Journal of Korea Multimedia Society Vol. 20, No. 8, August 2017(pp. 1395-1405)
https://doi.org/10.9717/kmms.2017.20.8.1395

Stress Identification and Analysis using Observed Heart Beat Data from Smart HRM Sensor Device

SPL Aditya Pramanta[†], Myonghee Kim[††], Man-Gon Park[†††]

ABSTRACT

In this paper, we analyses heart beat data to identify subjects stress state (binary) using heart rate variability (HRV) features extracted from heart beat data of the subjects and implement supervised machine learning techniques to create the mental stress classifier. There are four steps need to be done: data acquisition, data processing (HRV analysis), features selection, and machine learning, before doing performance measurement. There are 56 features generated from the HRV Analysis module with several of them are selected (using own algorithm) after computing the Pearson Correlation Matrix (p - values). The results of the list of selected features compared with all features data are compared by its model error after training using several machine learning techniques: support vector machine, decision tree, and discriminant analysis. SVM model and decision tree model with using selected features shows close results compared to using all recording by only 1% difference. Meanwhile, the discriminant analysis differs about 5%. All the machine learning method used in this works have 90% maximum average accuracy.

Key words: Stress, HRV, Machine Learning

1. INTRODUCTION

Nowadays, as the growth of internet demand increases, the technology is shifted towards Internet of Things (IoT). A smartphone is a common example of the well-known product of IoT. It helps people to connect to each other thus creating a vast link of digital information. This huge amount of information then refers as Big Data.

Plenty of researchers works to process the Big Data in order to discover a pattern or invisible solution of human's problem. This race among researchers outcome is the data analyzer technique called machine learning. As its name implies, this method made the machine (computer, mostly) to "learn" like what human did. Researcher inspired by the biology of human in the development of machine learning.

The combination of IoT and machine learning are being used in many ways. Several sectors which needs fast prediction will mainly utilize this system for example, healthcare system. Doctors and engineers collaborate to create an easy monitoring and diagnosis system in order to prevent the disease symptoms to appear or to observe the patient's condition. The IoT component used here are sensors to acquire data from the patients in the term of the physical phenomenon for example blood pressure (BP), galvanic skin response (GSR), body temperature, and heart rate (HR).

Commercial IoT sensors are available after developer finds another application of the IoT in

※ Corresponding Author : Man-Gon Park, Address : (48513) Yongso-Ro 45, Nam-Gu. Busan, Rep. of Korea, TEL: +82- 51-629-6240, FAX : +82-51-629-6230, E-mail : mpark@pknu.ac.kr
Receipt date : Mar. 31, 2017, Revision date : Jun. 26, 2017
Approval date : Jul. 3, 2017

[†] School of Electrical Engineering, Institute of Technology Bandung, Indonesia
(E-mail : spladityapramanta@gmail.com)
[††] Dept. of IT Convergence and Application Engineering, PuKyong Nat. Univ., Rep. of Korea
(E-mail: mhgold@pknu.ac.kr)
[†††] Dept. of IT Convergence and Application Engineering, PuKyong Nat. Univ., Rep. of Korea

healthcare as a business prospect. Most of them are used to track user's heart rate while they are exercising so they are called the heart rate monitor (HRM) device. Several studies found that the performance of those HRMs is competing with the electrocardiogram (ECG) measurement. Although ECG still remains the best in acquiring heart beat data, the mobility, simplicity, and relatively cheap drives researcher to try implementing HRM as disease analyzer tools.

Mental stress identification has a promising potential to be solved using those systems mentioned above because this "disease" or problem affect the sympathetic nervous systems (SNS) work. SNS is the part of the autonomic nervous system (ANS) and responsible for the variation of the heart beat. Even though mental stress is considered to be physiological disease, the way it disturbs the physical system should not be ignored. Therefore, simple and fast detection will help to prevent the occurrence of the disease.

2. Related Works

The very first documented studies about the beat-to-beat variability of heart rate (HR) was seeking out the correlation of HR and fetal death in 1963 [16]. Since this finding, the term "variability" was being used widely to obtain the relationship between the HR and certain disease or health-related problem. In last decade, there were more than 5000 publications on Heart Rate Variability (HRV) showing that this technique has a promising solution [1].

The main source of HRV analysis is derived from RR interval data of heart beat (mainly could be acquired by using ECG recording). The RR interval data can be classified into short-term recording (5 minutes long minimum) and long-term recording (24 hours minimum) [10]. The length of recording determines the analysis used. Time-domain analysis (statistical) typically needs the long-term recording meanwhile frequency -domain (spectral) could be performed well using a short-term recording. Generally, it is not appropriate to compare the result of HRV analysis obtained from different recording durations [17].

Another issue with recording's duration is stationary of the recording. In the long recording, it is inevitable that subject's physical condition is changed thus influence the recording. A possible solution is by dividing the recording into shorter segments. Besides, preprocessing the RR interval should also be considered as it may contain noise, ectopic beats, and arrhythmias (artifacts). Kauffman [19] in his research conclude that even single artifact will possibly influence the results of HRV analysis. Therefore, they present a software tool for processing ECG and Inter - Beat Interval (IBI) data from pre - processing until performing HRV analysis called ARTiiFACT (MATLAB).

Valderlei [23] compared the recording results of Polar S810i HRV (rMSSD, pNN50, LFnu, HRnu, LF/HF) at rest and during exercise with LYNX signal conditioner (BIO EMG 1000 model, ECG signal) [8]. The results show that Polar S80i is able to obtain data as reliable as those captured using ECG signal. Similarly, Nunan reported that even in short - term HRV analysis (10 minutes maximum recording duration), Polar S810i showed a decent result. Additionally, Luiz who also did short - time comparison (but with time - domain HRV only) provided identical results [24] as reported by Nunan. However, after those research, Polar S80i had not produced again.

The smart HRM sensor device Polar H7 was used to assist the Polar V800 in 2015 assessment of the validity of the RR interval and short-term HRV (linear and non - linear parameter) data obtain from V800 by D. Giles, N. Draper, and W. Neil [26]. The study used intra - class correlation coefficient (ICC), Bland - Altman limits of agreement (LoA), and effect size (ES) as parameters to compare the ECG and PolarV800 data. The study found that Polar V800 has an ability to produce RR interval recording which is consistent compared to ECG. In addition, HRV parameters derived were also remarkably comparable. However, as V800 is really expensive smart watch and the purpose of the using V800 was to collect the data from Polar H7. Therefore, it can be concluded that by using Polar H7 only, similar study's results (with the studies mentioned above) might possibly be obtained.

Numerous studies had used the combination of machine learning and HRV analysis to classify subjects into healthy or not healthy in hope for an accurate and agile diagnosis. Support vector machine (SVM) and radial basis function (RBF) were used to classify patients who died because of Chagas disease by Sady [27, 30]. They gathered the data from 24 hours monitoring and extract time - frequency indices of HRV as the input of the machine learning system (SVM and RBF). In that study, they found that it is possible to distinguish between two classes (binary classification) with the accuracy of 95%.

Stress monitoring and recognition also feasible to be performed using HRV parameter and machine learning methods. However, as the studies differ in experiment environment, scheme, and methods they obtain relatively different results of which HRV parameters are related to stress. It is due to lack of fundamental theory of stress and also the proper and internationally agreed experiment's standard. Regardless of those facts, those studies show promising results by average of 80% binary classification of stress and non-stress subject [1-8], [12-15].

Machine learning could perform better when we have a reliable and trustworthy data base. Various studies reported the usage of PhysioNet database as their training and test data for the machine learning system they proposed. The difficulty in finding the medical data from hospital history is the main cause of the lack of data source. Most of the data available in PhysioNet are heart - related diseases and in the form of ECG data recording. However, there is only one data set for stress related problem. That data is taken from car driving condition of the subjects.

3. System Design

The system will have four process modules: data acquisition, HRV analysis, feature selection, and supervised machine learning by order as shown in Fig. 1. The output of the prior module will be input for the next module.

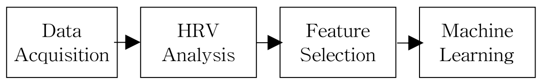

Fig. 1. Overall system design process modules.

Table 1. Overall system brief description

Module	Input	Process	Output
Data Acquisition	Raw Heart Beat Data	1. Record the raw heart beat data of subject 2. Generate .csv file with heart rate data 3. Generate (.ibi) file from .csv file	Inter-beat interval recording data (.ibi) file
HRV Analysis	.ibi files	1. Preprocess the (.ibi) file: correct the ectopic beat, detrending, and resampling 2. Extract all HRV parameter	All HRV parameters
Feature Selection	All HRV parameters from all recording	1. Apply correlation analysis. 2. Select the uncorrelated features.	1. Correlation Matrix 2. Selected Features
Machine Learning	1. All HRV features 2. Selected features	[1] Create and train the models [2] Measure the accuracy	Classifiers model and accuracy

The input - output and description process for each module are listed in the Table 1 below:

3.1 Data Acquisition

As the system will binary classify subjects into two states stress and not stress, two types of data are needed also. Table 2 will describe more about the methods used to collect those needed data:

Table 2. Data Acquisition Methods and Requirement

	Stress State	Non-Stress State (Relax)
Task	Subjects play several brain games from Lumosity web apps.	Subjects listen to two of their lately favorite music.
Recording Duration	More than 5 minutes	Range from 5 minutes to 10 minutes

3.2 HRV Analysis

Before doing the HRV analysis, we should detect and correct those ectopic beats. Besides, frequency domain analysis requires data to be stationary and randomly sampled in prior [9]. Thus, signal detrending and signal resampling are also compulsory in preprocessing steps. This Fig. 2 shows design for the pre-processing part.

The preprocessed signal then become the input of HRV analysis module. There are three types of analysis used: time-domain analysis, frequency - domain analysis, and non - linear analysis. In frequency - domain analysis, we use three methods to estimate the power spectral density (PSD) of the pre - processed heart beat signal: Welch, autoregressive (AR), and Lomb. The formula for estimating PSD for each method are given as follows [9]:

1. $P_{Welch}(f) = \frac{1}{N}\sum_{i=1}^{N-1} P_{M,i}(f)$, , where $P_{M,i}(f)$ is the i^{th} modified periodogram from the pre-processed heart beat signal.

2. $P_{AR}(f) = \frac{1}{f_s}\frac{\epsilon_p}{\left|1+\sum_{k=1}^{p} a_p(k) e^{-\frac{2\pi jkf}{f_s}}\right|^2}$,, where ϵ_p is total least square error, f_s is sample rate, a_p is Burg AR model parameters.

3. $P_{LS}(f) \equiv \frac{1}{2\sigma^2}\left\{\frac{\left[\sum_{n=1}^{N}(X(t_n)-\overline{X})\cos(2\pi f(t_n-\tau))\right]^2}{\sum_{n=1}^{N}\cos^2(2\pi f(t_n-\tau))}+\frac{\left[\sum_{n=1}^{N}(X(t_n)-\overline{X})\sin(2\pi f(t_n-\tau))\right]^2}{\sum_{n=1}^{N}\sin^2(2\pi f(t_n-\tau))}\right\}$,

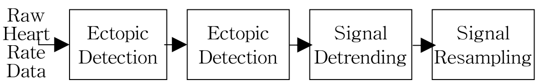

Fig. 2. Pre-processing steps.

where $\overline{X}$ and σ^2 are the mean and variance of the time series, and $\tau \equiv \tan^{-1}\left(\frac{\sum_{n=1}^{N}\sin(4\pi f t_n)}{\sum_{n=1}^{N}\cos(4\pi f t_n)}\right)$ is a frequency dependent time delay defined to make the periodogram insensitive to time shift.

Table 3 below show the features generated from each analysis mentioned above:

Table 3. HRV Features

Parameter	Unit	Formula	Description
Time Domain Analysis			
Mean RR	ms	$\frac{\sum_{i=1}^{N}(RR_i)}{N}$	The mean of RR intervals
SDNN	ms	$\sqrt{\frac{1}{N-1}\sum_{n=2}^{N}(RR_i-\overline{RR})^2}$	Standard deviation of RR intervals
RMSSD	ms	$\sqrt{mean((RR_{i+1}-RR_i)^2)}$	Square root of the mean squared differences between successive RR intervals
NN50	count	$count(\lvert RR_{i+1}-RR_i\rvert)_{>50ms}$	Number of successive RR interval pairs that differ more than 50 ms
pNN50	%	$\frac{NN50}{N-1}\times 100$	NN50 divided by total number of RR interval (percentage)
Frequency Domain Analysis			
VLF power	ms2	Power spectrum from 0.003 - 0.04 Hz	Absolute powers of very low frequency band
LF power	ms2	Power spectrum from 0.04 - 0.15 Hz	Absolute powers of low frequency band
HF power	ms2	Power spectrum from 0.15 - 0.4 Hz	Absolute power of high frequency band
LF / HF ratio	-	$\frac{P_{LF}}{P_{HF}}$	Ratio between LF and HF band power
Nonlinear analysis : Poincare Plot			
SD1, SD2	ms	SD1 : the standard deviation of the Poincare Plot (PP) perpendicular to the line of identity SD2 : the standard deviation of the PP along to the line of identity	Short - term and long - term variability standard deviation. Relation between consecutive RR interval (plot of RRi+1 as a function of RRi)
Nonlinear analysis: Sample Entropy			
SampEn	-	$-\log\frac{d[X_{m+1}(i), X_{m+1}(j)]}{d[X_m(i), X_m(j)]}$	Sample entropy, quantifies extent to which a sequence of m RR intervals can predict the next RR interval duration, based on the knowledge of degree of the similarity for sequences of length m to that for sequence of length m + 1.
Nonlinear analysis: Detrended Fluctuation Analysis			
α_1, α_2	-	$y(k)=\sum_{j=1}^{k}(RR_j-\overline{RR})$, $k=1,\ldots,N.$	Short - term and long - term fluctuations. Using RMS(root - mean - square) from the integrated version of the original time series to quantify the time series in fractal form or self -

		$F(n) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\gamma(k) - \gamma_n(k$	similar properties of non - stationary time series. (Longer time series data). α_1: slope of regression line relating log(F(n)) to log(n) with n within 4-16. α_2: slope of regression line relating log(F(n)) to log(n) with n within 16 - 64.

3.3 Feature Selection

Since we are looking for the association between variables and HRV features are consist of either interval or ratio data, then we will choose the parametric method which is Pearson's correlation coefficient. As shown in Fig. 3, this algorithm is used to select the most irredundant variables from the Pearson matrix correlation result.

3.4 Supervised Machine Learning

All HRV parameter and selected features are the input for support vector machine(SVM), decision tree, and discriminant analysis used in this module. Besides, because the possibly encountered problem is over fitting of data due to lack number of sample, we implemented the *k - cross validation* method in this research as over fitting prevention method.

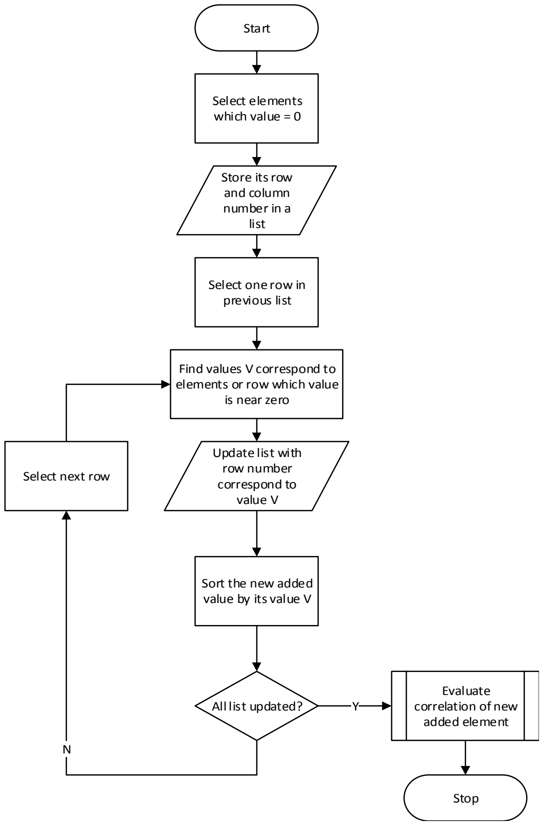

Fig. 3. Main algorithm to select the features based on Pearson matrix correlation result.

4. IMPLEMENTATION AND RESULTS

4.1 Overview of the System

After reviewing all games available in Lumosity, there are five games selected based on their simple to follow game mechanics, and based on their aspect of assessment of different area in brain: "Speed Match", "Memory Match", "Lost in Migration", "Brain Shift", and "Chalkboard Challenge". Besides, the mean play time of the total game is 5 minutes, enough for the recording requirement. To acquire data from Polar, we utilize the BLE Sensor Connect apps from GitHub (and also available in free in apps store) and our Dropbox as database gathering point. The output of that apps is .csv file needs to be formatted into .ibi file. Here we create MATLAB code to convertthe .csv files of each subject to be .ibi files. In addition, starting from HRV Analysis module to Machine Learning module, the implementation is being made in MATLAB using MATLAB script and HRVAS open source code as a reference.

4.2 Results

There are 21 data gathered from 5 men subjects for both stress and non-stress states. Fig. 4 is shown preprocessing results of theraw heart beat data of one subject:

Based on the figure above, the ectopic can be detected by seeing the fluctuation of the signal in time t compared to the signal in time t - 1. Cubic spline interpolation technique is used to replace the detected ectopic beat. In 3.Detrending, the orange

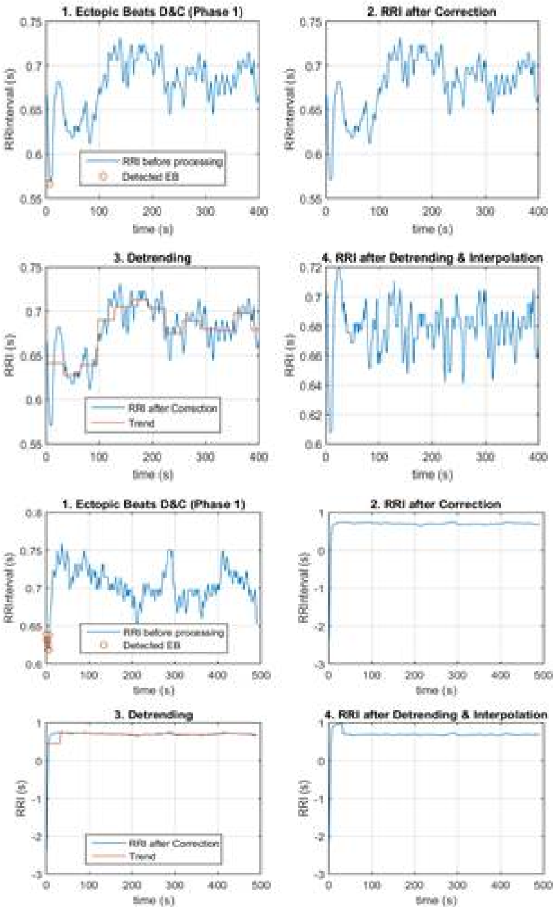

Fig. 4. Preprocessing stress (left) and non-stress (right) state results.

line indicates the trend of the RRI after ectopic beat correction. Decompose RRI after correction signal by using discrete wavelet transform (DWT) for Wavelet. Then set all lowest frequency trend (the highest level approximation of each decomposed sub - band) to zero. Finally, do inverse DWT or DWPT. The results are shown in 4.RRI after Detrending & Interpolation. There are 56 features extracted from each preprocessed heart beat data: 9 from time - domain, 8 from non-linear (6 from Detrended Fluctuation Analysis (DFA), and 2 from Poincare analysis), and 39 from frequency - domain (13 features using 3 different analysis method). The results of HRV Analysis are shown in Table 3 and Table 4.

Table 4. HRV analysis time domain and non-linear domain to subject Gilang in Non-Stress State

Time Domain		Non - Linear Domain	
Features	Value	Features	Value
Mean	689.000000	DFA. alpha (1)	0.522000
Median	705.900000	DFA. alpha (2)	1.565000
SDNN	187.300000	DFA. alpha1(1)	1.410000
SDANN	81.800000	DFA. alpha1(2)	0.688000
NNx	7.000000	DFA. alpha2(1)	0.337000
pNNx	1.400000	DFA. alpha2(2)	1.888000
RMSSD	63.700000	Poincare. SD1	44.900000
SDNNi	282.300000	Poincare. SD2	261.000000
meanHR	83.800000		

Table 4. HRV Analysis Frequency domain to subject Gilang in non-stress state

Features	Welch	AR	Lomb
aVLF	2145.460000	88.100000	0.019000
aLF	2054.880000	136.040000	0.025000
aHF	107.290000	8.800000	0.012000
aTotal	4307.630000	232.940000	0.056000
pVLF	49.800000	37.800000	33.700000
pLF	47.700000	58.400000	44.600000
pHF	2.500000	3.800000	21.700000
nLF	0.950000	0.939000	0.673000
nHF	0.050000	0.061000	0.327000
LFHF	19.152000	15.459000	2.057000
PeakVLF	0.000000	0.040000	0.020000
Peak LF	0.040000	0.050000	0.060000
PeakHF	0.150000	0.150000	0.180000

4.3 Feature Selection

Based on feature selection for all data gathered from every experiment until now, here is the correlation table gathered by using MATLAB function. The matrix resultsare 56 × 56 matrix with several values are NaN (not a number) because there are zero values for every experiment seen.

As the number above represent the linear correlation between one parameter with the other, with the number closer to zero means the two variables do not have any correlation, so the selection is performed by comparing all the data with one range. The features tuple that has 0 correlation values are: <Peak LF (Welch), LFHF (AR)>, < Peak HF (Welch), LFHF (AR) >, <Peak LF (Welch), SD2>, < Peak HF (Welch), SD2>, < Peak LF (Welch), SDNN>, <Peak HF (Welch), SDNN>. This concludes that every analysis domain will contribute in the model.

After selecting the tuple above, for each tuples features, next selection step is performed. In this case, $-0.1 < \times < 0.1$ and $-0.2 < \times < 0.2$ are used as the range parameter for selection. The first range creates 56 combinations of features ranging from 3 features to 6 features in one list. In contrary, the second range creates 152 combinations of features with maximum features is 7. Several results of the selected featuresare shown in Table 5 and Table 6.

Table 5. Examples of first range selected features

First Range (-0.1 < × < 0.1) Selected Features					
PeakLF (W)	LFHF (AR)	LFHF(L)	PeakLF (L)		
SDNNi	nLF(W)	nHF(W)	PeakLF (W)	LFHF(L)	PeakVLF (L)
SDNNi	nLF(W)	PeakLF (W)	LFHF(L)	PeakVLF (L)	PeakHF (L)
PeakLF (W)	LFHF (L)	PeakLF (L)			
SDNNi	nLF(W)	PeakLF (W)	LFHF(L)	PeakVLF (L)	

Table 6. Examples of second range selected features

First Range (-0.2 < × < 0.2) Selected Features					
Mean	SDNN	pLF (W)	PeakHF (W)	PeakHF (L)	
Mean	SDNN	pLF (W)	PeakHF (W)		
SDNN	PeakHF (W)	nLF(L)	nHF(L)	PeakVLF (L)	PeakLF (L)
SDNN	pLF(W)	PeakHF (W)			
aTotal (W)	PeakHF (W)	pHF(AR)	pVLF (L)	LFHF (L)	PeakHF (L)

4.4 Machine Learning

In order to check the result of feature correlation, the result of training both with and without parameter gathered from feature correlation result are compared. The implementation of SVM, decision tree, and discriminant analysis are using the MATLAB functions. For all the algorithm, the 10-fold cross validation is used and the error are also computed using average error of all the folds (implemented in MATLAB function). In order to gather the best parameter for every function, a hyper-parameter optimization that is newly introduced in MATLAB 2016b optimization option is used. The error after 10-fold cross validation will be computed using the best parameter from the hyper-parameter optimization iteration. The term minimum error is used as the average misclassification result of the 10-fold cross validation.

$$\text{Model Minimum Error} = \sum_{k=1}^{10} Misclassifiction\,Result$$

$$\text{Model Accuracy} = 100\% - \text{Model Minimum Error}$$

4.4.1 Support Vector Machine

The minimum error could be generated from the first model, using all data, is 10.56% by setting the box constraint to 5.8614 and kernel scale to 7.3937 (and there are another value pairs that leads to similar result). On the other hand, the feature selected model of first range ($-0.1 < \times < 0.1$) shows 14.29% error by specifying the box constraint parameter to 213.96 and kernel scale to 85.465 with the features used are Peak VLF (Lomb), LFHF (Lomb), Peak HF (Lomb), nHF (Welch), and SDNNi. And there are another 6 combinations of features that lead to similar error results as mentioned above. Better results appear in the second range selection with only 9.52% error. The features used in this model are: PeakHF (Lomb), LF/HF (Lomb), pVLF (AR), pHF (Welch), PeakHF (Welch). And there is one combination of features that lead to similar error results as mentioned above: LFHF (Lomb), pVLF (Lomb), pHF (AR), PeakHR (Welch). We summarized best-generated SVM model as depicted in Table 7.

Table 7. Summary of best generated SVM model

Feature Used	Parameter Leads to Minimum Error			
	Minimum Error	Kernel Function	Box Constraint	Kernel Scale
All	10.56%	RBF	5.8614	7.3937
1st Range Selection	14.29%	RBF	213.96	85.465
2nd Range Selection	9.52%	RBF	86.081	93.748

4.4.2 Decision Tree

The first model, which is decision tree with data from all of the recording, gave 10% error with setting the minimum leaf size parameter to 1. The first tree shows that the feature number 42 and number 8 are the most important features that can classify the stress and non-stress subject. Features number 42 is power Low Frequency (pLF) using AR algorithm, and feature number 8 is SDNNi.

In contrary, the selected features showed different results compared to all feature. The first range selected features show only 85.71% accuracy by also specifying the minimum leaf size into one. Features used in this model are Peak VLF (Lomb), LFHF (Lomb), Peak HF (Welch), nHF (Welch), and SDNNi. And there are another 6 features that have a similar result.

Better results appear in the second range selection with only 9.52% error. The features used in this model are Peak LF (Lomb), LF/HF (Lomb), LF/HF (AR), PeakHF (Welch), nLF (Welch), SDNNi. And there is one combination of features that lead to similar error results as mentioned above: Peak VLF (Lomb), LF/HF (Lomb), PeakHF (Welch), nLF (Welch), pHF (Welch) SDNNi. We summarized best - generated decision tree model as depicted in Table 8.

Table 8. Summary of best generated decision tree model

Feature Used	Parameter Leads to Minimum Error	
	Minimum Error	Minimum Leaf Size
All	10%	1
1st Range Selection	14.29%	1
2nd Range Selection	9.52%	1

4.4.3 Discriminant Analysis

When implementing this algorithm, there is a requirement for the data that doesn't contain any zero value. Both the all features and selected features data contains zero variance, therefore only two discriminant type that could be used to create the classifier model: diagonal linear and pseudo - linear method. The parameters that caused zero variance among the data are peakVLF, peakLF, and peakHF from Welch algorithm. Therefore, for the selection algorithm, all three features listed above is not used and we create another criterion which is $0.01 < \times < 0.1$ and $-0.1 < \times < -0.01$. Based on that criteria, there are 791 combinations of features obtained ranging from 3 to 6 features for every list. The discriminant analysis model in MATLAB uses two most important parameter for optimization, delta and gamma. Both delta and gamma are regularization parameters. Gamma(γ) can be obtained from this equation.

$$\hat{E} = (1-\gamma)E + \gamma(diag(\hat{X}^T * \hat{X})),$$

where, E represent covariance matrix of data X, $\hat{X}$ is the centered data (X minus mean by class), and $\hat{E}$ is regularized covariance matrix.

For the first model using all the recording data, the minimum error value observed is 4.56% by using pseudo-linear discriminant with parameter delta set is to 1.4231e-05 and parameter gamma is set to 0.6546. Meanwhile, the selected features show inaccurate results compared to the previous model with only 90.5% accuracy with using 0.00011201 delta value and 0.050069 gamma value. We summarized best - generated discriminant analysis model as depicted in Table 9.

Table 9. Summary of best generated discriminant analysis model

Feature Used	Parameter Leads to Minimum Error			
	Minimum Error	Discriminant Type	Delta	Gamma
All	4.56%	Pseudo - linear	1.4231e-05	0.6546
Feature Selected	9.52%	Pseudo - linear	0.00011201	0.050069

5. CONCLUSION

The design of this work has been implemented until the end of the module. There are 56 features generated by implementing HRVAS MATLAB code. Feature selection method gave p - values correlation matrix among those 56 features. With using an own algorithm for choosing the important value, there are several lists of features combination selected based on specified range. There are three range used: $-0.1 < \times < 0.1$; $-0.2 < \times < 0.2$; and $-0.1 < \times < -0.01$ and $0.01 < \times < 0.1$. The first and second range are used to create SVM and decision tree model and the third range is used for discriminant analysis because there is a restriction that features with 0 value has to be removed in order to perform the analysis. The first ranges create 56 lists of feature combination ranging from 3 to 6 features per list, the second range creates 152 lists of feature combination ranging from 3 to 6 features per list, and the third range creates 791 lists of feature combination ranging from 3 to 6 features per list.

Utilizing both the selected features and all features as the input data for the three different kinds of machine learning modeling techniques results to nearly similar value for SVM and decision tree. Both shows their minimum error value is 9.52% using the second range selected feature. This shows the feature selected perform better than using all data. In contrary, in discriminant analysis, using all data leads to better results by 5%. All recording data using pseudo -linear discriminant type results in 4.56% error only. There are two possible cause of the under performed model. The first one is related with the number of observations, and the other one is due to the assumption taken in the very first step thus affecting the label of the data (state: stress and non-stress). However, with 90% average best accuracy for all the implemented model and only 5 minutes recording, the system could possibly be implemented in real life as the prevention method of stress.

To overcome those mentioned problems, the very next step for the future work of this research will be gathering more data by looking for another subject and possibly a woman. Lastly, a further literature review about stress is neededto be done in order to create more general experiment procedure to assess the stress and non -stress condition of the subject. Thus, more stress level model could possibly be created instead of doing binary classification only.

REFERENCE

[1] M. Subramanya, J. Vivek, T. Vijay Narayan, and N.M. Mithun, "Continuous Monitoring of Stress on Smartphone Using Heart Rate Variability," *Proceeding of IEEE International Conference on Bioinformatics and Bioengineering*, pp. 1-5, 2015.

[2] S. Akane and P. Rosalind, "Stress Recognition Using Wearable Sensors and Mobile Phones," *Proceedings of Humane Association Conference on Affective Computing and Intelligent Interaction*, pp. 671-676, 2013.

[3] B. Sansanee and P. Sukanya, "Comparison of Heart Rate Variability Measures for Mental Stress Detection," *Journal of Computing in Cardiology*, Vol. 38, pp. 85-88, 2011.

[4] Listen to Your Heart: Stress Prediction Using Consumer Heart Rate Sensors, http://cs229.stanford.edu/proj2013/Liu Ulrich-ListenToYour Heart-Stress Prediction UsingConsumerHeart Rate Sensors.pdf (accessed Sep., 17, 2015).

[5] C. Eggert, O.D. Lara, and M.A. Labrador, "Recognizing Mental Stress in Chess Players Using Vital Sign Data," *Proceedings of IEEE Southeastcon*, pp. 1-4, 2013.

[6] A. Bert and T. Gerhard, "Towards Measuring Stress with Smartphones and Wearable Devices During Workday and Sleep," *Journal of Research Gate Bio Nano Science*, Vol. 3, No. 2, pp. 172-183, 2013.

[7] M. Wu, H. Cao, H.L. Nguyen, K. Surmacz, and C. Hargrove, "Modeling Perceived Stress Via HRV and Accelerometer Sensor Streams," *Proceeding of Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1625-1628, 2015.

[8] L. Vanitha and G.R. Suresh, "Hybrid SVM Classification Technique to Detect Mental Stress in Human Beings Using ECG signals," *Proceeding of International Conference on*

Advanced Computing and Communication Systems, pp. 1-6, 2013.

[9] J.T. Ramshur, *Design, Evaluation, and Application of Heart Rate Variability Analysis Software*, Master's Thesis of University of Memphis, United States, 2010.

[10] V. Timothy, A.S. Prihatmanto, and K.H. Rhee, "R-to-R Extraction and Preprocessing Procedure for an Automated Diagnosis of Various Disease from ECG Data," *Journal of Multimedia and Information System*, Vol. 3, No. 3, pp. 303-310, 2016.

[11] Stress and Cognitive Appraisal, https://explorable.com/stress-and-cognitive-appraisal (accessed Sep., 19, 2016).

[12] M. Paolo, B. Marcello, and P. Leandro, "Nonlinear Heart Rate Variability Features for Real - Life Stress Detection. Case Study: Students Under Stress Due to University Examination," *Journal of Biomedical Engineering Online*, Vol. 10, No. 1, pp. 96-109, 2011.

[13] V. Vesna and G. Vera, "Heart Rate Variability in Mental Stress Aloud," *Journal of Medical Engineering and Physics*, Vol. 29, No. 3, pp. 344-349, 2007.

[14] L. Salahuddin, J. Cho, M.G. Jeong, and D. Kim, "Ultra Short Term Analysis of Heart Rate Variability for Monitoring Mental Stress in Mobile Setting," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* pp. 4656-4659, 2007.

[15] M.V. Kamath, M. Watanabe, and A. Upton, *Heart Rate Variability (HRV) Signal Analysis: Clinical Applications*, CRC Press, Boca Raton, Florlda, 2013.

[16] A.H. Khandoker, C. Karmakar, M. Brennan, M. Palaniswami, and A. Voss, *Poincaré Plot-Methods for Heart Rate Variability Analysis*, Springer, New York, 2013.

[17] M. Schönfelder, G. Hinterseher, P. Peter, and P. Spitzenpfeil, "Scientific Comparison of Different Online Heart Rate Monitoring Systems," *International Journal of Telemedicine and Application*, Vol. 2011, No. 11, pp. 1-6, 2011.

[18] G. Liu, L. Wang, Q. Wang, G. Zhou, Y. Wang, and Q. Jiang, "A New Approach to Detect Congestive Heart Failure Using Short-Term Heart Variability Measures," *Journal of PLoS ONE*, Vol. 9, No. 4, pp. 1-8, 2014.

[19] T. Kaufmann, S. Sutterlin, S.M. Schulz, and C. Vogele, "ARTiiFACT: A Tool for Heart Rate Artifact Processing and Heart Rate Variability Analysis," *Journal of Behavior Research Method*s, Vol. 43, No. 4, pp. 1161-1170, 2011.

[20] B. Mali, S. Zulj, R. Magjarevic, D. Miklavcic, and T. Jarm, "Matlab-Based Tool for ECG and HRV Analysis," *Journal of Biomedical Signal Processing and Control*, Vol. 10, pp. 108-116, 2014.

[21] L. Rodríguez Liñares, A.J. Méndez, M.J. Lado, D.N. Olivieri, X.A. Vila, and I. Gómez Conde, "An Open Source Tool for Heart Rate Variability Spectral Analysis," *Journal of Computer Methods and Programs in Biomedicine*, Vol. 103, No. 1, pp. 39-50, 2011.

[22] M.P. Tarvainen, J.P. Niskanen, J.A. Lipponen, P.O. Ranta-Aho, and P.A. Karjalainen, "Kubios HRV-Heart Rate Variability Analysis Software," *Journal of Computer Methods and Programs in Biomedicine*, Vol. 113, No. 1, pp. 210-220, 2014.

[23] L.C.M. Vanderlei, R.A. Silva, C.M. Pastre, F.M. Azevedo, and M.F. Godoy, "Comparison of the Polar S810i Monitor and the ECG for the Analysis of Heart Rate Variability in the Time and Frequency Domains," *Brazilian Journal of Medical and Biological Research*, Vol. 41, No. 10, pp. 854-859, 2008.

[24] D. Nunan, G. Donovan, D.G. Jakovljevic, L.D. Hodges, G.R. Sandercock, and D.A. Brodie, "Validity and Reliability of Short-Term Heart-

Rate Variability from the Polar S810," *Journal of Medicine and Science in Sports and Exercise*, Vol. 41, No. 1, pp. 243-250, 2008.

[25] L.G. Porto and L.F. Junqueira, "Comparison of Time-Domain Short-Term Heart Interval Variability Analysis Using a Wrist-Worn Heart Rate Monitor and the Conventional Electrocardiogram," *Journal of Pacing and Clinical Electrophysiology*, Vol. 32, No. 1, pp. 43-51, 2009.

[26] D. Giles, N. Draper, and W. Neil, "Validity of the Polar V800 Heart Rate Monitor to Measure RR Intervals at Rest," *European Journal of Applied Physiology*, Vol. 116, No. 3, pp. 563-571, 2016.

[27] C.C.R. Sady and A.L.P. Ribeiro, "Symbolic Features and Classification Via Support Vector Machine for Predicting Death in Patients with Chagas Disease," *Journal of Computer in Biology and Medicine*, Vol. 70, pp. 220-227, 2016.

[28] MathWorks, *Machine Learning Challenges: Choosing the Best Model and Avoid Overfitting,* MATLAB White Paper, 7371_92974v00, 2016.

[29] B.K. Nayak and A. Hazra, "How to Choose the Right Statistical Test?," *Indian Journal of Ophthalmology* Vol. 59, No. 2, pp. 85-86, 2011.

[30] K. A. Kim and M. G. Park, "A Study on the Methods of Fault Analysis to Improve Safety in U-Healthcare System for Managing Emergency Rescue for Seniors," *Journal of Korea Multimedia Society*, Vol. 17, No. 2, pp.170-179, 2014.

SPL Aditya Pramanta

He graduated with Bachelor's Degree in Electrical Engineering at the Institute Technology Bandung (ITB), Indonesia. He is a graduate student in School of Electrical Engineering and Informatics at the Institute Technology Bandung, Indonesia, and also a research member of the Software Engineering and Multimedia Information System (SEMI) Laboratory as well as a Dual Degree Graduate Student of Graduate School, the Pukyong National University, Rep. of Korea. His research interests are in System Analysis and Design. e-Healthcare, Software Security and Safety.

Myong Hee Kim

She graduated with M.S. degree in Computer Science and Ph. D. in Information Systems from the Pukyong National University, Rep. of Korea. She was a Post Doctoral Researcher at the University of Colorado-Denver, USA. She is a lecturer of the Department of Information Systems, Graduate School, Pukyong National University. Also she was a lecturer of the Department of Computer Science and Engineering at the University of Colorado-Denver, USA. She served as an Assistant Faculty Consultant and an ICT Professional specialist for CPSC which is an Inter-Governmental International Organization for Human Resources Development in Asia and the Pacific Region.

Man-Gon Park

He is a head professor of the Dept. of IT Convergence and Application Engineering, Pukyong National University, Rep. of Korea since 1981. Also he was the president (2008) and chairman of the Korea Multimedia Society. He served as the Director General and CEO of the Colombo Plan Staff College for Technician Education (CPSC) from 2002 to 2007, which is an intergovernmental international organization of 29 member governments for Human Resources Development in Asia and the Pacific Region. His main areas of research are software reliability engineering, software safety and security engineering, BPR, Internet and web technology, multimedia information processing technology, and ICT-based HRD.