

Stress Classification Using K-means Clustering and Heart Rate Variability from Electrocardiogram

Mingu Kang, Siho Shin, Jaehyo Jung, Youn Tae Kim

*IT Fusion Technology Research Center, Department of IT Fusion Technology, Chosun University,
309 Pilmun-daero Dong-gu Gwangju 61452, Korea. petruskim@chosun.ac.kr*

Abstract— In this study, we propose a method to classify individuals under stress and those without stress using k-means clustering. After extracting the R and S peak values from the ECG signal, the heart rate variability is extracted using a fast Fourier transform. Then, a criterion for classifying the ECG signal for the stress state is set, and the stress state is classified through k-means clustering. In addition, the stress level is indicated using the $R - S_{peak}$ value. This method is expected to be applied to the U-healthcare field to help manage the mental health of people suffering from stress.

Keywords— *K-means Clustering, Electrocardiogram (ECG), Heart Rate Variability (HRV), Fast Fourier Transform (FFT)*

I. INTRODUCTION

Stress is known to be a significant threat to mental health. In reality, modern people experience high levels of stress owing to an uneasy future, excessive work, and various social risk factors. These stresses usually occur when the environment in which they are in is threatening, or when they are in a situation that is difficult to overcome.

Stress can be broadly classified into mental and physiological stress. Mental stress refers to worries regarding relationships, financial anxiety, and career paths. Physiological stress is classified as physical fatigue or a feeling of hunger.

Therefore, various studies are being conducted to solve this type of stress. As an example, by observing the change in the interval between the R peak values in the ECG signal, it is possible to check how much stress a person has received. However, the accuracy of stress classification using the change in the R peak interval is 45.39%, which is low [1].

Previously, the study of stress signal classification using k-means clustering mainly used electroencephalogram (EEG) signals [2]. However, the method for measuring an EEG is extremely complex, and the system is quite expensive. In addition, it is difficult to classify the stress signals because the criteria for classifying such signals are not clear. To compensate for these shortcomings, this study proposes a stress signal classification method using k-means clustering using ECG signals.

II. METHOD

In this study, an electrocardiogram signal is required to check the state of stress. Feature points according to the difference between the R-R interval and the $R - S_{peak}$ value are extracted from the acquired ECG signal.

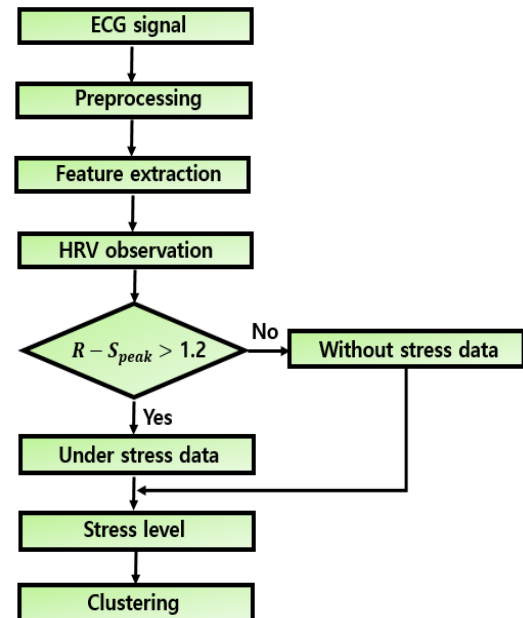


Fig. 1. Stress classification algorithms

To classify an ECG signal, the R-S peak value is set to 1.2 as the reference point. If it is greater than this value, the ECG signal moves to the right, and if it is smaller, it moves to the left. By using these feature points, how much stress the user is experiencing is indexed. Finally, K-means clustering is applied and classified into cases of stress and no stress.

A. Data acquisition

In this study, data according to the stress conditions are classified using the MIT-BIH Noise Stress Test Database [3]. In the database, ECG data from a total of 30 people are stored.

B. Preprocessing

We used the median filter and a low-pass filter to improve the performance of the stress classification algorithm [4]. The median filter is a nonlinear filter that removes high-frequency noise from an ECG signal. After applying a median filter, a low-pass filter was used to remove the noise [5].

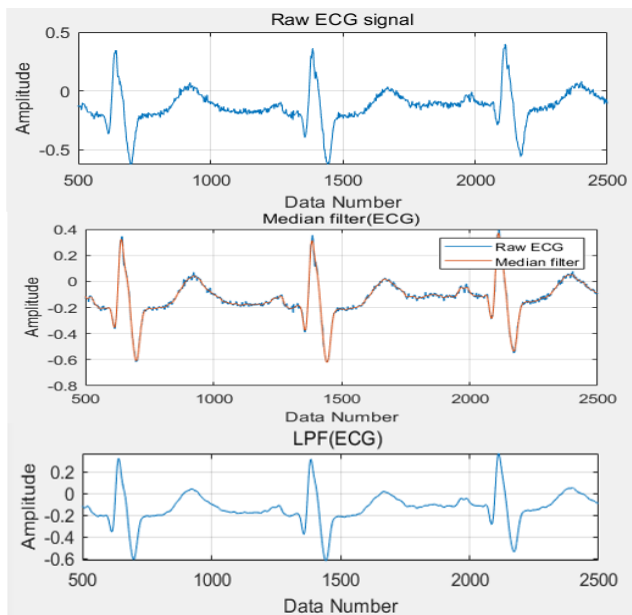


Fig. 2. Filter design for noise reduction

C. Feature extraction

Figure 3 shows the $R - S_{\text{peak}}$ according to each state of stress. To extract the feature points, a threshold was set from the ECG signal, and the peak values of R and S were extracted [6].

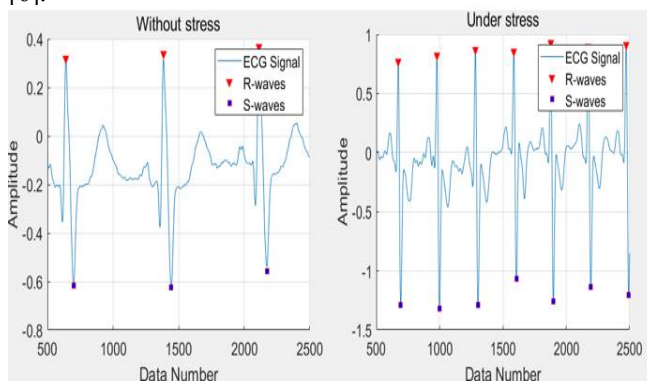


Fig. 3. Feature extraction from ECG signal

Table 1 shows the average of the $R - S_{\text{peak}}$ values in the ECG data acquired under stress and without stress. When this was confirmed, under stress, the heart rate increased and the interval of the ECG signal narrowed [7]. By contrast, without stress, the heart rate slowed.

TABLE I. FEATURE VALUES EXTRACTED ACCORDING TO STRESS

	Without stress	Under stress
R peak (mV)	0.38	0.83
S peak (mV)	-0.57	-1.33
$R - S_{\text{peak}}$ (mV)	0.95	2.16

D. HRV observation

Figure 4 shows the power spectrum confirming the frequency domain under stress and without stress. It can be

confirmed that the 0–0.15 Hz band is activated when under stress. The heart rate variability is a tool that can observe changes in the heart rate [8]. In this study, the heart rate variability under stress and without stress was observed using a fast Fourier transform (FFT).

When observing the activity of the heart, the frequency bands were set to a very low frequency (VLF) of 0.003–0.04 Hz, low frequency (LF) of 0.04–0.15 Hz, and high-frequency (HF) of 0.15–0.5 Hz [9]. By comparing the frequency signals when under stress and without stress, the effect of stress on the heart can be confirmed.

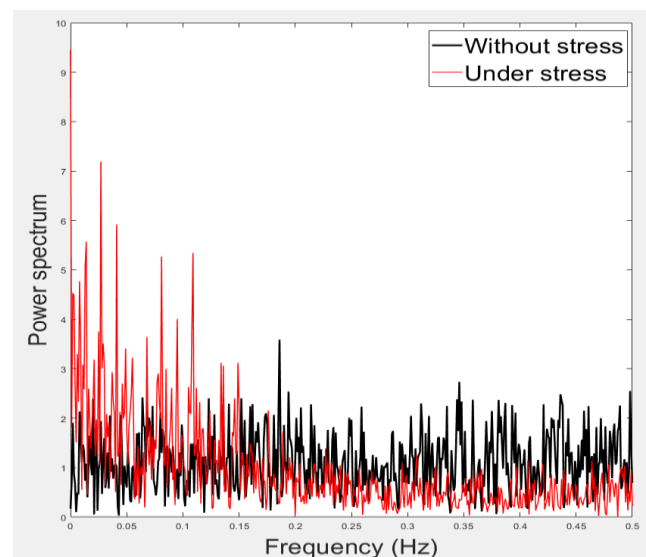


Fig. 4. Power spectrum of ECG signal

When a person is under stress, the sympathetic nerve corresponding to an LF is relatively activated, whereas under a state of non-stress, the parasympathetic nerve corresponding to an HF is activated [10].

Therefore, Fig. 4. and Table II show that an LF is activated. By contrast, it was confirmed that an HF was activated without stress.

TABLE II. MEAN FREQUENCY WITHOUT STRESS AND UNDER STRESS

FFT (Frequency domain)	LF (ms ²)	HF (ms ²)
Under stress	1.312	0.046
Without stress	0.164	1.156

E. Classification of stress signal

First, two-dimensional graphs of the R and S peaks of ECG data are shown for clustering according to the data classification. Then, to classify the ECG data, based on the $R - S_{\text{peak}}$ value, if the value of the new data is greater than 1.2, it is moved to the right. Conversely, if this value is less than 1.2, it is shifted to the left [11].

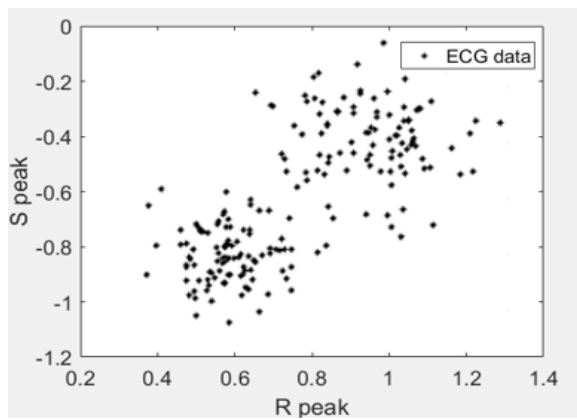


Fig. 5. Plot with distributed R and S peaks

Fig. 5 shows the R and S peaks extracted from the ECG data and displayed as a graph. Subsequently, the ECG data were placed to the left or right using the calculation of (1), which is shown in Fig. 6.

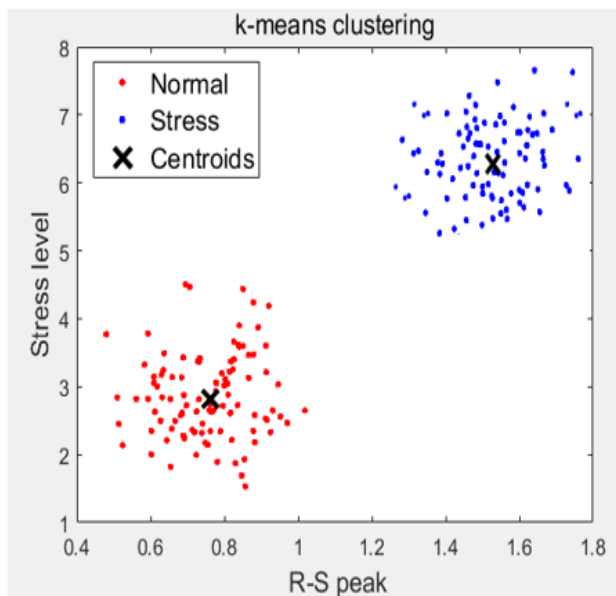


Fig. 6. Result of k-means clustering

K-means clustering is divided into an initialization, an assignment, and an update. An initialization is the step used to determine the number of clusters [12].

In (1), r_{ik} represents a case in which the i -th input data corresponds to the k -th cluster [13]. If the $R - S_{\text{peak}}$ value of the ECG data value is greater than 1.2, it corresponds to a cluster under stress. Conversely, if the $R - S_{\text{peak}}$ value is less than 1.2, it corresponds to a cluster without stress.

$$r_{ik} = \begin{cases} \text{Under stress} & k = x_i > 1.2 \\ \text{Without stress} & k = x_i < 1.2 \end{cases} \quad (1)$$

An assignment is the step of calculating the center point of the cluster [14]. Here, y_i is the ECG data before moving, r_{ik} is the ECG data after moving, n is the number of data, and c_k is the center point [15].

$$c_k = \sum_{i=1}^N \left(\frac{y_i - r_{ik}}{n} \right) \quad (2)$$

Therefore, it is possible to display the centroid of each cluster using (2).

An update is the process of reviewing whether all ECG data belong to the nearest cluster [16]. If any data are misclassified, the cluster closest to the data is found and the data are added into that cluster. Through this process, signals can be classified as under stress and without stress.

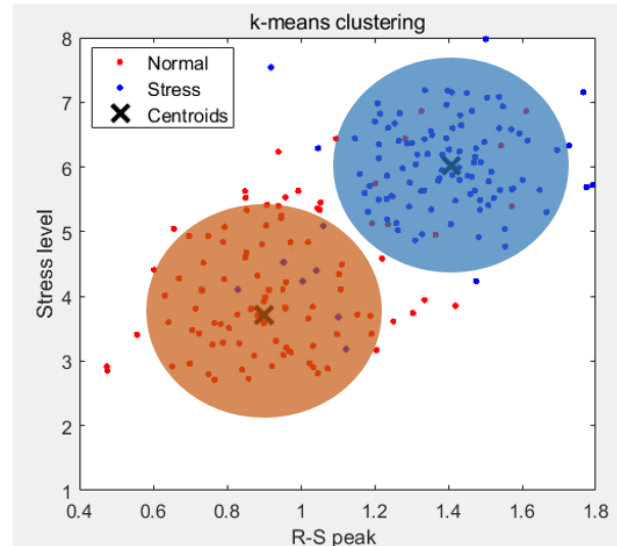


Fig. 7. Classification of k-means clustering

The stress intensity can be expressed according to the $R - S_{\text{peak}}$ value. Table 3 shows the level according to each size, and it can be seen that the higher the level is, the more severe the state of stress [17].

TABLE III. STRESS INDEX ACCORDING TO $R - S_{\text{PEAK}}$ VALUE

ECG signal	$R - S_{\text{peak}}$ (mv)	Stress level
Without stress	≥ 0.59	1
	0.6–0.79	2
	0.8–0.99	3
	1–1.19	4
Under stress	1.2–1.39	5
	1.4–1.59	6
	1.6–1.79	7
	≤ 1.8	8

The accuracy of the stress classification algorithm proposed in this study was calculated by comparing all ECG data and the distances to all centroids. This accuracy was 85.07%. Compared to previous studies on classifying stress signals, it was confirmed that the proposed algorithm achieves a better performance.

III. CONCLUSION

In this study, we developed an algorithm to classify the stress of an individual using k-means clustering from ECG signals. In addition, it was possible to check whether the user was exposed to stress by indexing the stress according to the $R - S_{\text{peak}}$ value. The accuracy of the developed algorithm was 85.07%. The proposed algorithm is expected to be helpful in mental health management by providing a value to the level of stress, which is a problem faced by modern society. In addition, it is expected to contribute to various studies on stress by incorporating medical ECG measuring devices and the U-healthcare field.

Acknowledgments

This research was supported by the Mid-career Researcher Program (No. 2016R1A2B3009423) through NRF grant funded by the MSIT (Ministry of Science and ICT) and the Basic Science Research Program through the NRF funded by the Ministry of Education of Korea (No. 2018R1A6A1A03015496).

REFERENCES

- [1] HAMILTON-CRAIG, Christian R, "Accuracy of quantitative echocardiographic measures of right ventricular function as compared to cardiovascular magnetic resonance," *IJC Heart & Vasculature*, pp. 38-44, Dec 2016.
- [2] Azhari, Ahmad, and Leonel Hernandez, "Brainwaves feature classification by applying K-Means clustering using single-sensor EEG." *International Journal of Advances in Intelligent Informatics* 2.3, pp.167-173, Nov 2016.
- [3] Mehra, Anu, "Study of ECG signals based on gender and heart abnormalities." *International Journal of Computational Complexity and Intelligent Algorithms*, pp. 277-291, Feb 2020.
- [4] B. Pagkalos, Ilias, et al, "A High - Performance Application Specific Integrated Circuit for Electrical and Neurochemical Traumatic Brain Injury Monitoring," pp. 12-15, Mar 2018.
- [5] Barrios-Muriel, Jorge, "A simple SSA-based de-noising technique to remove ECG interference in EMG signals." *Biomedical Signal Processing and Control*, pp. 117-126, Sep 2016.
- [6] Elgendi M, Mohamed A, Ward R, "Efficient ECG compression and QRS detection for E-health applications," *Sci Rep*, pp.459-471, Mar. 2017.
- [7] Venkatesan, C, P. Karthigaikumar, and S. Satheeskumaran, "Mobile cloud computing for ECG telemonitoring and real-time coronary heart disease risk detection," *Biomedical Signal Processing and Control* pp.138-145, Jul 2018.
- [8] Kim, Hye-Geum, "Stress and heart rate variability: a meta-analysis and review of the literature," *Psychiatry investigation*, pp. 235-245, Mar 2018.
- [9] Behar, Joachim A, "A universal scaling relation for defining power spectral bands in mammalian heart rate variability analysis," *Frontiers in physiology*, pp. 1001-1005, Aug 2018.
- [10] Sbröllini, Agnese, et al, "Evaluation of the low-frequency components in surface electromyography," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp.3622-3626, Aug 2016.
- [11] Krasteva, Vessela, Irena Jekova, and Ramun Schmid, "Simulating arbitrary electrode reversals in standard 12-lead ECG," *Sensors*, pp. 20-29, Jul 2019.
- [12] Balouchestani, Mohammadreza, and Sridhar Krishnan, "Advanced K-means clustering algorithm for large ECG data sets based on a collaboration of compressed sensing theory and K-SVD approach," *Signal, Image and Video Processing*, pp. 113-120, Oct 2016.
- [13] Roopa, C. K., B. S. Harish, and SV Aruna Kumar, "Classification of ECG Arrhythmia using symbolic dynamics through fuzzy clustering neural network," *Third International Workshop on Pattern Recognition International Society for Optics and Photonics*, pp.108-114, Jul 2018.
- [14] Yaohui, Liu, Ma Zhengming, and Yu Fang, "Adaptive density peak clustering based on K-nearest neighbors with aggregating strategy," *Knowledge-Based Systems*, pp. 208-220, Oct 2017.
- [15] Altayeva, Aigerim, Suleimenov Zharas, and Young Im Cho, "Medical decision making diagnosis system integrating k-means and Naïve Bayes algorithms." *International Conference on Control, Automation and Systems (ICCAS)*, pp. 16-19, Oct 2016.
- [16] Tanweer, Safdar, and Naseem Rao, "Novel Algorithm of CPU-GPU hybrid system for health care data classification," *Journal of Drug Delivery and Therapeutics*, pp. 355-377, Jan 2019.
- [17] Remy, J., P. Velmani, and T. C. Rajakumar, "Heart beat peak detection using signal filtering in ECG data," *International Journal of Advanced Technology and Engineering Exploration*, pp. 12-24, Jan 2019.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en_US