2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Female–male Differences Should be Considered in Physical Pain Quantification based on Electrodermal Activity: Preliminary Study*

Youngsun Kong, Hugo F. Posada-Quintero, Ki H. Chon, *Senior Member, IEEE*

***Abstract*—Objective pain quantification is an important but difficult goal. Electrodermal activity (EDA) has been widely explored for this purpose, given its reported sensitivity to pain. However, cognitive stress can hinder successful estimation of physical pain when using EDA signals. We collected EDA signals from ten subjects (5 male and 5 female) undergoing pain stimulation, and calculated phasic, tonic, and frequency-domain features. Each subject experienced pain with and without stress. Three low and three high pain sessions were induced using two thermal grills (low-level for visual analog scale [VAS] 4 or 5 and high-level for VAS 7 or more). The Stroop test was performed for inducing cognitive stress. Significant differences between EDA features of painless and pain segments were observed. Significant differences between no pain and stress were also observed. Furthermore, we compared differences in EDA features between females and males under pain and cognitive stress. Frequency-domain EDA features of pain increased with stress for both females and males. Frequency-domain features derived from females also showed higher standard deviation than did those derived from males. We performed machine learning analysis and evaluated the models using leave-one-subject-out cross-validation. We obtained balanced accuracies of 63.5%, 72.4%, and 53.2% (combined, male, and female) when using training data of the same sex and 47.6%, 57.4%, and 42.7% (combined, male, and female) when using different sex for training.**

***Clinical Relevance*—Our preliminary results suggest that sex of patients should be considered to increase the accuracy of pain quantification based on EDA in the presence of cognitive stress.**

I. Introduction

Accurate detection of pain intensity is an important factor in clinical diagnosis. However, assessment of pain intensity has always been challenging due to its subjectiveness. To help communicate relative pain intensity, assessment tools have included numbers, color codes, or facial expressions [1], [2]. However, they do not provide continuous monitoring, they are still subjective, and they can be gamed. If pain could truly be assessed objectively and accurately, medication could be appropriately prescribed and the use of opioids could be better-controlled.

For these reasons, there have been many attempts to develop objective pain quantification methods using noninvasively measured physiological signals. As the sympathetic nervous system's response is the most sensitive to pain intensity [3]–[5], electrodermal activity (EDA) has been explored to detect and quantify pain [6]–[8], because it measures sympathetic-related sweat gland activations. However, other sympathetic-induced activities such as cognitive stress can affect EDA, hence hindering accurate detection of pain [9], [10]. It has also been observed that pain is influenced by stress [11]. Logan et al. found that pain caused by capsaicin during the Stroop test led to stress-induced analgesia in men but stressed-induced hyperalgesia in women (i.e., enhanced sensitivity to pain) [12]. Also, several studies showed significant difference in EDA signals between males and females [13], [14]. Therefore, we have intended to objectively estimate physical pain in the presence of stress in healthy subjects, and hypothesized that sex must be considered for more accurate pain quantification.

II. Methods

A. Experiments

Five females and five males, a total of 10 subjects (22-34 years old), were recruited. Shimmer 3 was used to collect EDA signals with electrodes placed on the index and middle fingers of the left hand (Shimmer, Dublin, Ireland). The experimental protocol consisted of a period of 1) 2-min baseline, 2) 2-min cognitive stress (the Stroop test), 3) pain-only stimulation, then 4) simultaneous cognitive stress and pain stimulation. For the Stroop test, subjects were asked to speak the color of the font shown on the screen, whose text spelled out a different color; the font color changed every 1-3 seconds. The accuracy of naming the correct color was not evaluated as we only aimed to induce cognitive stress. The colors of the font were red, blue, green, yellow, purple, and black.

We built two aluminum thermal grills to induce high and low pain levels without any tissue injury. The pipes of the thermal grills for the cold (icy) and warm (50-58 °C) water are interlaced so that the contrast in temperatures causes pain sensations without any tissue injury [15]. For low and high levels of pain, the warm water temperature was set to induce visual analogue scale (VAS) levels 4-5 and 7-8 out of 10, respectively, for each subject. After setting the water temperature, three low- and three high-pain levels were tested for 30-60 second randomized intervals. Subjects were asked to put their right hands on the thermal grill for five seconds or less until they could not bear the pain. VAS was reported for each pain stimulus. The study protocol was approved by the Institutional Review Board of the University of Connecticut.

Y. Kong, H.F. Posada-Quintero, and K.H. Chon are with the Department of Biomedical Engineering, University of Connecticut, Storrs, CT 06269 USA (corresponding author to provide phone: 959-777-2623; e-mail: youngsun.kong@uconn.edu, h.posada@uconn.edu, ki.chon@uconn.edu).

B. EDA features

1) Preprocessing

EDA signals were resampled to 4 Hz using a cubic spline algorithm from 120 Hz, the original data sampling rate. We then applied a median filter with a 1-sec window to remove motion artifacts. Finally, the signal was resampled to 2 Hz.

2) Time-Varying Index of Sympathetic Activity (TVSymp)

TVSymp was shown to be an effective quantitative index of sympathetic response elicited by cognitive stress and pain [6], [8], [9], [16]. First, EDA signals were normalized to unit variance. Then, dynamics of the signal in the frequency range between 0.08-0.24 Hz were extracted using the variable frequency complex demodulation technique [17]. The extracted components were reconstructed using the Hilbert transform to obtain the instantaneous amplitude of the signal. We also calculated modified TVSymp (MTVSymp) to minimize other sympathetic arousal from TVSymp $a(t)$ by the following equation [8]:

$$MTVSymp = max\left(0, a(t) - mean\left(a(t-10:t)\right)\right) \quad (1)$$

3) Phasic and Tonic Features

EDA signals can be decomposed into phasic and tonic components; they represent fast and slow dynamics of the signal, respectively. We chose cvxEDA as it showed in our previous studies better performance in classifying different pain levels than other methods [8], [18]. The cvxEDA also estimates phasic drivers that supposedly represent the underlying sympathetic activation. Finally, we calculated derivatives of the phasic (dPhEDA) and tonic (dTonEDA) components using the five-point stencil:

$$y(n) = \frac{x(n-2) - 8 \cdot x(n-1) + x \cdot P(n+1) - x(n+2)}{12 \cdot (1/fs)} \quad (2)$$

where the sampling frequency fs was set to 2 Hz. To summarize, we calculated phasic, tonic, phasic driver, dPhEDA, dTonEDA, TVSymp, MTVSymp indices as shown in Fig. 1.

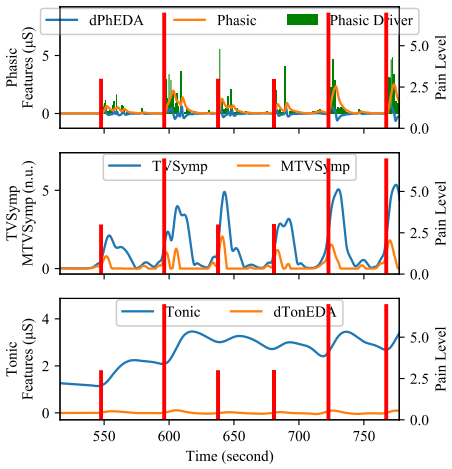

Figure 1. An example of EDA features with pain stimuli (red bars)

C. Statistics and Machine Learning

For statistical evaluation and machine learning, we put 30-second segments into one of the following categories: no pain, stress, low pain, low pain with stress, high pain, and high pain with stress. We calculated the mean and standard deviation of each of the feature indices for each category. Non-overlapping 30-second windows were extracted from 2-min cognitive stress and 2-min baseline for no pain and stress. 30-second windows after each pain stimulus were considered to represent pain data segments. Pain segments were categorized as low (if VAS was between 1-6) and high pain (VAS >=7). We also calculated significant difference between the categories for only both female and male subjects as the number of subjects was not enough to compare between females and males. First, normality was examined using the Kolmogorov–Smirnov test. Wilcoxon signed-rank test and t-test were then used for non-normally and normally distributed data, respectively.

We also performed machine learning with leave-one-subject-out cross-validation using a training set with: 1) the same sex as the test sets and 2) the opposite sex. We used mean, standard deviation, skewness, and kurtosis values of each EDA feature index. We examined support vector machine with linear kernel (SVM), random forest (RF) with 100 estimators and the Gini impurity criterion, and multi-layer perceptron (MLP) with the following parameters: 100 hidden units, rectifier with unit activation function, Adam optimizer, and 0.001 learning rate. The parameter C of SVM was chosen using the grid-search cross-validation with a group four-fold cross-validation consisting of 1, 10, 100, and 1000. We calculated feature importance using the same classifier methods consisting of SVM and RF, and RF for MLP with a grid-search cross-validation with a group four-fold cross-validation. Features that had greater importance than the average were selected for each fold.

III. Results

Table 1 shows the mean and standard deviation of EDA features. All EDA features except for dPhEDA and Tonic showed significant differences between no pain and stress. Phasic and phasic driver for the painless segments showed higher mean values than did those for the pain segments, while all other features for the pain segments exhibited higher mean values. Low pain with stress showed lower values for phasic features than without stress, while the mean values of high pain with stress were higher than those of high pain without stress. Phasic component and phasic driver of pain segments showed significant difference with those of painless segments except for low pain, which was significantly different with no pain only. Moreover, dPhEDA showed significant difference in values between painless segments and high pain with stress.

Tonic showed significant difference between painless segments and low pain with stress. No pain for dTonEDA showed significant differences with all other categories but low pain with stress. Pain segments of TVSymp and MTVSymp showed significant difference between painless and pain segments. Only MTVSymp showed a significant difference between low pain and high pain with stress. TVSymp and MTVSymp also exhibited increased mean values with stress for all categories (i.e., no pain, low pain, and high pain with stress showed higher mean values of the features than each of those without stress.)

TABLE I. MEAN ± STANDARD DEVIATION OF EDA FEATURES

	Segments		Painless Segments		Low Pain Segments		High Pain Segments	
	Category		No Pain^{0}	Stress^{0+}	Low Pain^{1}	Low Pain^{1+} w/ Stress	High Pain^{2}	High Pain^{2+} w/ Stress
Phasic Features	Phasic	All	0.917±2.561	0.601±1.989 ^{0}	0.395±0.892 ^{0}	0.186±0.167 $^{0,0+}$	0.552±1.086 $^{0,0+}$	0.967±1.586 $^{0,0+}$
		Male	0.012±0.018	0.040±0.028	0.188±0.078	0.267±0.160	0.304±0.206	0.257±0.182
		Female	1.821±3.389	1.222±2.756	0.644±1.276	0.052±0.051	0.751±1.414	1.361±1.860
	Phasic Driver	All	0.702±1.964	0.458±1.514 ^{0}	0.312±0.709 ^{0}	0.146±0.133 $^{0,0+}$	0.427±0.840 $^{0,0+}$	0.766±1.247 $^{0,0+}$
		Male	0.008±0.010	0.031±0.023	0.145±0.061	0.211±0.127	0.235±0.160	0.207±0.150
		Female	1.396±2.599	0.930±2.098	0.512±1.014	0.038±0.037	0.582±1.094	1.076±1.462
	dPhEDA	All	-0.001±0.004	-0.003±0.012	0.004±0.011	0.001±0.005	0.003±0.010	0.012±0.024 $^{0,0+}$
		Male	-0.001±0.002	0.000±0.002	0.000±0.003	0.002±0.006	0.000±0.001	0.004±0.005
		Female	-0.001±0.005	-0.006±0.017	0.008±0.016	-0.001±0.002	0.005±0.014	0.016±0.028
Tonic Features	Tonic	All	1.622±0.731	1.763±0.818	2.917±2.710	2.823±1.427 $^{0,0+}$	3.008±3.120	4.746±4.745
		Male	1.755±0.643	1.830±0.579	2.686±0.786	3.443±1.116	2.610±0.706	3.029±1.535
		Female	1.489±0.787	1.688±1.014	3.194±3.908	1.791±1.285	3.326±4.111	5.700±5.583
	dTonEDA	All	-0.001±0.004	0.004±0.006 ^{0}	0.006±0.014 ^{0}	0.003±0.009	0.009±0.010 ^{0}	0.001±0.017 ^{0}
		Male	-0.003±0.004	0.003±0.006	0.005±0.010	0.003±0.010	0.015±0.011	0.008±0.011
		Female	0.000±0.003	0.004±0.007	0.007±0.018	0.003±0.004	0.005±0.005	-0.002±0.019
Frequency domain Features	TVSymp	All	0.243±0.444	0.591±0.769 ^{0}	1.182±0.641 $^{0,0+}$	1.368±0.621 $^{0,0+}$	1.348±0.716 $^{0,0+}$	1.631±0.653 $^{0,0+}$
		Male	0.186±0.240	0.454±0.283	1.591±0.373	1.636±0.442	1.880±0.426	1.975±0.489
		Female	0.300±0.575	0.743±1.054	0.691±0.542	0.922±0.620	0.922±0.608	1.440±0.654
	MTVSymp	All	0.040±0.090	0.102±0.130 ^{0}	0.215±0.122 $^{0,0+}$	0.271±0.147 $^{0,0+}$	0.244±0.140 $^{0,0+}$	0.325±0.141 $^{0,0+,1}$
		Male	0.027±0.038	0.081±0.051	0.295±0.084	0.325±0.126	0.317±0.110	0.408±0.110
		Female	0.052±0.120	0.127±0.177	0.120±0.087	0.183±0.137	0.186±0.134	0.279±0.136

Superscript indices indicate which segments show significant difference with that level for that approach ($p < 0.05$). Significant difference was only calculated for all subjects (both female and male).

The phasic features of females showed higher standard deviation than did those of males except for low pain with stress. Tonic, TVSymp, and MTVSymp exhibited higher standard deviation for females than for males. TVSymp and MTVSymp derived from females showed greater mean during painless segments than did those of males, while male subjects exhibited greater mean for pain segments.

Table 2 shows comparison of classification results between female and male subjects. Balanced accuracies were higher when the same sex was used for training except for a case. Balanced accuracy values of 63.5% and 53.2% were obtained for the same and different sex, respectively, when both stress and no stress datasets were used. For male and female subjects, balanced accuracies of 72.4% and 53.2% were obtained, respectively, as shown in Table 3.

TABLE II. CLASSIFICATION RESULTS

	Train Dataset		No Stress	Stress	All		
	Test Dataset		No Stress	Stress	No Stress	Stress	All
Same Sex between Train and Test data	*All*	*N*	120	120	120	120	240
		SVM	0.562	0.576	0.513	0.639	0.576
		RF	0.563	0.613	0.560	0.591	0.574
		MLP	0.633	0.529	0.624	0.651	0.635
	Male	*N*	60	60	60	60	120
		SVM	0.731	0.634	0.602	0.651	0.624
		RF	0.694	0.601	0.685	0.602	0.644
		MLP	0.833	0.651	0.704	0.734	0.724
	Female	*N*	60	60	60	60	120
		SVM	0.378	0.505	0.411	0.612	0.507
		RF	0.419	0.537	0.419	0.492	0.457
		MLP	0.433	0.415	0.544	0.528	0.532
Different Sex between Train and Test data	*All*	*N*	120	120	120	120	240
		SVM	0.462	0.438	0.421	0.461	0.440
		RF	0.487	0.443	0.510	0.430	0.468
		MLP	0.422	0.462	0.423	0.531	0.476
	Male	*N*	60	60	60	60	120
		SVM	0.630	0.456	0.556	0.459	0.512
		RF	0.583	0.589	0.648	0.490	0.565
		MLP	0.565	0.639	0.519	0.623	0.574
	Female	*N*	60	60	60	60	120
		SVM	0.311	0.432	0.311	0.460	0.388
		RF	0.426	0.410	0.404	0.448	0.427
		MLP	0.307	0.339	0.352	0.469	0.409

N = the number of samples

IV. DISCUSSIONS

TABLE III. CONFUSION MATRICES OF FEMALE AND MALE SUBJECTS

MLP		Predicted Pain					
		Female (53.2%)			Male (72.4%)		
		No	Low	High	No	Low	High
True Pain	No	64.9%	21.6%	13.5%	97.4%	2.6%	0.0%
	Low	22.2%	37.0%	40.7%	7.9%	60.5%	31.6%
	High	18.2%	24.2%	57.6%	4.5%	36.4%	59.1%

Many factors can affect sympathetic nervous activities and EDA signals. We tested the feasibility of pain detection in the presence of cognitive stress, and how including sex differentiation in the model improved the performance of the models. All features except for tonic features showed higher values for high pain than for low pain. Also, females and males showed different trends and distributions of EDA features throughout the pain and stress categories, as shown in Table 1 and Figure 2. Females showed higher standard deviation than males in most EDA features, which resulted in poorer performance of classifiers, as shown in Figure 3. We observed higher balanced accuracies of most classifiers when the same sex was used for training classifiers.

We induced cognitive stress using the Stroop task. dTonEDA and the frequency-domain features (TVSymp and MTVSymp) for pain with stress showed higher and lower mean values, respectively, than for pain without stress. EDA features for no pain segments were significantly different from at least one of pain segments. Notably, MTVSymp showed a significant difference between low and high pain

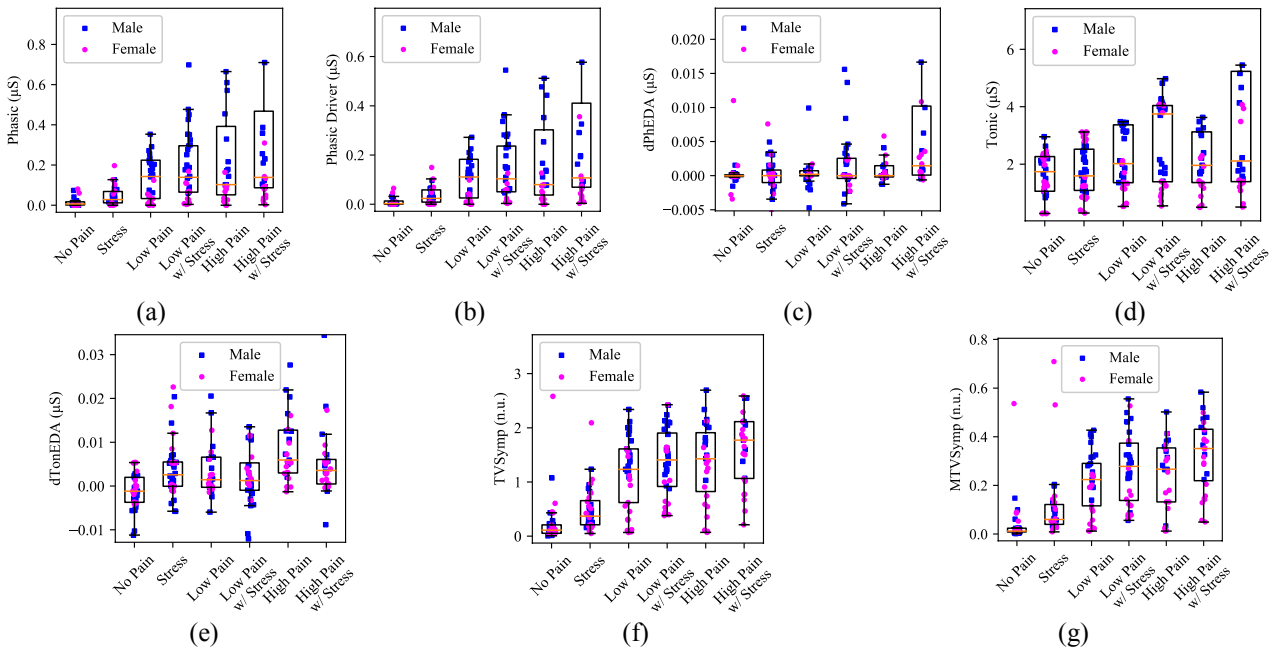

Figure 2. Box plots of (a) Phasic, (b) Phasic Driver, (c) dPhEDA, (d) Tonic, (e) dTonEDA, (f) TVSymp, and (g) MTVSymp.

with stress. dTonEDA between low pain with stress and high pain was also marginally significant ($p = 0.051$). This resulted in poorer classification performance for low and high pain, compared to that for no pain, as shown in Table 3.

The main limitation of the study is the low number of subjects, as this study is preliminary. Moreover, more factors such as age and race might affect pain and stress [19], [20]. Finally, different pain behaviors can be observed based on the types of stress (e.g., stress-induced analgesia or stress-induced hyperalgesia) [11]. These must be considered in the future for more accurate classifications.

REFERENCES

[1] M. J. Hjermstad *et al.*, "Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review," *J. Pain Symptom Manage.*, vol. 41, no. 6, pp. 1073–1093, 2011.

[2] M. P. Jensen and P. Karoly, "Self-report scales and procedures for assessing pain in adults.," 2011.

[3] D. R. Bach, "Sympathetic nerve activity can be estimated from skin conductance responses—A comment on Henderson et al.(2012)," *NeuroImage*, vol. 84, pp. 122–123, 2014.

[4] M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system.," 2017.

[5] H. F. Posada-Quintero and K. H. Chon, "Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review," *Sensors*, vol. 20, no. 2, p. 479, 2020.

[6] Y. Kong, H. F. Posada-Quintero, and K. H. Chon, "Pain Detection using a Smartphone in Real Time," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 4526–4529.

[7] D. Lopez-Martinez and R. Picard, "Continuous pain intensity estimation from autonomic signals with recurrent neural networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 5624–5627.

[8] Y. Kong, K. H. Chon, and H. F. Posada-Quintero, "Sensitive Physiological Indices of Pain based on Differential Characteristics of Electrodermal Activity," *IEEE Trans. Biomed. Eng. (forthcoming).*, 2021.

[9] H. F. Posada-Quintero, J. P. Florian, Á. D. Orjuela-Cañón, and K. H. Chon, "Highly sensitive index of sympathetic activity based on time-frequency spectral analysis of electrodermal activity," *Am. J. Physiol.-Regul. Integr. Comp. Physiol.*, vol. 311, no. 3, pp. R582–R591, Jul. 2016, doi: 10.1152/ajpregu.00180.2016.

[10] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón, T. Aljama-Corrales, S. Charleston-Villalobos, and K. H. Chon, "Power spectral density analysis of electrodermal activity for sympathetic function assessment," *Ann. Biomed. Eng.*, vol. 44, no. 10, pp. 3124–3135, 2016.

[11] A. H. Ahmad and R. Zakaria, "Pain in times of stress," *Malays. J. Med. Sci. MJMS*, vol. 22, no. Spec Issue, p. 52, 2015.

[12] H. Logan, S. Lutgendorf, P. Rainville, D. Sheffield, K. Iverson, and D. Lubaroff, "Effects of stress and relaxation on capsaicin-induced pain," *J. Pain*, vol. 2, no. 3, pp. 160–170, 2001.

[13] J. Buthmann, J. Finik, and Y. Nomura, "Sex differences in the relations between infant temperament and electrodermal responses in early childhood," *Int. J. Behav. Dev.*, vol. 42, no. 6, pp. 535–542, 2018.

[14] D. S. Bari, "Gender differences in tonic and phasic electrodermal activity components," *Sci. J. Univ. Zakho*, vol. 8, no. 1, pp. 29–33, 2020.

[15] H. F. Posada-Quintero *et al.*, "Using electrodermal activity to validate multilevel pain stimulation in healthy volunteers evoked by thermal grills," *Am. J. Physiol.-Regul. Integr. Comp. Physiol.*, vol. 319, no. 3, pp. R366–R375, 2020.

[16] H. F. Posada-Quintero, Jeffrey B. Bolkhovsky, Michael Qin, and Ki H. Chon, "Human Performance Deterioration Due to Prolonged Wakefulness Can Be Accurately Detected Using Time-Varying Spectral Analysis of Electrodermal Activity," *Hum. Factors*, vol. 60, no. 7, pp. 1035–1047, Nov. 2018

[17] H. Wang, K. Siu, K. Ju, and and K. H. Chon, "A High Resolution Approach to Estimating Time-Frequency Spectra and Their Amplitudes," *Ann Biomed Eng*, vol. 34, no. 2, pp. 326–338, Feb. 2006.

[18] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 797–804, 2015.

[19] P. A. Parmelee, T. L. Harralson, J. A. McPherron, J. DeCoster, and H. R. Schumacher, "Pain, disability, and depression in osteoarthritis: effects of race and sex," *J. Aging Health*, vol. 24, no. 1, pp. 168–187, 2012.

[20] K. M. Woodrow, G. D. Friedman, A. B. Siegelaub, and M. F. Collen, "Pain tolerance: differences according to age, sex and race," *Psychosom. Med.*, vol. 34, no. 6, pp. 548–556, 1972.