IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 23 January 2024, accepted 29 January 2024, date of publication 2 February 2024, date of current version 8 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3361684

RESEARCH ARTICLE

Stress Classification Using ECGs Based on a Multi-Dimensional Feature Fusion of LSTM and Xception

CHEOL HO SONG, JIN SU KIM, JAE MYUNG KIM, AND SUNGBUM PAN, (Member, IEEE)
Research Institute of IT, Chosun University, Gwangju 61452, South Korea

Corresponding author: Sungbum Pan (sbpan@chosun.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2017R1A6A1A03015496, and in part by NRF Grant funded by the Ministry of Science and ICT (MSIT) under Grant NRF-2021R1A2C1014033.

ABSTRACT In the information age, people are increasingly being exposed to stress as societies are experiencing sudden changes fueled by advancements in cutting-edge scientific technology and the Information Technology (IT) industry. Consequently, research has been actively conducted on stress classification to mitigate psychological and physical diseases caused by constantly feeling stressed. Specifically, the number of studies examining electrocardiograms (ECGs), which record biosignals that provide insight into the response level of the body's autonomic nervous system, has increased. However, previous studies on stress classification based on ECG used only one-dimensional feature data, thus entailing difficulties in analyzing the data more closely and comprehensively owing to bias toward a specific aspect. Therefore, to overcome the limitations of conventional stress classification based on ECGs, this paper developed a stress classification method based on multi-dimensional feature fusion of LSTM and Xception using ECGs from which outliers have been removed. Experimental results showed that applying multi-dimensional feature fusion of the weighted average method using ECG data with outlier signals removed resulted in a stress classification of 99.51%, a 1.25% improvement from previous studies which used only one-dimensional feature data of ECGs, thus highlighting the excellent performance of the proposed stress classification method using ECGs based on multi-dimensional feature fusion of LSTM and Xception.

INDEX TERMS Electrocardiography, stress, multidimensional systems, network systems.

I. INTRODUCTION

Humans are increasingly experiencing stress in a constantly changing modern society as the IT industry advances rapidly. Stress refers to complex physiological and psychological responses of an individual induced by mental and physical stimuli [1]. Prolonged exposure to stress leads to various health-related issues, such as cardiovascular diseases, weakened immune systems, and mood disorders [2], [3]. Therefore, more research efforts are required to accurately classify and quantify stress states to develop treatment methods or programs optimized for individuals to effectively manage stress and reduce the risk of developing certain diseases [4], [5], [6].

Among the representative biosignals used in stress classification research, electrocardiogram offers various advantages over other biosignals, including non-invasiveness, ease of use, and continuous real-time monitoring feature. Furthermore, heart rate variability (HRV)—a derivative parameter of an ECG—provides extensive information on activities of the autonomic nervous system exhibiting the body's response to stress.

An analysis of HRV examines the changes in the time intervals of heartbeats and is performed based on the periodic characteristics of ECG. An HRV analysis, which examines the changes in time interval between the R-peaks, is a useful stress classification tool for diagnosing different heart

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

diseases and evaluating stress levels [7]. However, ECGs are highly sensitive to motion artifacts—disturbances caused by movements of patients or external vibrations—ultimately affecting the accuracy and reliability of ECGs. Motion artifacts disrupt the periodical characteristics of ECG signals and induce errors in detecting R-peaks. Biased or unreliable HRV estimates can result owing to inaccuracies in the calculation of the RR interval [8], [9]. Therefore, outlier signals generated by motion artifacts must be removed for accurate stress classification using ECGs.

Previous studies on stress classification using ECGs mostly employed models based on one-dimensional features, thus making it difficult to capture the overall complexity of physiological changes associated with stress as misclassification or performance degradation may result from bias toward a specific aspect. In other words, using one-dimensional features leads to the second-best classification accuracy owing to the incapability of a stress classification model and non-optimal performance [10]. Therefore, multi-dimensional feature information must be inspected to accurately analyze complex stress pattern data.

This study developed a stress classification method of using multi-dimensional feature fusion of LSTM and Xception to solve the problem of performance degradation resulting from using only one-dimensional feature information when analyzing ECG signals reflecting stress conditions. The objective of this study is to complement the problems that arise from using only single-dimensional feature information by embracing both temporal and spatial features when classifying stress using ECGs. For this purpose, we developed a stress classification method that utilizes multi-dimensional feature fusion of LSTM and Xception. This study's approach stands out in the dual network architecture. The sequential dependencies of ECG signals are appropriately captured through the LSTM network, while the Xception network extracts spatial features from the spectrogram images. Afterwards, a weighted average mechanism is used to obtain a holistic view of the ECG signal and simultaneously capture complex temporal patterns and spatial nuances. This synergistic approach distinguishes this study from existing research methods and enriches our understanding of physiological state classification. In addition, unlike existing studies that focus only on temporal or spatial aspects, this research method is excellent at integrating both dimensions. Additionally, it allows for more nuanced and comprehensive analysis, resulting in superior classification accuracy compared to approaches that emphasize only a single dimension. The proposed method demonstrated a classification accuracy of 99.51% when multi-dimensional feature fusion of the weighted average method was applied using ECG data with outliers removed. The classification performance for the method using one-dimensional feature information and LSTM and the method using two-dimensional feature information and Xception improved by 18.09% and 0.59%, respectively. In addition, the performance improved by at least 1.25% compared with other studies that used the same database.

The remainder of this paper is organized as follows. In Section II, existing technologies using ECG signals are analyzed. Section III explains stress classification based on multi-dimensional feature fusion and outlier signal removal in ECG proposed in this paper. The method and result of the experiment conducted for the proposed stress classification system are analyzed in Section IV. Finally, Section V presents the conclusions drawn.

II. RELATED STUDIES

This section analyzes the technologies applied to studies using one-dimensional feature vectors and studies using two-dimensional feature information among previous stress classification studies using ECGs. Studies on stress classification using the HRV of ECG signals commonly utilize the HRV parameter as a feature by analyzing the changes in the time interval between the RR interval of ECG signals.

Mekruksavanich [11] conducted an experiment using the ECG data of the wearable stress and affect detection (WESAD) dataset and conducted preprocessing by applying the Butterworth filter and min-max normalization. Their proposed StressNext network demonstrated a classification accuracy of 87.10%. To remove noise in ECG signals of the WESAD dataset, Schmidt [12] applied a band-pass filter and extracted HRV features, including pNN50, nLF, nHF, and mHR. Based on the extracted feature information, a stress classification accuracy of 85.44% was obtained using linear discriminant analysis (LDA) as a classifier. In the study by Kang [13], ECG signals from the CLAS Database were used. To remove noise, the sampling frequency was set to 330Hz and the cutoff frequency was set to 120Hz, and a Butterworth filter and Low Pass filter were applied. Features were extracted for R-S peak, RR Interval, and Q-T Interval, respectively, and stress classification accuracy of 97.6% was confirmed through SVM. Chandrasekaran [14] extracted HRV features without applying preprocessing to ECG signals. Five types of HRV features, including SDNN, RMSSD, and NN50, were extracted, with the value combining NN50, pNN50, and SDNN additionally used as a feature. A classification accuracy of 93% was obtained using reservoir computing and logistic regression. In a study by Munla [15], a band-pass filter was applied to reduce muscle noise and interferences of 60 Hz included in ECG signals using the Drive open database provided by MIT-BIH. Subsequently, the R-peaks were detected after detecting the QRS complex using the Pan–Tompkins algorithm, based on which HRV features were extracted. A total of 12 HRV parameters, including SDNN, VLF, LF, and mHR, which represent the mean heart rate in time and frequency bands, were obtained. A stress classification accuracy of 83.33% was obtained via support vector machine (SVM) that used an RBF kernel.

Table 1 presents the details of previous studies on stress classification that used one-dimensional feature information.

TABLE 1. Previous studies on stress classification using one-dimensional feature information.

Author	Data base	Preprocessing	Feature extraction	Classifier	Performance (%)
S. Mekruksavanich et al. [11]	WESAD [12]	Butterworth filter	Segmented Signal	Stress Next	87.1
P. Schmidt et al. [12]	WEASD	Band-pass filter	HRV	LDA	85.44
Kang et al. [13]	CLAS [16]	Butterworth filter Low-pass filter	R-S peak, RR Interval Q-T Interval	SVM	97.6
S.T. Chandrasekaran. et al. [15]	WEASD	-	HRV	Reservoir computing, logistic regression	93
N. Munla et al. [16]	Drive [17]	Band-pass filter	HRV	SVM-RBF	83.33

TABLE 2. Previous studies on stress classification using one-dimensional feature information.

Author	Database	Feature extraction	Classifier	Performance (%)
L. Liakopoulos et al. [18]	WESAD	Spectrogram	CNN	96.79
M. G. Kang et al. [19]	WEASD, ST Change [20]	Signal Image, FFT	CNN-LSTM	94.13 98.26
Z. Ahmad et al. [21]	Self-acquired	GWT, DWT, Signal Image	CNN	85.45
M. Amin et al. [22]	Drive	CWT, Scalogram	Xception	98.11
S. Ishaque et al. [23]	WESAD	FFT	VGG16	90.62

In previous studies on ECG-based stress classification that used a one-dimensional feature vector as an input of a classifier, noise was removed mostly through digital filtering in the preprocessing step. Furthermore, various HRV parameters were used in the feature extraction step by analyzing the HRV in time and frequency bands.

Next, previous studies on stress classification that used feature information extracted from ECG as a two-dimensional input of a network are analyzed. Two-dimensional feature information frequently used in many studies can be divided into a form of an image of preprocessed ECG signals and the time–frequency band to which the Fourier transform is applied. Liakopoulos [18] segmented signals by one-minute window to the ECG data of the WESAD dataset. The segmented ECG signals were converted to a spectrogram to form two-dimensional image. The two-dimensional image was then input into a convolutional neural network (CNN), obtaining a stress classification accuracy of 96.79%. Kang et al. [19] used both the WESAD dataset and ST Change database, with a low-pass filter with a cut-off frequency of 150 Hz applied for noise removal. The information of a stress state was classified based on the difference between the R-wave and S-wave of ECGs. Two-dimensional image data used as a network input were analyzed in the time and frequency bands. The waveform of ECG signals was expressed as an image in the time band, with a fast Fourier transform (FFT) applied to express the data as a spectrogram image in the frequency band. CNN-LSTM was used to verify the stress classification performance wherein accuracies of 94.13% and 98.26% were observed in the time and frequency bands, respectively. Ahmad and Khan [21] created a stress-induced environment for six minutes and then collected ECG data from 15 subjects. To create a network input, two-dimensional images were generated through the signal image, discrete Fourier transform (DFT), and Gabor wavelet transform. The successive RR intervals of ECG were arranged into time-series segments to form an image from ECG signals. DFT was applied to convert a space domain image into a frequency domain image. The Gabor wavelet transform was applied to analyze the data in the time–frequency band. A CNN network was used for stress classification, with decision-level fusion applied to three image inputs, resulting in an accuracy of 85.45%. Amin et al. [22] conducted a study on the stress level detection of a driver to reduce the risks caused by a driver's health conditions. Hence, the Drive database—a public database provided by Physionet—was used. A Butterworth band-pass filter of 0.5–100 Hz and a notch filter of 50 Hz were applied to remove noise in the preprocessing

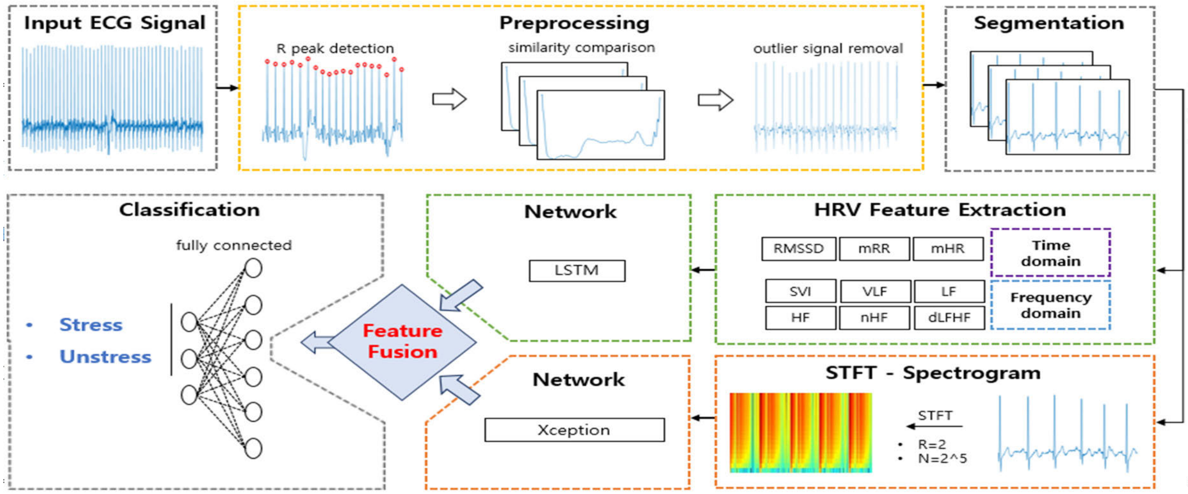

FIGURE 1. Flowchart of proposed stress classification using multi-dimensional feature fusion-based ECGs.

step. The continuous wavelet transform (CWT) was applied to analyze the time–frequency components of ECG signals in the feature extraction step before the time–frequency components were converted to scalogram images to form an image input. In the study by Ishaque et al. [23], a stress classification study was conducted using the WESAD Dataset, and the conversion of one-dimensional electrocardiogram data into two-dimensional images was emphasized. For this purpose, FFT was applied to the one-dimensional ECG signal and the spectrum image converted to two dimensions was used as a network input. VGG16 and network compression technology were used, and the stress and non-stress classification results showed 90.62% performance.

Table 2 presents the details of previous studies on stress classification that used two-dimensional feature information. In previous studies on stress classification based on ECGs that used two-dimensional feature information as a classifier input, a Fourier transform or wavelet transform was applied to analyze the time–frequency band. Subsequently, a two-dimensional image was generated by converting it to a spectrogram.

Previous studies on stress classification using ECG mostly did not remove noise during the preprocessing step or removed the noise using digital filtering. However, such preprocessing does not remove outlier signals of which periodic characteristic of ECG has been disrupted, ultimately resulting in performance degradation of stress classification. Additionally, after analyzing ECG containing stress state data, stress classification performance was verified using one-dimensional and two-dimensional feature information. However, using only one-dimensional feature information cannot accurately analyze the stress information reflected in ECG. In other words, while one-dimensional feature information can be used to analyze the overall trend or mean value of signals, it cannot be applied to analyze frequency variations over time. In contrast, two-dimensional feature information can be used to analyze frequency information that varies over time but cannot investigate the information that can be analyzed using one-dimensional feature information. Therefore, the analysis should be performed based on the fusion of one-dimensional and two-dimensional feature information to accurately analyze ECG data that contain complex stress state information.

III. STRESS CLASSIFICATION USING THE PROPOSED MULTI-DIMENSIONAL FEATURE FUSION-BASED ELECTROCARDIOGRAM

Figure 1 [24] shows the overall flowchart of stress classification using the multi-dimensional feature-fusion-based ECG method proposed in this paper. Preprocessing is performed to remove outlier signals for resolving the disruption of periodic characteristics of signals caused by motion artifacts of ECG signals. This process involves detecting the R-peaks of ECG signals and then comparing the similarity between the RR interval to remove outliers. Subsequently, stress classification is proceeded by applying feature level fusion using one-dimensional and two-dimensional feature information to precisely analyze complicated stress states.

A. OUTLIER SIGNAL REMOVAL BASED ON THE RR INTERVAL OF ECGs

Removing noises is particularly important in applications that require using biosignals and must be performed during the preprocessing step. Various digital filtering methods are used, such as a band-pass filter [25], median filter [26], and moving average filter [27], which are useful and easy tools for removing noises recorded in ECGs including interference of power cables or electrodes. However, removing noises by applying only digital filters to original signals cannot effectively remove outlier signals of which the morphological nature of signals has been damaged by motion artifacts, as shown in Figure 2. Therefore, noises containing unnecessary information and outlier signals inducing errors in the information analysis must be removed when performing stress classification using ECGs.

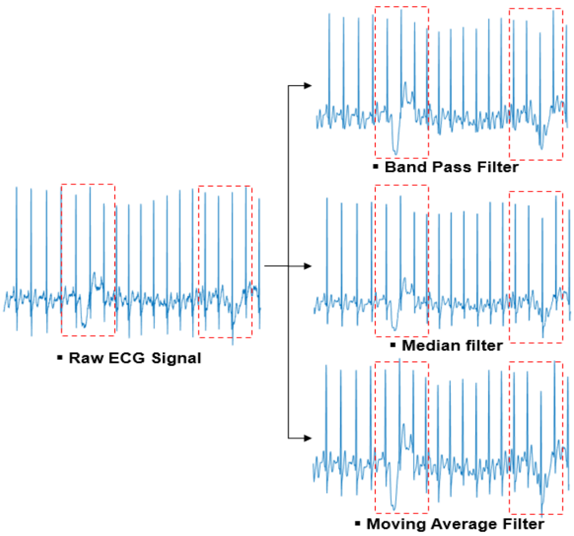

FIGURE 2. Results of applying various digital filters to ECG signals containing noise.

The Pan–Tompkins algorithm is applied to original ECG signals in the WESAD dataset to detect a band-pass filter and the R-peaks. First, a band-pass filter consisting of a low-pass filter and a high-pass filter is used to remove unnecessary noises. The cut-off frequency of the band-pass filter used in this study is 0.001–100 Hz, where frequency information within the cut-off frequency band is allowed; frequency information outside the cut-off frequency band is attenuated. This process enables analyzing the information of a very low frequency (VLF) ECG containing stress. R-peaks are detected by applying the Pan–Tompkins algorithm to the ECG signals in which noise has been removed through filtering. In this process, the QRS complex is separated in the 5–15 Hz band of the ECG. Subsequently, R-peaks of ECG are detected by applying the algorithm consisting of the information of the steepest part by finding the slope of the QRS complex. Based on the detected R-peaks, the RR interval of ECG is extracted to compare the similarity. Cosine similarity—used to determine the cosine angle between two vectors—is used to calculate the similarity between the RR-intervals as shown in Equation (1) [28].

$$Cosine\ Similarity = \frac{\boldsymbol{AB}}{\|\boldsymbol{A}\|\,\|\boldsymbol{B}\|} \tag{1}$$

where vector AB represents the inner product between two vectors where cosine similarity output has a value between –1 and 1; -1 indicates that two vectors are completely opposite, 0 indicates that two vectors are independent, and 1 indicates that two vectors are completely identical. A representative signal is selected among the extracted successive RR-interval signals, and then the threshold is set by comparing the similarity between the representative signal and ECG signal of the remaining RR-intervals. The threshold is computed using Equation (2).

$$Th = Cosine\ Similarity\ Average \times \alpha \tag{2}$$

where Th is the threshold, and α indicates the weight for setting the threshold. Therefore, the threshold is set by assigning a weight to the mean of the overall similarity after comparing the cosine similarity between the RR-intervals of ECG. A weight of 0.95 is assigned, and then outlier signals of the RR-interval with the similarity below the threshold are detected and removed. Figure 3 shows the process of detecting and removing outlier signals by comparing the similarity between successive RR-intervals.

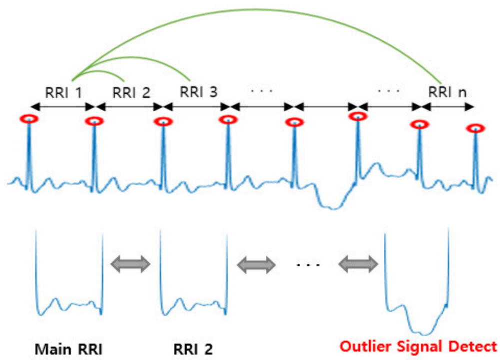

FIGURE 3. Process of detecting outlier signals of ECG through RR-interval similarity comparison.

B. STRESS CLASSIFICATION USING MULTI-DIMENSIONAL FEATURE FUSION-BASED ECG

In this paper, the accuracy and reliability of stress classification are improved by extracting features from preprocessed ECG signals and applying fusion to the extracted features. The feature extraction process consists of two main steps. In the first step, HRV features are extracted and used as an input of the LSTM network. In the second step, Short Time Fourier Transform (STFT) is applied to ECG to generate a spectrogram image to be used as an input of the Xception network.

HRV features reflecting the body's reaction under stress are extracted to compose one-dimensional feature data. In the HRV feature extraction process, nine major features are computed in the time and frequency bands of the preprocessed ECG data. Three features—mRR, mHR, and RMSSD—are extracted in the time band. mRR, representing the mean time interval between R-peaks, is computed using Equation (3). Furthermore, mHR, representing the mean heart rate, is computed using Equation (4), while RMSSD is calculated as the root mean square of the differences between successive RR-intervals, as shown in Equation (5). In Equations (3)–(5) [29], [30], i represents a natural number, whereas N represents the last interval of successive RR-intervals. In addition,

RR_i indicates the i-th RR-interval.

$$mRR = \frac{\sum_{i=1}^{N}(RR_i)}{N} \tag{3}$$

$$mHR = \frac{\sum_{i=1}^{N}(60000/RR_i)}{N} \tag{4}$$

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(RR_{i+1}-RR_i)^2} \tag{5}$$

TABLE 3. HRV feature information of the frequency band used in the study.

Features	Description	Frequency band (Hz)	Units
HF	Power in the high-frequency band	0.15 ~ 0.4	ms^2
LF	Power in the low-frequency band	0.04 ~ 0.15	ms^2
VLF	Power in the very low-frequency band	≤0.04	ms^2
nHF	Normalized HF $nHF = (HF/(Total\ Power - VLF)) \times 100$		%
dLFHF	$LF - HF$		
SVI	LF/HF		

A total of six features are extracted in the frequency band to which the FFT is applied. A high frequency (HF) can be used to analyze the activities of the parasympathetic system, while a low frequency (LF) can be used to analyze the activities of the sympathetic nervous system. Six types of features extractable based on the HF and LF from different frequency bands are shown in Table 3 [31]; the extracted HRV feature information of time and frequency bands is input in the LSTM network as a one-dimensional feature vector.

Time–frequency components are analyzed using STFT to form two-dimensional feature data. ECG signals are non-stationary signals in which the signal attributes change over time. Thus, the varying attributes are difficult to explain simply based on the information of a single frequency band [32]. In this paper, STFT is applied to extract time–frequency features for analyzing ECG signals that are non-stationary. The STFT is shown in Equation (6) [33].

$$STFT\ (\tau,\omega) = \int_{-\infty}^{\infty} x(\tau)\omega(\tau - t)e^{-j\omega t}dt \tag{6}$$

$\omega(\tau)$ is the analyzed input signal, which is computed through the FFT using the window function $\omega(\tau - \mathrm{t})$ for analyzing the time–frequency features. When applying the STFT, the Hamming window length (R) is set to 2^5, while the resolution of FFT (N) is set to 2^5. The STFT is applied to the final preprocessed ECG signals, and then the signals are converted to a spectrogram to generate a two-dimensional image shown in Figure 4; the converted data are then used as an input of the Xception network.

Using one-dimensional and two-dimensional feature information separately can result in missing important information

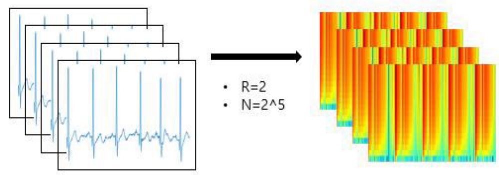

FIGURE 4. STFT application and spectrogram conversion of ECG signals.

contained in data because each dimension of feature information captures different aspects of ECG signals. The HRV feature, which is a one-dimensional feature, provides important information variability and regularity of heartbeats. However, obtaining information on the changes in the signal frequency over time is impossible. A two-dimensional feature such as a spectrogram enables the frequency variations of ECG signals to be analyzed more closely with respect to time. However, analyzing the overall trend of signals, mean values, or specific patterns of the time domain that can be analyzed based on one-dimensional HRV features is difficult. Thus, this paper examines stress classification performance by applying the fusion of one-dimensional feature information on variability and regularity of heartbeats and two-dimensional feature information from the time–frequency analysis of signals.

The strengths and characteristics of each network must be considered when selecting a network to process diverse types of data. The LSTM network is a type of recurrent neural network (RNN) highly capable of processing sequential data and appropriate for processing physiological signals with temporal dependency. Therefore, LSTM is suitable for time-series data such as HRV and for the sequence of variations in HRV features that indicate the temporal changes between heartbeats. The Xception network is a type of CNN that processes spatial hierarchy in data by utilizing convolutions that can be separated depthwise. This network is particularly efficient and effective for numerous types of image data because it separately trains feature of each channel and spatial feature. Accordingly, it can adequately train a spatial hierarchical structure in a spectrogram image by mapping temporal changes and spatial correlation between different frequencies. A two-dimensional spectrogram image of ECG signals expresses the time–frequency band; thus, it enables effective training of the spatial hierarchical structure through Xception.

LSTM and Xception networks analyze one-dimensional and two-dimensional feature information, respectively, before feature-level fusion is performed. The fused feature information is then used as an input of a fully connected layer to test the final stress classification performance. Feature fusion applied in this study involves six fusion methods of concatenation, mean, maximum, weighted average, multiplication, and addition to compare the performance. The weight for feature information of each network for the weighted average method is computed using Equations (7) and (8) based on the accuracy of each network. W_{LSTM} and $W_{Xception}$ indicate the weight of each network, and A_{LSTM} and $A_{Xception}$ indicate the accuracy of each network. Feature fusion of the weighted average method is computed using F_{LSTM} and $F_{Xception}$, which represent feature information analyzed by each network, as shown in Equation (9) [34]. Figure 5 shows the process of classifying stress through multi-dimensional feature fusion using the weighted average method.

$$W_{LSTM} = \frac{A_{LSTM}}{A_{Xception} + A_{LSTM}} \tag{7}$$

$$W_{Xception} = \frac{A_{Xception}}{A_{Xception} + A_{LSTM}} \tag{8}$$

$$Fusion = \left(F_{Xception}\right)\left(W_{Xception}\right) + (F_{LSTM})(W_{LSTM}) \tag{9}$$

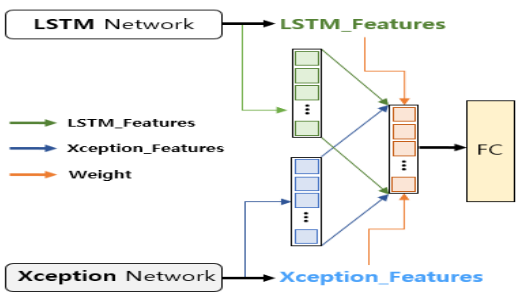

FIGURE 5. Multi-dimensional feature fusion process using the weighted average method.

IV. EXPERIMENTAL RESULTS

The WESAD dataset was used to investigate the proposed stress classification method using multi-dimensional feature fusion. The total number of subjects was 15, and the ECG data were collected at the sampling frequency of 700 Hz. The collected ECG data contained three different emotional states: neutral, stress, and amusement. Specifically, the stress state was acquired by exposing the subjects to the trier social stress test (TSST) [35] for about 10 min. This study aimed to classify the stress and non-stress states (neutral, amusement) using the ECG data of three emotional states (neutral, stress, amusement).

The experiment was conducted as shown in Table 4 to verify the performance of the proposed stress classification system. The network performance was tested using window length in experiment 1, whereas the performance was tested with respect to outlier signal removal in experiment 2. Finally, in experiment 3, the performance after applying multi-dimensional feature fusion was tested using the ECG data with outlier signals removed.

The experiment 1 analyzes performance according to changes in window length for signal segmentation. The experimental conditions are as follows. The window length is set to 5 seconds, 10 seconds, and 60 seconds and compared. The network uses an LSTM network suitable for one-dimensional feature information analysis and an Xception network suitable for two-dimensional image data analysis. As a result of the experiment, as shown in Figure 6, both LSTM and Xception showed the highest performance of 78.25% and 97.12%, respectively, when the window length was the shortest, 5 seconds. It was confirmed that the performance of LSTM improves significantly as the window length becomes shorter. The performance of Xception showed a slight improvement. In particular, this shows that setting the window length is important when analyzing one-dimensional feature information. Also, when analyzing 2D feature information, the shorter the window length, the better the performance, so this study conducts experiments with a window length of 5 seconds. In this experiment, it was shown that stress classification performance deteriorated as the window length increased, indicating that the shorter the window length, the more accurately the stress information of the ECG could be captured. Figure 7 shows the results of experiment 2, which analyzed the performance before and after removing outlier signals from the ECG signal. The experiment uses ECG signals segmented with a window length of 5 seconds, which showed the highest performance, according to previous experimental results. Removal of outlier signals is performed through similarity comparison based on the RR interval of the ECG in the same manner as Figure 3. Afterwards, the stress classification performance of each network is checked using the ECG from which outlier signals have been removed. The experiment result showed that the performance of LSTM and Xception improved by 3.17% and 1.8% to 81.42% and 98.92%, respectively, after outlier signals were removed compared with before the outlier signals were removed. This indicates that the outlier signal was not removed through digital filtering alone. This indicates that removal of outlier signals is essential when analyzing stress using an ECG.

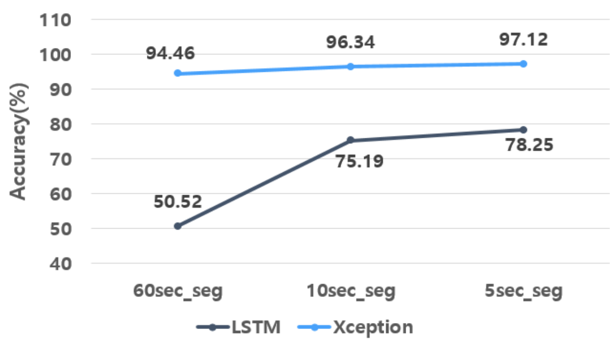

FIGURE 6. Network performance according to window length.

Subsequently, the experiment was conducted to test the performance after applying multi-dimensional feature fusion. First, HRV features were extracted using the ECG data with outlier signals removed. A total of nine HRV features were extracted, including mRR, mHR, and RMSSD from the time domain, and nHF, HF, LF, VLF, dLFHF, and SVI from the frequency domain. The extracted HRV feature information was input into the LSTM network as a one-dimensional feature vector. Subsequently, STFT was

TABLE 4. Performance comparison based on experiment method.

Step	Experiment details	Used data	Experiment result (%) Before	After	Note
1	Testing the performance of LSTM and Xception by window length	WESAD dataset (60 s, 10 s, 5 s)	Performance analysis according to the changes in the window		-
2	Testing outlier signal removal by window length	WESAD dataset (5 s)	97.12	98.92	Xception
3	Testing the performance after applying multi-dimensional feature fusion	WESAD dataset	LSTM: 81.42 Xception: 98.92	99.51	After removing outlier signals

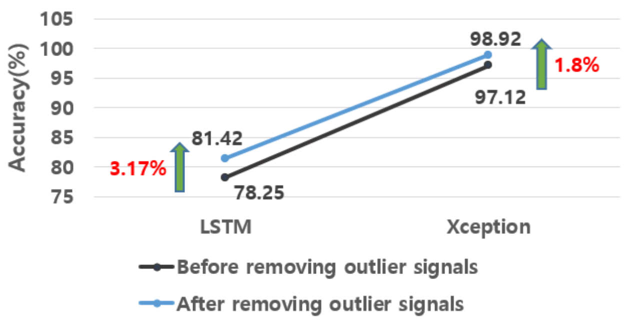

FIGURE 7. Comparison of before and after removing outlier signals from ECG signals.

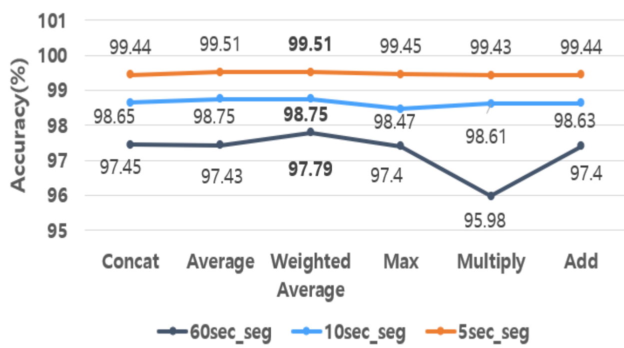

FIGURE 8. Stress classification performance according to the feature fusion method.

applied to the ECG signals to analyze the changes in the signal frequency over time. The analyzed information was converted to a spectrogram into a two-dimensional image to be input into the Xception network. Feature level fusion was applied using the fully connected layer of each network based on the information trained and analyzed by the networks. The proposed multi-dimensional feature level fusion tested the classification performance by applying six fusion methods (concatenation, mean, maximum, weighted average, multiplication, and addition). Figure 8 shows the result of testing the performance of the multi-dimensional feature fusion according to window length for signal segmentation. The performance of the multi-dimensional feature fusion method using LSTM and Xception improved as the window length decreased; the highest performance of 99.51% was obtained when multi-dimensional feature fusion was applied based on the weighted average method using the signals segmented into 5 s window lengths.

Table 5 presents a comparison of stress classification performance between the proposed method and previous methods using ECG signals. Kang et al. [13], [19] performed stress classification via an ECG analysis using the same WESAD dataset. Kang [13] performed preprocessing using a Butterworth filter and a low-pass filter. Afterwards, R-Speak, RR Interval, and Q-T Interval features were extracted and a stress classification performance of 97.6% was confirmed using SVM, a machine learning-based classifier. Kang et al. [19] removed noises using a low-pass filter with a cut-off frequency of 150 Hz and utilized the signals segmented into 1 min window. An accuracy of 98.26% was confirmed when a two-dimensional image of the signals and the image applied with FFT were input in a deep-learning-based network. Therefore, a proper noise removal technique is required to analyze ECG signals reflecting the stress state, and stress classification performance was higher as the window length for signal segmentation became shorter. Furthermore, the result signified that analyzing the data through deep learning

TABLE 5. Comparison of stress classification performance using ECG signals.

Author	Data base	Sub-jects	Signal length	Feature extraction	In put	Perfor mance (%)
Kang et al. [13]	WE SAD	15	Full-signal	R-S peak, RR Interval, Q-T Interval	1D	97.6
Kang et al. [19]	WE SAD	15	1 min	Signal Image, FFT	2D	94.13 98.26
Pro-posed Method	WE SAD	15	5 sec	HRV, STFT	1D 2D	99.51

networks is more appropriate than using machine learning. Compared with previous studies, the proposed stress classification method produced an accuracy of 99.51%, an improvement of more than 1.25%, in addition to the various strengths mentioned above.

V. CONCLUSION

Most previous studies on stress classification using deep learning networks employed single-dimension feature information. Consequently, the analyses tended to be biased toward a specific aspect, and stress states exhibiting complicated information could not be comprehensively analyzed. Thus, this study developed a stress classification method based on the multi-dimensional feature fusion of LSTM and Xception to overcome such limitations. Outlier signals were removed from ECG data to solve the problem of morphological damage in ECG caused by motion artifacts, which can degrade the performance of the proposed method. The proposed method overcame the limitations of one-dimensional and two-dimensional feature information by applying multi-dimensional feature fusion. Furthermore, the advantages of unique features in a neural network architecture were used by applying feature-level fusion to the outputs of a neural network appropriate for each dimension.

The experimental results showed that the classification performance improved by 3.17% and 1.8% using LSTM and Xception, respectively, after outlier signals in the ECG data were removed. The performance of six feature fusion methods was tested, with the weighted average method demonstrating the highest performance of 99.51%. Therefore, the proposed stress classification method of using multi-dimensional feature fusion of LSTM and Xception demonstrated outstanding performance. Future studies will focus on optimizing and lightening deep learning networks used for ECG-based stress classification.

REFERENCES

[1] H. Y. Lee and H. B. Joo, "Stress and mental illness," *Korean Med. Assoc.*, vol. 56, no. 6, pp. 471–477, Jun. 2013.

[2] S. Hey, A. Gharbi, B. von Haaren, K. Walter, N. König, and S. Löffler, "Continuous noninvasive pulse transit time measurement for psycho-physiological stress monitoring," in *Proc. Int. Conf. eHealth, Telemed. Social Med.*, 2009, pp. 113–116.

[3] 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.

[4] S. Cohen, J. D. Denise, and G. E. Miller, "Psychological stress and disease," *JAMA*, vol. 298, no. 14, pp. 1685–1687, Oct. 2007, doi: 10.1001/jama.298.14.1685.

[5] S. L. Noble, "Control-theoretic treatment scheduling for posttraumatic stress disorder," in *Proc. Amer. Control Conf.*, 2013, pp. 722–727.

[6] C. H. Song, J. S. Kim, G. H. Choi, and S. Pan, "Xception-based stress classification by removing ECG signal's outliers," *J. Korean Inst. Inf. Technol.*, vol. 21, no. 5, pp. 75–81, May 2023.

[7] M. Umair, N. Chalabianloo, C. Sas, and C. Ersoy, "HRV and stress: A mixed-methods approach for comparison of wearable heart rate sensors for biofeedback," *IEEE Access*, vol. 9, pp. 14005–14024, 2021.

[8] M. Blanco-Velasco, B. Weng, and K. E. Barner, "Synthetic structure of industrial plastics," in *Proc. Comput. Cardiol.*, 2006, pp. 917–920.

[9] D. Berwal, C. R. Vandana, S. Dewan, C. V. Jiji, and M. S. Baghini, "Motion artifact removal in ambulatory ECG signal for heart rate variability analysis," *IEEE Sensors J.*, vol. 19, no. 24, pp. 12432–12442, Dec. 2019, doi: 10.1109/JSEN.2019.2939391.

[10] S. M. Shahid, S. Ko, and S. Kwon, "Performance comparison of 1D and 2D convolutional neural networks for real-time classification of time series sensor data," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Jan. 2022, pp. 507–511.

[11] S. Mekruksavanich, N. Hnoohom, and A. Jitpattanakul, "A deep residual-based model on multi-branch aggregation for stress and emotion recognition through biosignals," in *Proc. 19th Int. Conf. Electr. Eng./Electron. Comput. Telecommun. Inf. Technol. (ECTI-CON)*, Prachuap Khiri Khan, Thailand, May 2022, pp. 1–4.

[12] 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.*, New York, NY, USA, Oct. 2018, pp. 400–408.

[13] M. Kang, S. Shin, G. Zhang, J. Jung, and Y. T. Kim, "Mental stress classification based on a support vector machine and naive Bayes using electrocardiogram signals," *Sensors*, vol. 21, no. 23, p. 7916, Nov. 2021.

[14] S. T. Chandrasekaran, S. P. Bhanushali, I. Banerjee, and A. Sanyal, "A bio-inspired reservoir-computer for real-time stress detection from ECG signal," *IEEE Solid-State Circuits Lett.*, vol. 3, pp. 290–293, 2020.

[15] N. Munla, M. Khalil, A. Shahin, and A. Mourad, "Driver stress level detection using HRV analysis," in *Proc. Int. Conf. Adv. Biomed. Eng. (ICABME)*, Beirut, Lebanon, Sep. 2015, pp. 61–64.

[16] V. Markova, T. Ganchev, and K. Kalinkov, "CLAS: A database for cognitive load, affect and stress recognition," in *Proc. Int. Conf. Biomed. Innov. Appl. (BIA)*, Nov. 2019, pp. 1–4.

[17] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005, doi: 10.1109/TITS.2005.848368.

[18] L. Liakopoulos, N. Stagakis, E. I. Zacharaki, and K. Moustakas, "CNN-based stress and emotion recognition in ambulatory settings," in *Proc. 12th Int. Conf. Inf. Intell. Syst. Appl. (IISA)*, Chania Crete, Greece, Jul. 2021, pp. 1–8.

[19] M. Kang, S. Shin, J. Jung, and Y. T. Kim, "Classification of mental stress using CNN-LSTM algorithms with electrocardiogram signals," *J. Healthcare Eng.*, vol. 2021, pp. 1–11, Jun. 2021.

[20] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, Jun. 2000.

[21] Z. Ahmad and N. M. Khan, "Multi-level stress assessment using multi-domain fusion of ECG signal," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Montreal, QC, Canada, Jul. 2020, pp. 4518–4521.

[22] M. Amin, K. Ullah, M. Asif, A. Waheed, S. U. Haq, M. Zareei, and R. R. Biswal, "ECG-based driver's stress detection using deep transfer learning and fuzzy logic approaches," *IEEE Access*, vol. 10, pp. 29788–29809, 2022.

[23] S. Ishaque, N. Khan, and S. Krishnan, "Detecting stress through 2D ECG images using pretrained models, transfer learning and model compression techniques," *Mach. Learn. Appl.*, vol. 10, Dec. 2022, Art. no. 100395.

[24] C. H. Song, "A study on stress classification using electrocardiograms based on the multi-dimensional feature fusion of LSTM and xception," M.S. thesis, Dept. Electron. Eng. Chosun Univ., Dong-gu, Gwangju, Korean, 2023.

[25] G. S. Mihov and D. H. Badarov, "Application of a reduced band-pass filter in the extraction of power-line interference from ECG signals," in *Proc. 29th Int. Sci. Conf. Electron. (ET)*, Sep. 2020, pp. 1–4.

[26] S. Aadhirai and D. N. Jamal, "Feature extraction and analysis of renal abnormalities using fuzzy clustering segmentation and SIFT method," in *Proc. 3rd Int. Conf. Biosignals, Images Instrum. (ICBSII)*, Chennai, India, Mar. 2017, pp. 1–5.

[27] Y.-W. Bai, W.-Y. Chu, C.-Y. Chen, Y.-T. Lee, Y.-C. Tsai, and C.-H. Tsai, "The combination of Kaiser window and moving average for the low-pass filtering of the remote ECG signals," in *Proc. 17th IEEE Symp. Comput.-Based Med. Syst.*, Jun. 2004, pp. 273–278.

[28] D. Wang, H. Lu, and C. Bo, "Visual tracking via weighted local cosine similarity," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1838–1850, Sep. 2015, doi: 10.1109/TCYB.2014.2360924.

[29] Y. Shang, N. Yang, Y. Zhu, and X. Song, "Application of artificial neural network in HRV analysis," in *Proc. IEEE 4th Int. Conf. Comput. Commun. (ICCC)*, Chengdu, China, Dec. 2018, pp. 2138–2141.
[30] H. Costin, C. Rotariu, and A. Pasarica, "Application of artificial neural network in HRV analysis," in *Proc. EPE, Lasi, Romania*, Oct. 2012, pp. 591–596.
[31] K.-J. Park and H. Jeong, "Assessing methods of heart rate variability," *Korean J. Clin. Neurophysiol.*, vol. 16, no. 2, p. 49, 2014.
[32] J. Huang, B. Chen, B. Yao, and W. He, "ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network," *IEEE Access*, vol. 7, pp. 92871–92880, 2019.
[33] X. Wang, T. Ying, and W. Tian, "Spectrum representation based on STFT," in *Proc. 13th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Chengdu, China, Oct. 2020, pp. 435–438.
[34] W. Kabir, M. O. Ahmad, and M. N. S. Swamy, "A multi-biometric system based on feature and score level fusions," *IEEE Access*, vol. 7, pp. 59437–59450, 2019.
[35] A. P. Allen, P. J. Kennedy, S. Dockray, J. F. Cryan, T. G. Dinan, and G. Clarke, "The trier social stress test: Principles and practice," *Neurobiol. Stress*, vol. 6, pp. 113–126, Feb. 2017.

CHEOL HO SONG received the B.S. degree from the Department of Control, Instrumentation and Robot Engineering, Chosun University, Gwangju, South Korea, in 2018, and the M.S. degree from the Department of Electronic Engineering, Chosun University, in 2023, where he is currently pursuing the Ph.D. degree. His current research interests include biometrics, deep learning, and artificial intelligence.

JIN SU KIM received the B.S., M.S., and Ph.D. degrees from the Department of Control, Instrumentation and Robot Engineering, Chosun University, Gwangju, South Korea, in 2016, 2018, and 2022, respectively. He is currently a Postdoctoral Researcher with the Research Institute of IT, Chosun University. His current research interests include biometrics, deep learning, and deepfake detection.

JAE MYUNG KIM received the B.S. and M.S. degrees from the Department of Electronic Engineering, Chosun University, Gwangju, South Korea, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree. His current research interests include biometrics, deep learning, and artificial intelligence.

SUNGBUM PAN (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronics engineering from Sogang University, South Korea, in 1991, 1995, and 1999, respectively. From 1999 to 2005, he was the Team Leader of the Biometric Technology Research Team, ETRI. He is currently a Professor with Chosun University. His current research interests include biometrics, security, and VLSI architectures for real-time image processing.