Daily Stress Prediction Using Heart Rate Variability Metrics

Whitepaper by OMsignal
(Dated: Nov. 29, 2017)

***Abstract*—The OMsignal heart rate measurement algorithms were demonstrated to be accurate within 5 beats per minute during running and jogging. This accuracy was verified for 95% of the recording time with 95% statistical certainty. Data for the verification was taken from 20 recordings by women and men wearing the OMsignal bio-sensing garments. In this work, an accurate ECG-based daily mental stress level prediction strategy is presented. Multiple support vector machines (SVM) with linear kernel functions are individually trained to predict daily stress levels of women who participated in the OM- signal MyHeart project. In this study, participants are asked to answer a daily survey to determine the quality of their sleep, exercise, valence, control and rumination during the last 24-hour. Using the aforementioned items, a daily stress score was defined to be used as the target value for constructing the stress prediction model. The model is designed by the use of heart rate variability (HRV) metrics calculated from a 5-minute data window moving over daily ECG recordings. A 30 dimensional feature vector, including the first five minimum and maximum values of SDNN and RMSSD (two popular HRV metrics) as well as heart rate is extracted to represent each individual daily ECG record. The leave-one-out cross- validation method is used to train and validate our user-dependent SVM model. On validation data, an average accuracy of 85.26% is achieved for predicting daily stress scores of the users with sufficient number of daily survey data.**

I. Introduction

Chronic stress is one of the major risk factors of coronary heart disease and hypertension. When there is an acute stress, the body sympathetic system is activated to increase heart rate. The adrenal gland starts secreting high level of cortisol. When stress is stopped, the parasympathetic system takes over to decrease heart rate, sweating and breathing rate. To monitor chronic stress in daily life, heart rate variability (HRV) has been widely accepted in the literature as one of the popular physiological stress indicators. HRV variations due to the changes in mental stress levels are subjective to individuals. This difference among the individuals may arise due to different body conditions, gender, age, physical fitness and emotional states.

To study the effect of mental stress on HRV reduction, multiple HRV metrics in time and frequency domains are calculated with respect to the body position. HRV metrics are calculated for segments of ECG records with sufficient RR coverage after removing ectopic beats, which are defined as when one RR interval differs from the previous one by more than 20%. The metrics are calculated from a 5-minute data window moving along ECG recordings captured by OM apparel while the participants go about their daily activities. The participants are 30 women aged between 40 - 60 years. SDNN (standard deviation of normal RR intervals) and RMSSD (root mean square of SDNN) are commonly used in the literature to quantify HRV and to monitor stress level.

It is important to also consider stable and transient potential confounding variables influencing HRV to exclude participants prior to data collection or to understand the outliers within the data post collection. In this study, stable confounding variables such as habitual levels of alcohol and coffee consumption as well as smoking, weight, body size, list of medications and medical history are initially reported by the participants. Moreover, transient confounding variables such as the amount of sleep and exercise during the last 24-hour are reported by the participants through a daily questionnaire in the MyHeart App.

In this work, multiple features, namely SDNN, RMSSD and mean heart rate are extracted to design a user-dependent daily stress prediction model using a support vector machine (SVM). For this purpose, a daily stress score is defined for each participant using daily survey questionnaire answers. The average ac- curacy of daily stress score prediction over validation data for users with sufficient number of daily survey data is 85.26%.

II. Daily Questionnaire

To study the correlation of HRV metrics (mainly SDNN and RMSSD) with daily stress load, a daily questionnaire was designed in the MyHeart App in which five questions were answered by the participants. Two questions were used to quantify the amounts of sleep and exercise in the last 24-hour period and the last three questions were used for estimating participants' daily stress level. The following three stress related questions were respectively used to measure the valence, control and rumination factors on a daily basis.

1) Did you experience something emotionally in- tense in the last 24 hours?
2) I currently feel that I am on top of things
3) Today, I am thinking constantly about specific issues or problems

The daily stress score is defined by the use of the last two stress-related questions that are on a five-point likert scale.

Since control is negatively correlated to daily stress, the control score should be inverted to obtain daily stress score as follows:

$$\text{Stress Score} = \text{Inverted Control} + \text{Rumination} \quad (1)$$

where the inverted control is scaled from 5 to 1. Hence, the daily stress score range is between 2 to 10. Basically, high rumination combined with low control is an indicator of high stress days, whereas either high control or low rumination can be a sign of a positive day.

TABLE I. AVERAGE ACCURACY PERCENTAGE RATES OF USER-DEPENDENT SVM-BASED STRESS PREDICTION MODELS OVER TRAINING AND VALIDATION DATA SETS FOR MYHEART PARTICIPANTS.

User ID	# Daily Records	Training	Validation
1	11	100	92.73
2	10	98.04	86.00
3	13	90.75	80.00
4	17	71.39	55.24
5	13	73.97	60.00
6	10	99.88	86.00
7	14	94.20	85.71
8	6	100.00	100.00
9	7	97.06	80.00
10	7	100.00	100.00
11	7	100.00	94.29
12	6	100.00	93.33

III. STRESS PREDICTION MODEL

A fairly accurate stress prediction model is designed to predict daily stress score using multiple features extracted from participants' ECG recordings. The model is user-dependent and is trained separately for each participant. For each individual a 30 dimensional feature vector including the first five minimum and maximum values of SDNN, RMSSD and mean HR are extracted. Different approaches namely as support vector machines (SVM) with various kernel functions and feed-forward neural networks are investigated and their performances are similar. To evaluate the performance of our proposed classifier, a leave-one-out cross-validation is used. The aver- age accuracy over training and validation data is displayed in TABLE I for each individual MyHeart user. The number of daily records is equivalent to the number of days in which the user answered the daily questionnaire and it is also possible to calculate at least 10 HRV samples per those days considering the user acceleration intensity and data quality metrics.

Users 4 and 5 are the only subjects for which our stress prediction model is not quite accurate due to certain reasons that will be explained later in this section. Hence, excluding the aforementioned users, the average accuracy for other participants are sufficiently high. The accuracy associated with each subject depends on the distribution of the training data, the amount of available data for different daily stress scores as well as the emotional awareness of the user while answering daily survey. Moreover, the accuracy is only reported for those participants who have already answered daily surveys and were not excluded from our study due to certain medical conditions. The histogram of the users data for different daily stress scores is displayed in Fig. 1. The statistical information of user stress score data distributions is also summarized in TABLE II.

TABLE II. STATISTICAL FEATURES OF MYHEART USERS DAILY STRESS SCORE DATA

User ID	Minimum SCore	Maximum Score	#Distinct Scores
1	4	6	3
2	2	4	3
3	4	6	3
4	2	4	3
5	3	8	6
6	2	4	3
7	2	7	6
8	2	6	4
9	3	7	3
10	4	7	4
11	3	5	3
12	2	8	4

User 5 has the highest number of bins (distinct daily stress scores) with insufficient data per score which may explain the poor prediction accuracy indicated in Table 1. The prediction accuracy is expected to be improved by collecting more data from user 5. However, user 4 has only reported three distinct stress scores with sufficient data for each which does not support receiving poor prediction accuracy on her validation data. One possible reason could be an insufficient level of emotional awareness which causes this user to incorrectly report her daily stress.

IV. CONCLUSION

Achieving a fairly high accuracy on daily stress prediction using the HRV metrics extracted from ECG demonstrates that OM garments provides a sufficiently high quality signal for this purpose in a lifestyle context. Moreover, utilizing machine learning techniques do enable us to identify different ECG patterns that can be used for multiple intelligent applications and environments.

Fig. 1. Distribution of data associated with different daily stress scores for the users in TABLE I

.