Stress Detection With Single PPG Sensor by Orchestrating Multiple Denoising and Peak-Detecting Methods

SEONGSIL HEO[1], SUNYOUNG KWON[2], AND JAEKOO LEE [1], (Member, IEEE)

[1]College of Computer Science, Kookmin University, Seoul 02707, South Korea
[2]School of Biomedical Convergence Engineering, Pusan National University, Yangsan 50612, South Korea

Corresponding author: Jaekoo Lee (jaekoo@kookmin.ac.kr)

ABSTRACT Stress is one of the major causes of diseases in modern society. Therefore, measuring and managing the degree of stress is crucial to maintain a healthy life. The goal of this paper is to improve stress-detection performance using precise signal processing based on photoplethysmogram (PPG) data. PPG signals can be collected through wearable devices, but are affected by many internal and external noises. To solve this problem, we propose a two-step denoising method, to filter the noise in terms of frequency and remove the remaining noise in terms of time. We also propose an ensemble-based multiple peak-detecting method to extract accurate features through refined signals. We used a typical public dataset, namely, wearable stress and affect detection dataset (WESAD) and measured the performance of the proposed PPG denoising and peak-detecting methods by lightweight multiple classifiers. By measuring the stress-detection performance using the proposed method, we demonstrate an improved result compared with the existing methods: accuracy is 96.50 and the F1 score is 93.36%. Our code is available at https://github.com/seongsilheo/stress_classification_with_PPG.

INDEX TERMS Health, health care, time series analysis, signal processing, affective computing, feature extraction or construction, machine learning, mental health, feature engineering, PPG, ensemble method, denoising method, peck detection.

I. INTRODUCTION

When faced with environmental changes, we tend to keep our internal state constant. This mechanism is called homeostasis. Stress can be defined as a state of threatened homeostasis [1], [2]. Stress is one of the common problems in modern society. Long-term stress can lead to chronic activation of the stress response. Chronic stress even threatens physical and mental health. For example, it breaks down a body's immune system and causes cardiovascular disease, diabetes, depression, and other illness [3], [4]. Thus, it is important for us to detect and manage stress to improve the quality of life and reduce the threat of physical and mental health.

Accurately measuring stress has become an important task for people. In the past, stress was assessed by

The associate editor coordinating the review of this manuscript and approving it for publication was Jason Gu.
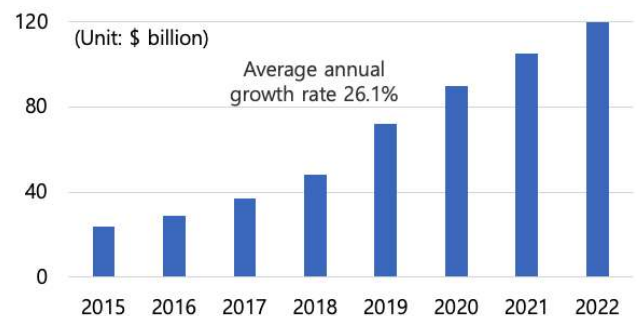
FIGURE 1. Global healthcare wearable device market revenue and growth.

directly answering questionnaires [5]. Recently, the demand for wearable devices that monitors our health condition in real-time has increased, as shown in Figure 1, and detecting stress by measuring physiological signals via
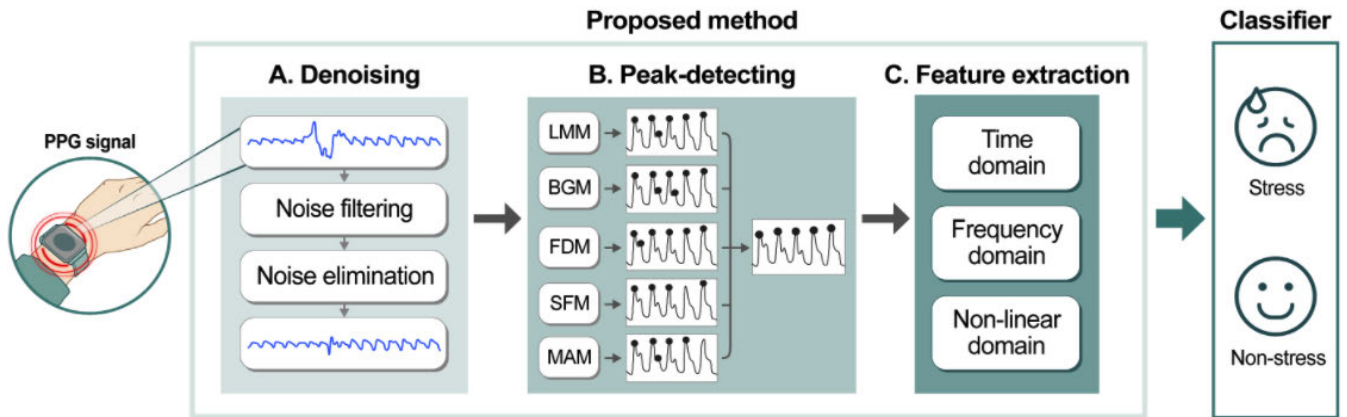
FIGURE 2. The overall process of the proposed method.

wearable devices has become possible. The typically used physiological signals are electrocardiogram (ECG) and photoplethysmogram (PPG).

ECG is an electrical signal in fine muscles around the heart activated during a cardiac cycle, and the data are collected by attaching sensors to the chest near the heart [6].

PPG is a signal that represents a changing arterial wave during each cardiac cycle, and the data are collected by attaching sensors to the wrist [7]. Therefore, setting up devices is inexpensive and convenient. The PPG signal contains more noise than the ECG signal. Most previous studies used relatively less noisy ECG signals or other types of signals [8], [9]. However, our current research focused on using only a single PPG signal, which is simpler and more convenient in everyday life.

Previous PPG studies focused mainly on denoising or peak-detecting method. Denoising techniques are typically performed in terms of frequency [10]–[12] and time [13], [14]. However, the current methods have limitations in noise reduction since the denoising techniques are applied independently.
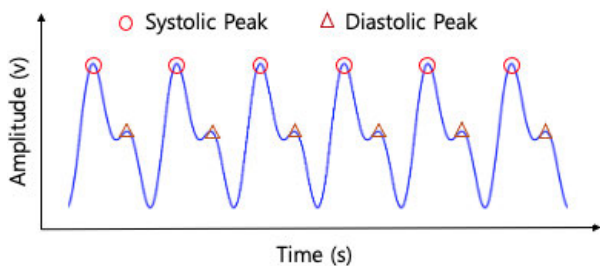

FIGURE 3. Typical PPG signal form.

The current peak-detecting methods mainly focus on attenuating the non-systolic peak, especially the diastolic peak or the peak triggered by noise. Figure 3 shows a typical PPG signal form. The circle indicates the systolic peak, and the triangle indicates the diastolic peak. The proposed method detects the peaks by extracting only a peak with the largest amplitude

value per block [15], [16], attenuating the non-systolic peak signals through signal transformation [16]–[19], and removing the non-systolic peak through thresholding [17]–[20]. However, each method can still detect incorrect peaks since the PPG signal comes in diverse forms, and the distribution depends on an individual characteristics.

The objective of this paper is to improve the performance of stress detection by extracting accurate features using orchestrating multiple denoising and peak-detecting methods (OMDP) with PPG signals to overcome the limitations of the previous PPG signal analysis. The contributions in this paper are as follows.

(1) We present an effective two-step denoising method in terms of frequency and time.

(2) We extract more accurate peak points by applying an ensemble-based multiple peak-detection method.

(3) We demonstrate the superiority of our proposed method through seven lightweight classifiers for stress detection.

II. METHODOLOGY

Figure 2 shows the entire process of this work. It consists of denoising, peak detection, and feature extraction. It finally measures the performance of stress detection by training seven lightweight machine-learning classifiers available on low power wearable devices.

A. ADAPTIVE TWO-STEP DENOISING METHOD

Figure 4 shows that the PPG signal-denoising process consists of two steps: noise filtering and noise elimination.

In the noise filtering process, we use a band-pass filter to compensate the signals in terms of the frequency [10], as shown in Figure 4(a). After applying the band-pass filtering, we use a three-point moving average filter for smoothing.

To deal with the remaining noise after the noise filtering process, we remove the noisy segment in terms of time using a statistical method [14]. Figure 4(b) shows the noise elimination process.
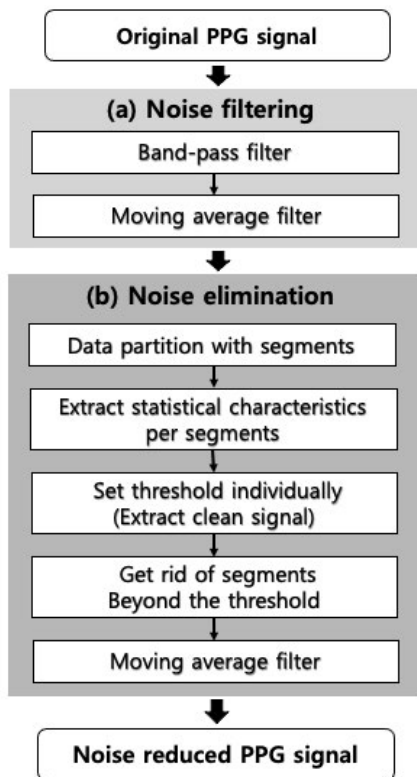
FIGURE 4. Adaptive two-step denoising process.

We first extract the valley from the signals and divide it into segments per cycle. The previously proposed valley-detecting method, an improved moving-window method, resulted in an inaccurate valley-detection problem because all signals continuously detect incorrect valley due to noise. To address this problem, we devise the method that calculates the mean of the entire signal amplitude and sets each segment from the previous to the current points, which contain the mean values of the entire signal amplitude, as suggested in [15]. We calculate the standard deviation, kurtosis, and skewness of all segmented data. The measurement method is expressed as follows:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (1)$$

$$kurtosis = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{std^4} \qquad (2)$$

$$skewness = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{std^3} \qquad (3)$$

where $\bar{x}$ is the mean value and σ is the standard deviation. When one of the three statistical parameters is beyond the threshold, the corresponding segment is eliminated. We extract the part with high-quality signals as a reference signal, and set it as the threshold. In this approach, we extract the reference signals by ourselves. We can obtain better performance when experts directly extract high-quality signals.

The threshold T_σ, T_k, and T_{s1} formulas are as follows:

$$T_\sigma = \bar{\sigma} + \alpha, \quad T_k = \bar{k} + \beta, \qquad (4)$$

$$T_{s1} = \bar{s} - \gamma, \quad T_{s2} = \bar{s} + \delta \qquad (5)$$

where $\bar{\sigma}$, $\bar{k}$, and $\bar{s}$ represent the mean of the three-reference signal, i.e., standard deviation, kurtosis, and skewness, respectively. The optimal parameters obtained from our experiment are as follows: $\alpha = 1.0$, $\beta = 2.0$, $\gamma = 1.8$, and $\delta = 1.5$. We reconstruct the signal by removing all the segments beyond the threshold. Finally, we utilize a three-point moving average filter for smoothing at the reconstructed signal.

B. ENSEMBLE-BASED PEAK-DETECTING METHOD
We apply the ensemble to five well-known peak-detecting methods and determine the final peak point by majority voting. Figure 5 shows each process of the five peak-detecting methods.
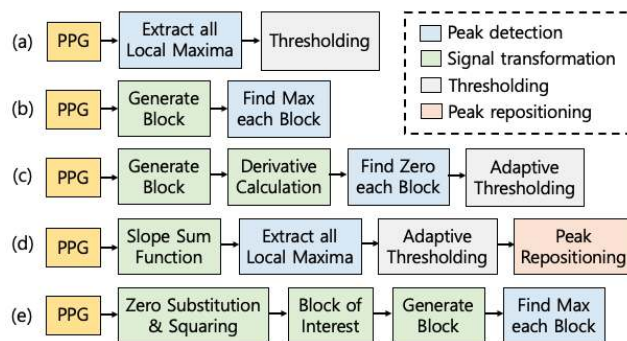

FIGURE 5. Procedures for five peak-detection methods. (a) LMM. (b) BGM. (c) FDA. (d) SFA. (e) MAD.

The *local maxima method (LMM)* [20] extracts all local maximum points from the signal. Among the extracted points, we remove the points with a lesser value than the mean of the entire signal amplitude.

The *block generation with the mean of the signal threshold method (BGM)* [15] generates blocks with the mean value of amplitude of the entire signal. The points that contain the largest value in each block become the peak point.

The *first derivative with an adaptive threshold method (FDA)* [17] divides the signal into every 5s and generates blocks. All signals are differentiated, and the points with a differential value of zero are extracted as peak candidates. The threshold is adaptively set according to the mean of amplitude in the 2s of the block. The peak candidates with larger amplitudes than the threshold become the peak point.

The *slope sum function with an adaptive threshold method (SFA)* [18], [19] applies the SSF, leaving only the ascending point through the differential process and setting the remaining points to zero. All the local maximum points are extracted from the modified signal. The threshold is constantly updated to 70% of the median value between the last five peaks, and only the point with amplitude is larger than

TABLE 1. Notation and description of used HRV features.

Domain	Parameter	Description	Unit
Time	MeanHR	Mean HR	$1/min$
	StdHR	Standard deviation of the HRs	$1/min$
	MeanNN	Mean of N-N intervals	ms
	SDNN	Standard deviation of N-N intervals	ms
	MedNN	Median of N-N intervals	ms
	MeanSD	Mean of successive N-N interval differences	ms
	SDSD	Standard deviation of successive N-N interval differences	ms
	RMSSD	Root mean square of successive N-N interval differences	ms
	pNN20	Percentage of successive N-N intervals that differ by more than 20ms	$\%$
	pNN50	Percentage of successive N-N intervals that differ by more than 50ms	$\%$
	TINN	Baseline width of the R-R interval histogram	Ms
Frequency	LF	Power in the low-frequency band(0.04 - 0.15Hz)	ms^2
	HF	Power in the high-frequency band(0.15 - 0.4Hz)	ms^2
	LF/HF	Ratio of LF to HF	-
	ULF	Power in the ultra-low-frequency band(≤ 0.003Hz)	ms^2
	VLF	Power in the very-low-frequency band(0.003 - 0.04Hz)	ms^2
	P(TotPow)	Total power 0.003-0.4Hz	ms^2
	LF / P	Ratio of low-frequency to total power	-
	HF / P	Ratio of high-frequency to total power	-
Non-linear	SD1	Poincaré plot standard deviation perpendicular to the line of identity	Ms
	SD2	Poincaré plot standard deviation along the line of identity	Ms
	pQ	pQ = SD1 / SD2	-
	pA	pA = SD1 * SD2	ms^2
	ApEn	Approximate entropy	-
	ShanEn	Shannon entropy	-
	D2	Correlation dimension, which estimates the minimum number of variables	-

the threshold is extracted. It is re-positioned into the original signal since the extracted peak point represents the peak point from the modified signal.

The ***moving averages with the dynamic threshold method (MAD)*** [16] set signals below zero, then square all signals. The block is generated by two moving averages. The points with the largest amplitude value in each extracted block ultimately become the peak point.

C. FEATURE EXTRACTION
The PPG characteristics for general stress detection are based on heart rate variability (HRV) [21]. HRV is quantified as the changes in the interval between successive peaks [22]. The signals must be divided into appropriate window sizes to calculate HRV. Since at least 2 min of the window length is required for accurate feature extraction, 2 min of the window length and 0.25 s of the sliding window are applied for HRV extraction. An inaccurate interval between successive peaks exists because not all the peak points of the signal can be perfectly extracted. Therefore, intervals of approximately 300 ms greater or lesser than the mean of the entire intervals are eliminated, and the rest are used to extract HRV [15].

The most widely used measurements, such as HRV time domain, frequency domain and nonlinear domain, are applied to extract the HRV features. We employ the most commonly used features, and these features are listed in Table 1 [21].

III. EXPERIMENTAL RESULT
A. EXPERIMENTAL SETUP
We employed wearable stress and affect detection (WESAD) public dataset to verify the proposed method. It provided

various types of physiological signals and was labeled with four emotional states: baseline, stress, amusement, and meditation [23]. The baseline condition aimed at inducing a neutral affective state. However, only a single PPG signal was used in our experiments, and the sampling rate was 64 Hz.

To set up the cut-off and threshold in the noise-elimination process of the adaptive two-step denoising method, a 0.1% high-quality signal with high-peak-detection performance was extracted and used as a reference signal. 952 windows were generated through feature extraction.

Figure 6 shows the changes of the PPG signal by applying the denoising method step-by-step: (a) the original, (b) after the noise-filtering step, and (c) after the proposed adaptive two-step denoising method. When the proposed denoising method is applied, the number of generated windows was reduced by approximately 8 % as shown in 6 (c).

To quantify the stress detection performance using the proposed method (OMDP), we utilized seven typical learning-based classifiers [23], [24]. The area under the receiver operating characteristic curve (AUC) [25] and F1 score were used to measure the proposed method performance. The accuracy was used for the performance comparison between the existing method and the proposed method. In addition, Leave-One-Subject-Out cross validation was used to verify the generalization performance [26], [27]. Our code is available at https://github.com/seongsilheo/stress_classification_with_PPG.

B. INTEGRATED APPROACH PERFORMANCE
We investigated the effectiveness of the proposed integrated method (OMDP) by comparing it with the non-integrated
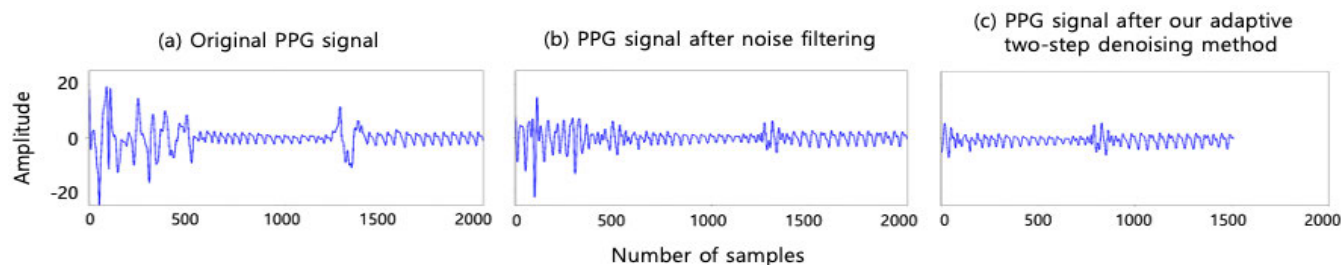
FIGURE 6. Visualization of a PPG signal when applying our denoising method step-by-step.

TABLE 2. Performance comparison between the proposed method and existing method (Abbreviations: 9-NN = 9-nearest neighbour, LDA = Linear discriminant analysis).

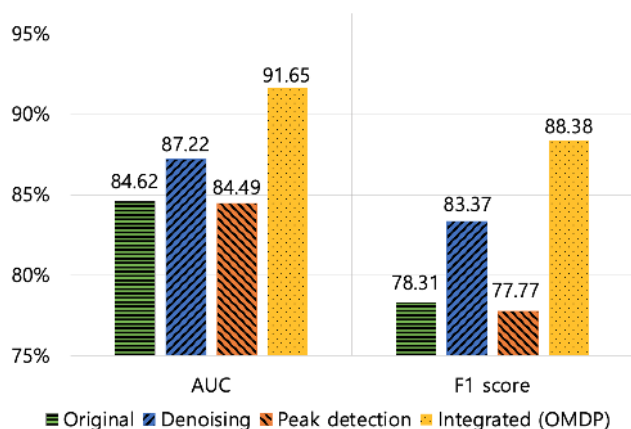|	Decision Tree	Random Forest	Ada-boosting	9-NN	LDA
	Accuracy (F1 score)				
Philip, et al [23]	81.39 (78.27)	84.18 (81.35)	84.10 (81.23)	82.06 (78.94)	85.83 (83.08)
OMDP (Our)	87.82 (81.09)	91.02 (87.02)	91.14 (87.43)	89.19 (89.19)	96.50 (93.36)

FIGURE 7. Performance results of ablation study of the proposed method.

methods: original, denoising without peak-detection, and peak-detection without denoising. OMDP is an integrated method that utilizes both two-step denoising and ensemble-based peak-detection. We performed a binary-classification, stress versus non-stress, and non-stress is defined by combining the state baseline and amusement [23]. Figure 7 shows the result of the ablation study of the proposed method by averaging all the result of the seven classifiers of AUC and F1 score.

When the ensemble-based peak-detecting method was not applied, the results of each proposed five peak-detection method were averaged. The integrated OMDP method exhibited the best performance with the result of 91.65 for AUC and 88.38 for the F1 score. The result indicated a performance improvement with 4.43% for AUC and 5.01% for the F1 score compared with the second best option that used only the adaptive two-step denoising method. Thus, we confirmed that the integrated method achieved better performance than any of the other methods. The peak detection without denoising showed the lowest performance because we applied the

peak-detection method to low-quality signals that contained a lot of noise.

Table 2 presents the performance comparison between the existing method proposed by *Schmidt, Philip, et al* under the same conditions [23] and our OMDP method. We measured the accuracy and F1 score using the same approach adopted in the previous method. The five classifiers demonstrated that the proposed OMDP outperforms the previous method by achieving a performance improvement at an average accuracy of 7.62% and F1 score of 7.04%.

The first table in Table 3 presents the results of each classifiers when the proposed OMDP method was applied. This table lists the AUC and F1 scores of each of the seven classifiers. The Linear discriminant analysis (LDA) classifier achieved the best result with 95.07 for AUC and 93.36 for the F1 score. The second table in Table 3 shows the overall performance results of the seven classifiers by applying the proposed method step-by-step. The bold marks represent the performance of the proposed OMDP method, and the gray marks represent the best results of each classifiers. The proposed OMDP method achieved the best results in all classifiers except the decision tree classifier. This result confirms the generalization performance that the proposed OMDP method performs well for any classifiers.

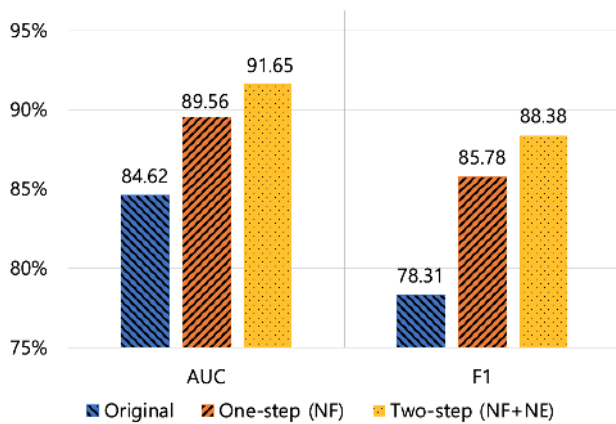## C. ADAPTIVE TWO-STEP DENOISING PERFORMANCE

To investigate the effect of the adaptive two-step denoising method, we measured the performance by applying our denoising method step-by-step: original, one-step noise filtering method, and two-step denoising method. The ensemble-based peak-detecting method was applied in every comparison. Figure 8 shows the adaptive two-step denoising method performance result by averaging all the result of the seven classifiers of AUC and F1 score.

We confirmed that the two-step denoising method achieved the best result with 91.65 for AUC and 88.38 for the F1 score.

TABLE 3. Summary of the performance of our OMDP method with seven classifiers, and the overall performance results with seven classifiers by applying our method step-by-step on the binary classification task. "Ens." on peak-detecting method section applied ensemble-based peak-detecting method. Bold marks represent the performance of our OMDP method, and the gray marks represent the best results for each classifiers. (Abbreviations: 9-NN = 9-nearest neighbour, LDA = Linear discriminant analysis, NF = Noise filtering, NE = Noise elimination).

OMDP (Our)	Decision Tree	Random Forest	Ada-boosting	9-NN	LDA	SVM	Gradient-boosting	Average of 7 classifiers
AUC	86.63	90.24	91.20	92.33	95.07	94.30	91.80	91.65
F1	81.09	87.02	87.43	89.19	93.36	91.95	88.61	88.38

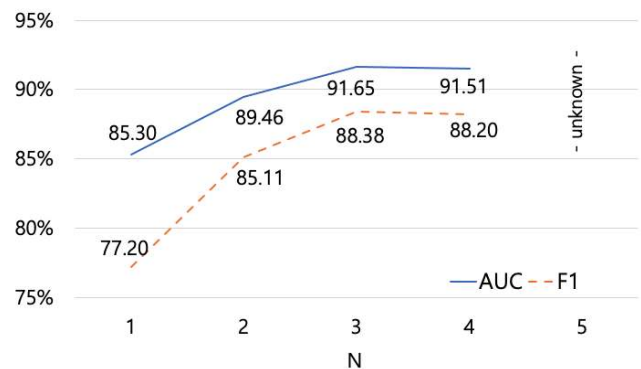| Denoising method	Peak-detecting method	Decision Tree [23]	Random Forest [23]	Ada-boosting [23]	9-NN [23]	LDA [23]	SVM [24]	Gradient-boosting [24]
					AUC (F1 Score)			
Base	LMM	80.47 (72.39)	80.57 (73.36)	80.77 (73.07)	83.56 (77.37)	82.82 (76.12)	83.74 (78.13)	79.26 (71.59)
NF		84.70 (77.77)	84.86 (78.92)	84.78 (79.72)	84.17 (76.59)	85.76 (78.07)	86.09 (80.11)	86.42 (81.55)
NF+NE		78.23 (66.76)	82.83 (74.48)	82.68 (73.42)	81.43 (70.18)	81.07 (69.38)	84.95 (77.15)	83.80 (76.07)
Base	BGM	79.62 (70.60)	82.57 (73.52)	85.36 (78.95)	84.58 (77.88)	88.02 (81.72)	88.81 (83.36)	84.32 (78.02)
NF		81.02 (73.40)	88.47 (84.54)	90.77 (86.74)	89.22 (84.39)	92.30 (89.26)	91.02 (87.28)	89.40 (85.58)
NF+NE		85.51 (79.86)	88.51 (83.42)	89.67 (85.80)	90.35 (84.84)	92.59 (89.70)	91.50 (87.30)	89.58 (85.84)
Base	FDA	80.20 (72.95)	81.19 (73.62)	86.45 (81.00)	84.68 (78.70)	88.83 (84.10)	88.19 (83.73)	84.35 (78.07)
NF		86.63 (81.51)	87.62 (83.28)	89.73 (85.73)	90.13 (86.42)	92.00 (88.71)	90.63 (87.22)	89.46 (85.66)
NF+NE		88.03 (81.78)	89.14 (84.55)	88.87 (83.29)	89.84 (84.97)	92.11 (87.99)	91.02 (86.70)	90.69 (86.19)
Base	SFA	84.24 (78.85)	83.18 (77.66)	85.55 (80.18)	86.24 (81.20)	89.92 (85.69)	88.97 (84.74)	87.33 (82.13)
NF		82.65 (76.71)	86.33 (82.34)	89.27 (84.97)	89.11 (84.90)	92.19 (88.66)	90.75 (87.08)	89.87 (86.29)
NF+NE		82.35 (74.68)	87.59 (81.89)	88.29 (82.72)	87.49 (81.56)	91.96 (87.74)	91.75 (87.79)	90.01 (85.00)
Base	MAD	82.39 (74.30)	84.81 (77.77)	83.48 (76.48)	84.94 (78.64)	86.49 (80.27)	88.94 (83.75)	86.93 (81.09)
NF		89.66 (85.00)	88.84 (83.99)	88.91 (84.96)	89.09 (84.55)	94.53 (92.48)	92.83 (89.73)	88.45 (83.68)
NF+NE		86.05 (80.02)	89.37 (85.05)	87.86 (83.41)	89.24 (84.77)	94.01 (91.28)	92.69 (89.85)	90.56 (86.19)
Base	Ens.	78.80 (70.53)	82.35 (74.51)	85.86 (79.01)	84.49 (78.34)	87.07 (80.72)	89.08 (84.30)	83.80 (76.98)
NF		87.29 (82.50)	88.97 (85.42)	89.05 (84.72)	90.62 (86.92)	92.21 (89.48)	90.94 (87.82)	87.83 (83.58)
NF+NE		86.63 (81.09)	90.24 (87.02)	91.20 (87.43)	92.33 (89.19)	95.07 (93.36)	94.30 (91.95)	91.80 (88.61)

FIGURE 8. Performance of an adaptive two-step denoising method (NF: noise filtering, NE: noise elimination).

FIGURE 9. Performance of an ensemble-base peak-detecting method according to N (N represents he number of at least how many peak points matched for the new peak point selection).

This result indicated an increase in the performance by 2.09% for AUC and 2.60% for the F1 score compared with the second best option that used only the one-step noise filtering method. An increase in performance of 6.44% for AUC and 10.61% for the F1 score compared with the worst option that did not use any of the denoising methods was also achieved.

D. ENSEMBLE-BASED PEAK-DETECTING PERFORMANCE

We set variable N to determine at least how many peak points should be matched for selection as the new peak point. We measured the performance of the ensemble-based peak-detecting method according to each N in the seven classifiers. The adaptive two-step denoising method was applied to every method. Figure 9 shows the result of the average of AUC and F1 scores with seven classifiers. We achieved the best result when N = 3, i.e., 91.65 for AUC and 88.38 for the F1 score.

When N = 4, the performance was comparable to that when N = 3. However, the number of extracted peaks was too small to extract features in some windows. Thus, we eliminated these windows, and the number of windows decreased

TABLE 4. Tri-classification task performance comparison between previous and our method.

	Decision Tree	Random Forest	Ada-boosting	9-NN	LDA
	Accuracy (F1 score)				
Philip, et al [23]	57.57 (51.15)	64.09 (53.83)	64.46 (53.29)	59.44 (50.97)	70.17 (54.72)
OMDP (Our)	67.12 (59.03)	76.66 (58.51)	71.89 (61.29)	66.67 (66.67)	78.68 (62.95)

TABLE 5. Overall performance results of the seven classifiers by applying our proposed method (OMDP) step-by-step on the multi-classification task (tri-class and quad-class). "Ens." on the peak-detecting method section when the ensemble with five peak-detection method is applied. The bold marks represent our OMDP method result, and the gray marks represent the best result for each classifiers. (Abbreviations: 9-NN = 9-nearest neighbour, LDA = Linear discriminant analysis, NF = Noise filtering, NE = Noise elimination).

# of classes	Denoising method	Peak-detecting method	Decision Tree [23]	Random Forest [23]	Ada-boosting [23]	9-NN [23]	LDA [23]	SVM [24]	Gradient-boosting [24]
			AUC (F1 Score)						
3									
	Base	LMM	65.53 (50.09)	71.16 (56.51)	65.35 (49.75)	70.35 (55.04)	70.80 (55.14)	71.37 (56.00)	70.96 (56.78)
	NF	LMM	67.56 (53.16)	71.20 (55.77)	68.88 (56.37)	70.41 (56.67)	71.47 (56.83)	72.16 (57.15)	73.85 (61.82)
	NF+NE		64.63 (49.43)	69.53 (53.12)	70.77 (57.88)	67.36 (52.46)	67.67 (53.56)	71.14 (55.75)	72.21 (59.11)
	Base	BGM	64.11 (48.94)	69.63 (51.83)	67.68 (53.53)	69.71 (54.73)	74.19 (61.15)	72.29 (56.24)	71.09 (56.70)
	NF		68.76 (54.67)	71.78 (55.87)	69.59 (56.13)	72.19 (58.58)	72.98 (59.16)	73.27 (58.22)	69.94 (55.89)
	NF+NE		71.39 (57.03)	72.11 (56.41)	73.60 (61.19)	72.14 (57.99)	72.57 (59.74)	73.91 (59.41)	72.30 (59.21)
	Base	FDA	68.39 (55.15)	71.04 (56.43)	70.86 (56.97)	70.83 (57.13)	75.14 (62.76)	75.09 (61.92)	70.68 (56.67)
	NF		71.09 (58.86)	72.98 (59.40)	70.03 (57.72)	74.55 (62.95)	76.79 (65.24)	74.97 (61.94)	74.18 (61.44)
	NF+NE		70.06 (57.91)	71.81 (57.50)	70.69 (58.61)	72.98 (61.61)	73.88 (61.54)	74.39 (61.23)	74.56 (62.19)
	Base	SFA	68.66 (56.19)	71.14 (56.85)	71.84 (58.34)	72.18 (59.68)	75.02 (61.34)	74.15 (59.75)	72.78 (59.35)
	NF		70.36 (57.83)	70.33 (54.20)	67.24 (54.16)	73.79 (61.19)	75.26 (62.27)	75.52 (62.19)	71.08 (57.20)
	NF+NE		70.91 (58.28)	70.69 (55.32)	68.10 (55.01)	72.86 (60.93)	75.01 (63.36)	75.38 (61.32)	72.83 (60.35)
	Base	MAD	64.83 (50.95)	69.04 (52.43)	63.15 (47.58)	70.08 (55.60)	71.58 (56.16)	72.76 (56.68)	70.57 (56.95)
	NF		71.31 (58.09)	71.76 (55.86)	71.48 (59.32)	74.13 (61.16)	77.68 (66.84)	75.57 (61.46)	74.11 (61.18)
	NF+NE		70.33 (58.49)	72.38 (57.17)	67.26 (54.04)	72.13 (59.77)	75.12 (63.03)	75.19 (61.71)	74.58 (62.44)
	Base	Ens.	67.67 (54.24)	70.20 (52.87)	69.49 (54.18)	71.53 (56.65)	71.68 (57.15)	73.90 (58.00)	70.50 (54.86)
	NF		68.40 (54.64)	72.18 (56.87)	70.49 (55.78)	75.27 (63.57)	74.76 (61.70)	75.23 (61.79)	72.78 (60.10)
	NF+NE		71.30 (59.03)	72.73 (58.51)	73.29 (61.29)	77.39 (66.67)	75.19 (62.95)	75.76 (62.49)	73.51 (61.14)
4									
	Base	LMM	62.54 (41.15)	67.50 (46.14)	63.67 (41.94)	65.52 (44.11)	66.97 (45.20)	67.86 (46.98)	66.75 (46.81)
	NF		61.81 (40.86)	62.77 (38.06)	62.95 (41.05)	63.37 (41.81)	62.99 (40.47)	67.52 (46.49)	65.52 (44.71)
	NF+NE		60.68 (38.71)	62.10 (38.28)	61.94 (40.25)	61.43 (38.93)	60.79 (37.56)	66.65 (45.31)	63.66 (42.21)
	Base	BGM	61.86 (40.83)	66.43 (43.45)	63.46 (42.91)	65.58 (45.41)	71.39 (53.18)	69.68 (49.59)	67.21 (46.29)
	NF		66.00 (46.83)	69.67 (49.14)	68.65 (50.51)	70.79 (53.19)	73.11 (55.64)	72.80 (54.37)	69.54 (49.88)
	NF+NE		67.23 (48.56)	71.43 (52.12)	67.14 (47.94)	70.98 (53.01)	73.55 (56.29)	73.14 (54.18)	71.63 (53.05)
	Base	FDA	66.23 (46.97)	68.03 (47.07)	64.83 (43.58)	70.30 (53.42)	70.72 (52.49)	70.99 (51.70)	70.09 (51.84)
	NF		68.54 (50.42)	69.79 (49.54)	66.99 (48.29)	70.40 (52.96)	71.59 (54.99)	69.80 (50.96)	71.52 (53.21)
	NF+NE		66.05 (45.68)	68.58 (47.91)	65.66 (45.75)	70.75 (53.11)	70.42 (52.91)	70.01 (50.85)	70.72 (51.42)
	Base	SFA	66.98 (47.94)	66.83 (45.17)	67.01 (48.72)	68.84 (51.59)	72.71 (55.22)	70.74 (51.56)	69.62 (51.03)
	NF		66.62 (47.82)	69.36 (49.01)	68.20 (49.59)	70.46 (53.37)	74.15 (57.62)	72.18 (53.68)	69.61 (50.80)
	NF+NE		68.25 (50.20)	69.97 (49.39)	70.24 (52.74)	71.93 (55.66)	74.79 (58.25)	73.61 (55.66)	69.93 (51.20)
	Base	MAD	65.94 (46.94)	66.66 (44.87)	64.29 (44.40)	67.26 (48.02)	69.92 (50.97)	69.47 (49.17)	67.75 (48.74)
	NF		67.23 (49.33)	70.30 (50.88)	68.94 (51.13)	71.03 (53.02)	75.11 (58.67)	73.71 (56.10)	71.37 (53.23)
	NF+NE		67.26 (49.20)	70.83 (51.93)	68.07 (49.00)	70.04 (53.05)	74.95 (58.60)	73.74 (56.66)	71.18 (53.36)
	Base	Ens.	62.87 (42.16)	66.68 (43.90)	63.57 (42.11)	68.06 (48.52)	68.92 (50.11)	70.77 (50.84)	67.50 (46.52)
	NF		67.96 (49.50)	70.53 (50.60)	68.92 (49.92)	71.93 (54.87)	74.21 (56.98)	73.38 (55.00)	70.28 (51.38)
	NF+NE		68.71 (50.81)	71.17 (51.71)	70.61 (52.06)	72.95 (56.93)	75.16 (58.67)	72.45 (54.00)	70.83 (52.60)

by approximately 2%. When N = 5, most of the windows could not extract the features; thus, we were unable to measure the performance. When N = 1, we achieved the worst result because majority voting was not performed, and many inaccurate peak points were reflected.

Figure 10 shows the performance comparison between the single-based and ensemble-based peak-detecting methods. We measured the performance using the seven classifiers and averaged all the results of each classifiers. We obtained that the ensemble-based peak-detecting method achieved the best result of 91.8 for AUC, and 88.6 for the F1 score compared with the single-based peak-detecting method. The ensemble-based peak-detecting method demonstrated an improved performance of 1.2% for AUC and 2.4% for the

F1 score compared with the second best option that applied the MAD peak detection method.

E. MULTI-CLASSIFICATION PERFORMANCE

Many studies have focused on the binary classification of stress detection [8], [28]–[30]. However, we need to develop a system to detect the stress levels above certain limit since stress is harmless at a certain level [31]. Thus, we extend our experiment to evaluate the performance using multi-classification.

Table 4 lists the performance comparison between the state-of-the-art method by *Schmidt, et al* [23], and the proposed OMDP method in the multi-classification task for
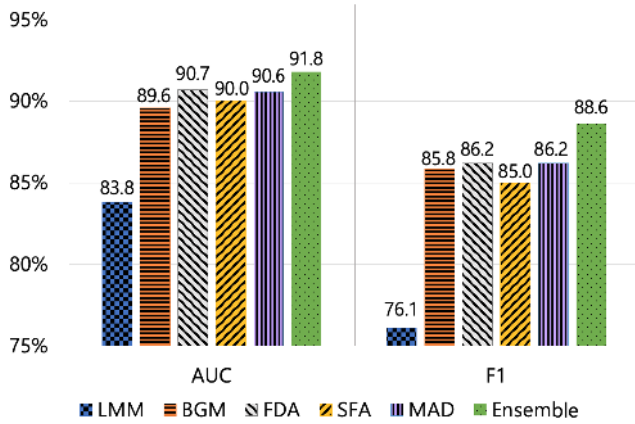
FIGURE 10. Performance comparison between single-based and ensemble-based peak-detecting methods.

tri-class: baseline, stress, and amusement. We verify that the proposed OMDP method demonstrates higher performance than the existing method. On average, the proposed method achieves a higher performance of about 9.05% for AUC, and 8.90% for the F1 score than the existing method.
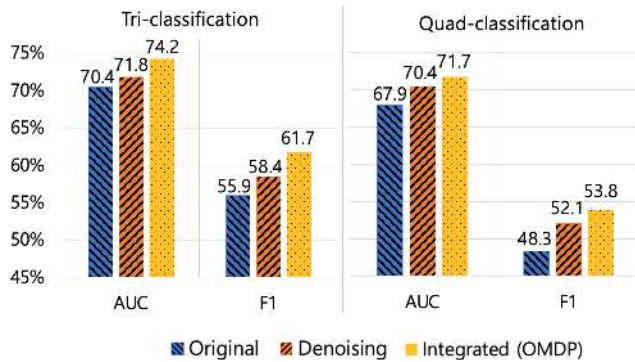

FIGURE 11. Performance of tri-classification and quad-classification tasks.

To precisely verify the generalized performance of the proposed OMDP method, we also performed the multi-classification task for quad-class similar to the tri-class. We used data that are labeled as stress, baseline, amusement, and meditation in the multi-classification. Figure 11 shows the result of the multi-classification task by applying our suggested method step-by-step: original, denoising without peak-detection, and the proposed OMDP method. We obtained the average result of the seven classifiers and showed the average result of each of the five proposed peak-detection methods when the ensemble-based peak-detecting method is not applied. When the proposed OMDP method is applied, the multi-classification task for tri-class achieves the best performance with 74.2 for AUC and 61.7 for the F1 score. The multi-classification task for quad-class also achieved the best performance with 71.7 for AUC and 53.8 for the F1 score.

Table 5 presents the result of the AUC and F1 score when the multi-classification task for tri-class and quad-class are performed using the proposed method step-by-step in the seven classifiers. The bold marks indicate the performance of the proposed OMDP method, and the gray marks indicate the best results of each classifier. The proposed OMDP method achieves the best results in four out of the seven classifiers in tri-class and quad-class. On average, we obtained that the proposed OMDP method achieves the best result in the multi-classification task as well as in the binary classification task.

IV. CONCLUSION
In this paper, we proposed an orchestrating multiple denoising and peak-detecting (OMDP) method by integrating various well-known lightweight denoising and peak-detecting methods. we utilized the advantages of the PPG signal, which is non-invasive, and can be widely used in wearable devices. We solved an inaccurate feature extraction problem due to low-quality signals with many noises by applying OMDP method. The noise was filtered and efficiently eliminated by analyzing the original PPG signals using a two-step process in terms of frequency and time. In addition, we applied the ensemble method when peaks were detected to determine the accurate peak point and extract the accurate features. The proposed OMDP method demonstrated superior stress-detection performance over the existing methods with multiple classifiers. This result opens up the possibility to monitor our health in real-time using only a single PPG signal only. The proposed OMDP method also provides the possibility of detecting elaborate stress levels by measuring the stress-detection performance at up to four levels as well as binary levels.

REFERENCES
[1] G. P. Chrousos and P. W. Gold, "The concepts of stress and stress system disorders: Overview of physical and behavioral homeostasis," *J. Amer. Med. Assoc.*, vol. 267, no. 9, pp. 1244–1252, Mar. 1992.

[2] G. P. Chrousos, "Stress and disorders of the stress system," *Nature Rev. Endocrinol.*, vol. 5, no. 7, p. 374, 2009.

[3] H. Thapliyal, V. Khalus, and C. Labrado, "Stress detection and management: A survey of wearable smart health devices," *IEEE Consum. Electron. Mag.*, vol. 6, no. 4, pp. 64–69, Oct. 2017.

[4] A. Mariotti, "The effects of chronic stress on health: New insights into the molecular mechanisms of brain–body communication," *Future Sci. OA*, vol. 1, no. 3, p. 23, Nov. 2015.

[5] E. Andreou, E. C. Alexopoulos, C. Lionis, L. Varvogli, C. Gnardellis, G. P. Chrousos, and C. Darviri, "Perceived stress scale: Reliability and validity study in Greece," *Int. J. Environ. Res. Public Health*, vol. 8, no. 8, pp. 3287–3298, Aug. 2011.

[6] A. Dupre, S. Vincent, and P. A. Iaizzo, "Basic ecg theory, recordings, and interpretation," in *Handbook of Cardiac Anatomy, Physiology, and Devices*. Springer, 2005, pp. 191–201.

[7] S. Greene, H. Thapliyal, and A. Caban-Holt, "A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health," *IEEE Consum. Electron. Mag.*, vol. 5, no. 4, pp. 44–56, Oct. 2016.

[8] H.-M. Cho, H. Park, S.-Y. Dong, and I. Youn, "Ambulatory and laboratory stress detection based on raw electrocardiogram signals using a convolutional neural network," *Sensors*, vol. 19, no. 20, p. 4408, Oct. 2019.

[9] J. Zhai and A. Barreto, "Stress detection in computer users through non-invasive monitoring of physiological signals," *Blood*, vol. 5, pp. 1–6, 2008.

[10] A. Esmaili, M. Kachuee, and M. Shabany, "Nonlinear cuffless blood pressure estimation of healthy subjects using pulse transit time and arrival time," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 12, pp. 3299–3308, Dec. 2017.

[11] K. A. Reddy, B. George, and V. J. Kumar, "Use of Fourier series analysis for motion artifact reduction and data compression of photoplethysmographic signals," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 5, pp. 1706–1711, May 2009.

[12] G. Joseph, A. Joseph, G. Titus, R. M. Thomas, and D. Jose, "Photoplethysmogram (PPG) signal analysis and wavelet de-noising," in *Proc. Annu. Int. Conf. Emerg. Res. Areas, Magn., Mach. Drives (AICERA/iCMMD)*, Jul. 2014, pp. 1–5.

[13] Z. Zhang, Z. Pi, and B. Liu, "TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 522–531, Feb. 2015.

[14] S. Hanyu and C. Xiaohui, "Motion artifact detection and reduction in PPG signals based on statistics analysis," in *Proc. 29th Chin. Control Decis. Conf. (CCDC)*, May 2017, pp. 3114–3119.

[15] P. van Gent. (2016). *Analyzing a Discrete Heart Rate Signal Using Python, a Tech Blog About Fun Things With Python and Embedded Electronics*. Google Scholar. [Online]. Available: http://www.paulvangent.com/2016/03/15/analyzing-a-discrete-heartrate-signal-using-python-part-1/

[16] M. Elgendi, I. Norton, M. Brearley, D. Abbott, and D. Schuurmans, "Systolic peak detection in acceleration photoplethysmograms measured from emergency responders in tropical conditions," *PLoS ONE*, vol. 8, no. 10, Oct. 2013, Art. no. e76585.

[17] B. N. Li, M. C. Dong, and M. I. Vai, "On an automatic delineator for arterial blood pressure waveforms," *Biomed. Signal Process. Control*, vol. 5, no. 1, pp. 76–81, Jan. 2010.

[18] D.-G. Jang, U. Farooq, S.-H. Park, and M. Hahn, "A robust method for pulse peak determination in a digital volume pulse waveform with a wandering baseline," *IEEE Trans. Biomed. Circuits Syst.*, vol. 8, no. 5, pp. 729–737, Oct. 2014.

[19] D.-G. Jang, S. Park, M. Hahn, and S.-H. Park, "A real-time pulse peak detection algorithm for the photoplethysmogram," *Int. J. Electron. Electr. Eng.*, vol. 2, no. 1, pp. 45–49, 2014.

[20] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, and S. J. van der Walt, "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, 2020.

[21] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers Public Health*, vol. 5, p. 258, Sep. 2017.

[22] R. Mccraty and F. Shaffer, "Heart rate variability: New perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk," *Global Adv. Health Med.*, vol. 4, no. 1, pp. 46–61, Jan. 2015.

[23] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interact.*, Oct. 2018, pp. 400–408.

[24] S. Heo, I. Kim, S. Kwon, H. Lee, and J. Lee, "PPG signal preprocessing and comparison study with learning-based model for stress detection," *J. Korean Inst. Commun. Sci.*, pp. 595–596, Aug. 2020.

[25] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.

[26] A. Koul, C. Becchio, and A. Cavallo, "Cross-validation approaches for replicability in psychology," *Frontiers Psychol.*, vol. 9, p. 1117, Jul. 2018.

[27] S. Saeb, L. Lonini, A. Jayaraman, D. C. Mohr, and K. P. Kording, "The need to approximate the use-case in clinical machine learning," *GigaScience*, vol. 6, no. 5, pp. 1–9, May 2017.

[28] J. R. M. Fernández and L. Anishchenko, "Mental stress detection using bioradar respiratory signals," *Biomed. Signal Process. Control*, vol. 43, pp. 244–249, May 2018.

[29] E. Vildjiounaite, J. Kallio, V. Kyllönen, M. Nieminen, I. Määttänen, M. Lindholm, J. Mäntyjärvi, and G. Gimel'farb, "Unobtrusive stress detection on the basis of smartphone usage data," *Pers. Ubiquitous Comput.*, vol. 22, no. 4, pp. 671–688, Aug. 2018.

[30] A. O. Akmandor and N. K. Jha, "Keep the stress away with SoDA: Stress detection and alleviation system," *IEEE Trans. Multi-Scale Comput. Syst.*, vol. 3, no. 4, pp. 269–282, Oct. 2017.

[31] Y. S. Can, N. Chalabianloo, D. Ekiz, and C. Ersoy, "Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study," *Sensors*, vol. 19, no. 8, p. 1849, Apr. 2019.

SEONGSIL HEO received the B.S. degree from the College of Computer Science, Kookmin University, South Korea, where she is currently pursuing the M.S. degree. Her research interests include time series analysis, machine learning, and artificial intelligence.

SUNYOUNG KWON received the B.S. degree in computer science from Pusan National University, South Korea, in 2001, and the M.S. degree in bioinformatics and the Ph.D. degree in electrical and computer engineering from Seoul National University, South Korea, in 2014 and 2018, respectively. From 2018 to 2020, she worked with NAVER Corporation, South Korea. Previously, she was with Korea Communication Agency and LG Electronics Inc., South Korea. She is currently an Assistant Professor with the School of Biomedical Convergence Engineering, Pusan National University, South Korea. Her research interests include bioinformatics, drug discovery, graph neural networks, machine learning, and big-data analytics.

JAEKOO LEE (Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Seoul National University, South Korea, in 2018. From 2011 to 2013, he was with LG Electronics Research and Development Campus, Seoul, South Korea. He was a Data Scientist with SK Telecom Corporation, in 2018. He is currently an Assistant Professor with the College of Computer Science, Kookmin University, South Korea. His research interests include artificial intelligence (AI), machine learning, deep learning, data science, and their applications.

• • •