et al.*, "Methodological guidelines for impedance cardiography," *Psychophysiology*, vol. 27, no. 1, pp. 1–23, 1990.
- [12] L.-Y. Shyu, Y.-S. Lin, C.-P. Liu, and W.-C. Hu, "The detection of impedance cardiogram characteristic points using wavelet transform," *Computers in biology and medicine*, vol. 34, no. 2, pp. 165–175, 2004.
- [13] J. Tomaka, J. Blascovich, R. M. Kelsey, and C. L. Leitten, "Subjective, physiological, and behavioral effects of threat and challenge appraisal," *Journal of Personality and Social Psychology*, vol. 65, no. 2, p. 248, 1993.
- [14] J. Tomaka, J. Blascovich, J. Kibler, and J. M. Ernst, "Cognitive and physiological antecedents of threat and challenge appraisal," *Journal of personality and social psychology*, vol. 73, no. 1, p. 63, 1997.
- [15] W. B. Mendes, B. Major, S. McCoy, and J. Blascovich, "How attributional ambiguity shapes physiological and emotional responses to social rejection and acceptance," *Journal of personality and social psychology*, vol. 94, no. 2, p. 278, 2008.
- [16] J. P. Jamieson, M. K. Nock, and W. B. Mendes, "Mind over matter: reappraising arousal improves cardiovascular and cognitive responses to stress," *Journal of Experimental Psychology: General*, vol. 141, no. 3, p. 417, 2012.
- [17] K. S. Quigley, L. F. Barrett, and S. Weinstein, "Cardiovascular patterns associated with threat and challenge appraisals: A within-subjects analysis," *Psychophysiology*, vol. 39, no. 3, pp. 292–302, 2002.
- [18] "A Biosignal-Specific Processing Tool," <http://www.northeastern.edu/ostadabbas/codes/>, accessed: 2017.
- [19] D. Tkach, H. Huang, and T. A. Kuiken, "Study of stability of time-domain features for electromyographic pattern recognition," *Journal of neuroengineering and rehabilitation*, vol. 7, no. 1, p. 1, 2010.
- [20] M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system," *Handbook of psychophysiology*, vol. 2, pp. 200–223, 2007.
- [21] M. S. Perez-Rosero, B. Rezaei, M. Akcakaya, and S. Ostadabbas, "Decoding emotional experiences through physiological signal processing," *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pp. 881–885, 2017.
- [22] J. Benedek and R. Hazlett, "Incorporating facial emg emotion measures as feedback in the software design process," *Proc. Human Computer Interaction Consortium*, 2005.
- [23] M. B. Reaz, M. Hussain, and F. Mohd-Yasin, "Techniques of emg signal analysis: detection, processing, classification and applications," *Biological procedures online*, vol. 8, no. 1, pp. 11–35, 2006.
- [24] P. Ekman, "Facial expression and emotion," *American psychologist*, vol. 48, no. 4, p. 384, 1993.
- [25] C. De Luca, L. Gilmore, M. Kuznetsov, and S. Roy, "Filtering the surface emg signal: Movement artifact and baseline noise contamination," *Journal of Biomechanics*, vol. 43, no. 8, pp. 1573–1579, 2010.
- [26] R. M. Kelsey and W. Guethlein, "An evaluation of the ensemble averaged impedance cardiogram," *Psychophysiology*, vol. 27, no. 1, pp. 24–33, 1990.
- [27] "Knowledge Base Impedance Cardiography," <https://support.mindwaretech.com/article-categories/imp/>, accessed: 2017.
- [28] R. M. Kelsey, S. Reiff, S. Wiens, T. R. Schneider, E. S. Mezzacappa, and W. Guethlein, "The ensemble-averaged impedance cardiogram: An evaluation of scoring methods and interrater reliability," *Psychophysiology*, vol. 35, no. 3, pp. 337–340, 1998.
- [29] J. R. Árbol, P. Perakakis, A. Garrido, J. L. Mata, M. C. Fernández-Santaella, and J. 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.
- [30] "Mathworks Classification Learner," <https://www.mathworks.com/products/statistics/classification-learner.html>, accessed: 2017.