Emotion Recognition using Short-Term Multi-Physiological Signals

Tae-Koo Kang[1*]
[1] Department of Human Intelligence and Robot Engineering, Sangmyung University
Cheonan, Chungcheongnam-do, Republic of Korea
[e-mail: tkkang@smu.ac.kr]
[*]Corresponding author: Tae-Koo Kang

Received March 2, 2021; revised November 16, 2021; revised January 27, 2022; accepted March 1, 2022; published March 31, 2022

Abstract

Technology for emotion recognition is an essential part of human personality analysis. To define human personality characteristics, the existing method used the survey method. However, there are many cases where communication cannot make without considering emotions. Hence, emotional recognition technology is an essential element for communication but has also been adopted in many other fields.

A person's emotions are revealed in various ways, typically including facial, speech, and biometric responses. Therefore, various methods can recognize emotions, e.g., images, voice signals, and physiological signals. Physiological signals are measured with biological sensors and analyzed to identify emotions. This study employed two sensor types. First, the existing method, the binary arousal-valence method, was subdivided into four levels to classify emotions in more detail. Then, based on the current techniques classified as High/Low, the model was further subdivided into multi-levels. Finally, signal characteristics were extracted using a 1-D Convolution Neural Network (CNN) and classified sixteen feelings. Although CNN was used to learn images in 2D, sensor data in 1D was used as the input in this paper. Finally, the proposed emotional recognition system was evaluated by measuring actual sensors.

Keywords: Convolutional Neural Network (CNN), Emotion Recognition, Physiological Signal

This research was funded by a 2019 research Grant from Sangmyung University.

http://doi.org/10.3837/tiis.2022.03.018 **ISSN : 1976-7277**

1. Introduction

Emotion research has impacted many fields in modern society [1]. Human engineering systems have developed enormously recently, and many recent studies have analyzed human conditions, such as human emotions [2]. Ergonomics is continuously integrated with biomechanics, cognitive engineering, human-computer interface (HCI), emotional engineering, and user experience (UX) to measure and analyze practical research areas [3]. For AI technology to effectively collaborate or respond to humans, recognizing human emotions or states is essential. In particular, biosignal-based state recognition technology is a core technology that can be the basis of remote medical treatment or smart healthcare technology. Therefore, it is necessary to develop a biosignal-based emotion recognition technology.

These areas of physical, cognitive, and emotional interactions between humans and systems increase the importance of user-friendly interface design to improve usability, stability, emotional quality, and system efficiency and provide differentiated use experiences [4]. Emotions can be recognized in various ways, including facial expressions, voice, signals, and brain waves. They can also show in facial expressions and voice, but these can be hidden just by wearing a mask [5-7]. Many recent advances in emotional recognition technology have been using physiological sensors to measure honest feelings [8]. Physiological sensors have also been actively investigated to help recognize emotions by combining technology and physiology. Since the extent to which feelings are expressed or the emotional criteria have various limitations as subjective indicators, different emotional models have been proposed and grouped through models [9].

Physiological sensors to recognize emotions typically include electromyography (EMG), photoplethysmography (PPG), and galvanic skin response (GSR) sensors. These sensors measure autonomic nervous system responses, which control heart muscle, smooth muscle, and external secretion, and are affected by changes in emotions and various body reactions. EMG can measure muscle tension due to muscle activity or stress [10], and PPG can now measure the amount of blood flowing through vessels to determine vessel contraction and heart rate [11]. GSR measures skin temperature and characteristic electrical changes [12]. Electrocardiogram (ECG) measures the heart's contraction and checks the heart rate and inter-beat intervals (IBI). Other sensors associated with the autonomic nervous system can also measure physiological signals.

Current emotional awareness is based on two emotional models [13]

- the six basic feelings of happiness, sadness, surprise, fear, anger, and disgust [14].
- categories based on the second arousal-valence axis [15].

Rather than classifying six emotions, I used the second method of recognizing emotions as arousal and valence values to express various feelings. Previous studies have used a two-stage classification model based on the arousal-valence axis. However, this has the disadvantage of classifying emotions into only four categories. On the other hand, most recent studies that categorize emotions using sensors combine EEG and other sensors. Deep learning methods have been recently applied to physiological processing signals, such as EEG or voice, achieving comparable results with conventional methods [24-26]. Martinez et al. were the first to propose CNNs to establish physiological models for emotion, with many subsequent deep emotion recognition studies [27]. However, EEG requires an algorithm that is too complex to

analyze brainwave signals, and the subject needs 10-20 sensors to be attached to provide reliable brain wave data. Hence, data collection is difficult to apply to everyday life, even if it can be well classified.

Therefore, this paper proposed a more granular arousal-valence 4 step model that will allow sixteen emotional categories and recognize new feelings by combining segmented ones. I used EMG and PPG sensors to measure psychologically relevant signals and provide data for emotion classification. I propose a two-dimensional (2D) classification system using a 2D emotional model separated by arousal-valence and a measurable model for instant emotional changes by using less than 1 s to extract characteristics. Emotional characteristics for each emotion category were extracted using a one-dimensional (1D) convolutional neural network (CNN) [16]. Consequently, I could verify the proposed four stages arousal-valence system better than previous studies. The remainder of this paper is organized as follows. Section 2 reviews Russel's emotion model and introduces the necessity of redefinition by subdividing the emotion model. Section 3 describes our emotion recognition model based on the CNN network. Section 4 presents our experimental results and detailed discussions. Section 5 summarizes our conclusions.

2. Definition of Extended Emotion Model Based on Russel's Emotion Model

This section firstly introduces Russel's emotion model. Then I describe the necessity of redefinition by the proposed subdivision model.

2.1 Russel's Emotion Model Analysis

Before recognizing emotion, I have to define an emotion model to enable emotion classification. An emotion is an action caused by own or another person feeling about something or a situation. Human emotions include subjective criteria; hence it is challenging to express emotions objectively or ultimately numerically. Various models have been developed to provide quantitative figures for feelings, such as the emotion variation detection or dimension approaches. As shown in **Fig. 1**, estimating emotional variation involves graphical representation and numerical verification of the six basic emotions (joy, disgust, anger, sadness, surprise, and fear). It can identify mood changes and incorporate new emotions from the basic emotions.

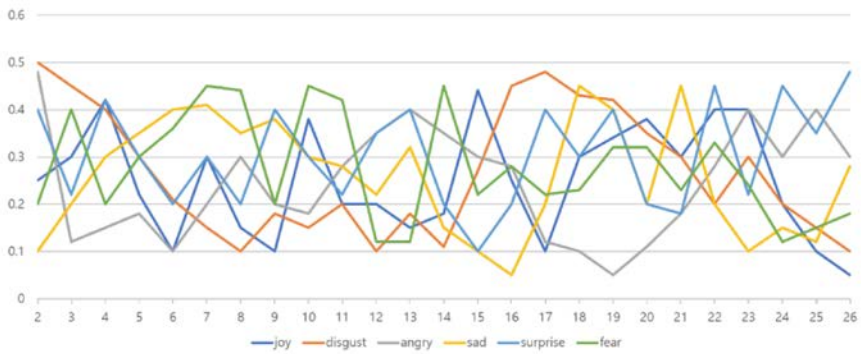

Fig. 1. Examples of physiological signal patterns for various emotions over time

The dimensional approach is a typical example of the Russell emotional model. Russell proposed a two-dimensional emotional model that divided emotions along the arousal and valence axes to express emotions quantitatively. Arousal indicates the level of emotional excitement, where smaller values represent more relaxed or more boring, whereas higher values indicate more excitement or anger. Valence indicates the positive or negative emotional level. For example, fear would be a very negative valence concept, whereas boredom would be a positive one. The Russell model allows emotions to be expressed in a two-dimensional plane [17]. However, I need a more precise criterion than "high" and "low" for physiological emotion signals to identify the various emotions precisely using the numerical method. Therefore, I propose a disaggregated emotional model based on emotional dimensionality. The following section describes the proposed redefined emotion model to exactly segment various emotions from physiological emotion signals.

2.2 Definition of Extended Emotion Model for Segmentation of Physiological Emotion Signals

This section proposes a disaggregated emotional model based on the arousal-valence approaches described in Section 2.1. Russell models can be expressed in the arousal and valence variables, divided into four quadrants. However, the Russell model is somewhat simplistic, and hence I disaggregate the model into multi-levels. To investigate the most efficient number of sub-classes, I conduct the simulation for emotion classification of various emotion signals, including EMG and PPG signals using fuzzy C-Means clustering.

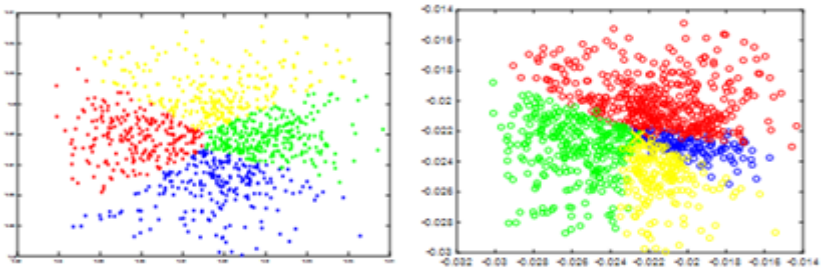

Fig. 2. Example results of emotion classification for physiological signals using fuzzy C- means clustering (left: 4 class, right: 5 class)

Fig. 2 shows example results of emotion classification for physiological signals. In **Fig. 2**, the left image illustrates the classification result in the case of dividing physiological signals into the four classes, and the correct image represents the result for 5-class clustering **Table 1** shows the classification accuracy to determine the appropriate number of levels in 2 to 6 steps using the fuzzy C-means algorithm. Sensor values were analyzed using characteristics under classification by the fuzzy C-means algorithm. In **Table 1**, classification accuracy declined sharply for the five-level model, whereas the four-level model achieves 80% and can express more complex emotions than the standard Russell model.

Table 1. Analysis of multi-class for emotion signals using Fuzzy C-Means

The number of class	2	3	4	5	6
Accuracy	90.8	88	84.2	50.8	36.1
Precision	0.77	0.56	0.23	0.37	0.31

From the results of **Table 1**, I can know that the appropriate number of levels is four and employed a four-level emotional model rather than the usual two-level arousal-valence model to recognize the emotion status precisely. The proposed emotion model allows more granular emotion expression than the conventional two-level model. Current models also struggle to express the level of emotion defined as an adjective. The proposed model can classify emotions into sixteen domains rather than four domains and describe the steps for a given emotion, as shown in **Fig. 3**.

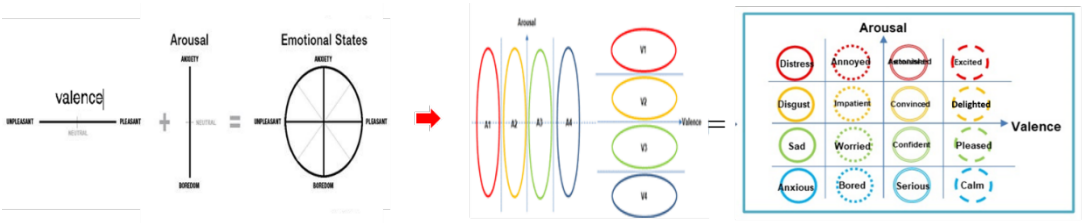

Fig. 3. The proposed emotion model for physiological signals by four-level division

3. Emotion Recognition with Multi-Physiological Signals

This section explains how to generate CNN input for learning and the normalization required to create it. Typical CNN inputs comprise two-dimensional images, whereas I propose that the one-dimensional signals are segmented into their underlying periodicity for input. The CNN model can be freely designed depending on the application. Detailed descriptions for these modules are introduced as follows.

3.1 Structure of Proposed Emotion Recognition System

Fig. 4 shows the structure of the proposed CNN network for emotion recognition. As shown in **Fig. 4**, our proposed system primarily consists of two parts: single pulse segmentation and personal normalization and classification modules. The first step is the functional division part of making single pulse data for physiological emotion data and functional normalization to reduce errors that cause biased signals. The classification part extracts the emotion features for multi-physiological signals, EMG and PPG. Then it recognizes the emotion status using these extracted features. I use the sixteen-emotion model defined in the previous section to classify emotions in the classification part.

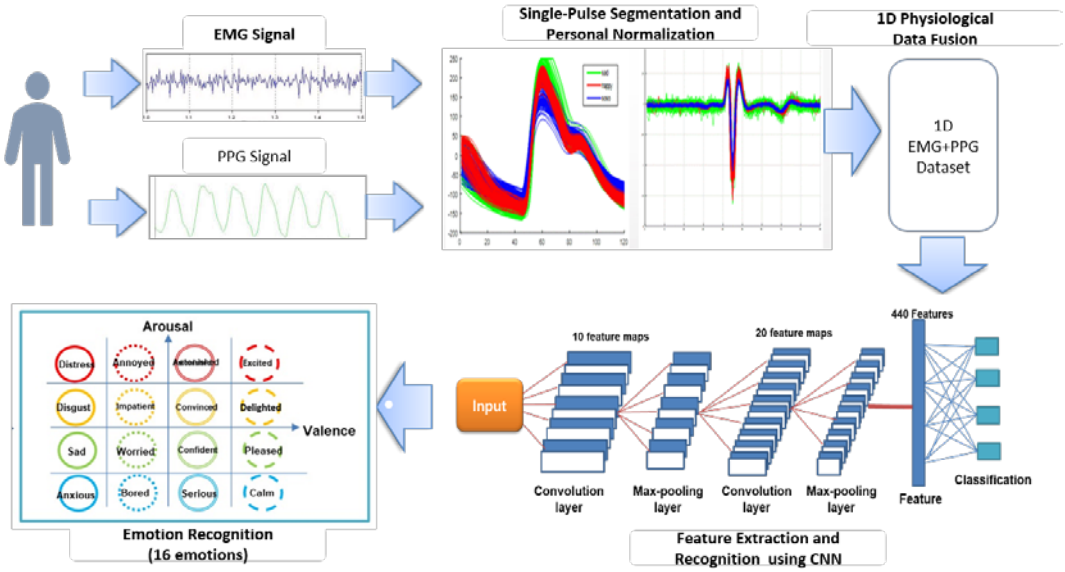

Fig. 4. Proposed emotion recognition system using multi-modal Physiological Signals

3.2. Single-Pulse Segmentation and Personal Normalization

Before training the physiological emotion data, I first have to make the physiological signal with a single-cycle length to extract the feature using CNN. Then I manufacture these data into uniform data using personal normalization. Details descriptions are presented as follows.

3.2.1 Single Pulse Segmentation using Peak Point Values

Most CNN inputs comprise images with constant width and length. Physiological signals have a constant cycle, which can be segmented following set criteria and used as inputs, as shown in **Fig. 5**. As shown in **Fig. 5**, EMG signals were segmented based on the low peak since there were two high peaks, and PPG signals were segmented based on the high peak. I truncated the 1D sensor data to pulse length, with both sensor data being the same length to generate training input data with the same length. As a result, I create the input physiological data for EMG and PPG. These one-pulse data of EMG and PPG for sixteen-emotion are used as the input data of CNN to extract features.

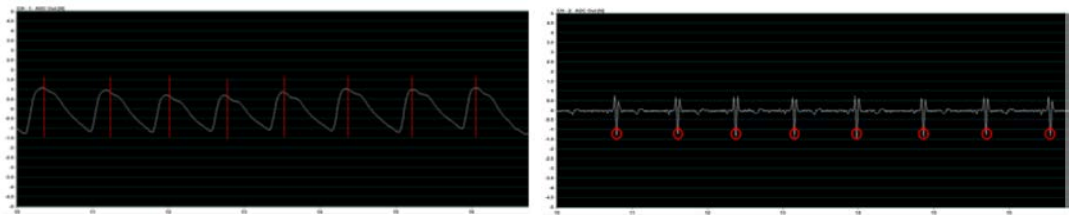

Fig. 5. Sensor data periodicity (left: PPG signal, right: EMG signal)

3.2.2 Personal Normalization for Segmented Single-Cycle Physiological Data

Normalization is usually required to work with data from multiple subjects rather than just one person [18] and can be achieved in several ways. Most of the data are normalized based on the mean, median, lowest, and highest values, as shown in **Fig. 6**. In **Fig. 6**, **Fig. 6(a)** illustrates a plot that does not normalized, and **Fig. 6(b)** represents the data distribution that signals do not match. **Fig. 6(c)** displays the normalized data using the appropriate criteria. The distribution is more consistent, and your CNN and CNN training can proceed correctly.

Since the minimum and maximum values of the sensor differ from person to person, the average value is different. Therefore, emotional training cannot proceed accurately without normalization. Therefore, terms of the EMG values, which are prone to many oscillations, except for the low and high peak values, were normalized based on the average by (1),

$$\widetilde{d_i} := \frac{d_i - E(d)}{\sigma_d} \tag{1}$$

$$\mu_d = \frac{1}{N}\sum_{i=1}^{N} d_i \tag{2}$$

$$\sigma_d = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i - {\mu_d}^2)} \tag{3}$$

where, $\widetilde{d_i}$ presents the normalized EMG data, μ_d and σ_d means average and standard variation among one-pulse EMG data respectively.

Although PPG data is relatively stable and smooth for a given participant, signal amplitude varies significantly between participants. Therefore, in terms of PPG data, normalization was performed using the lowest and highest values using (4),

$$\widetilde{d_i} = \frac{d_i - d_{min}}{d_{max} - d_{min}} \tag{4}$$

where, where, $\widetilde{d_i}$ presents the normalized PPG data, d_{max} and d_{min} means minimum and maximum value among one-pulse PPG data respectively.

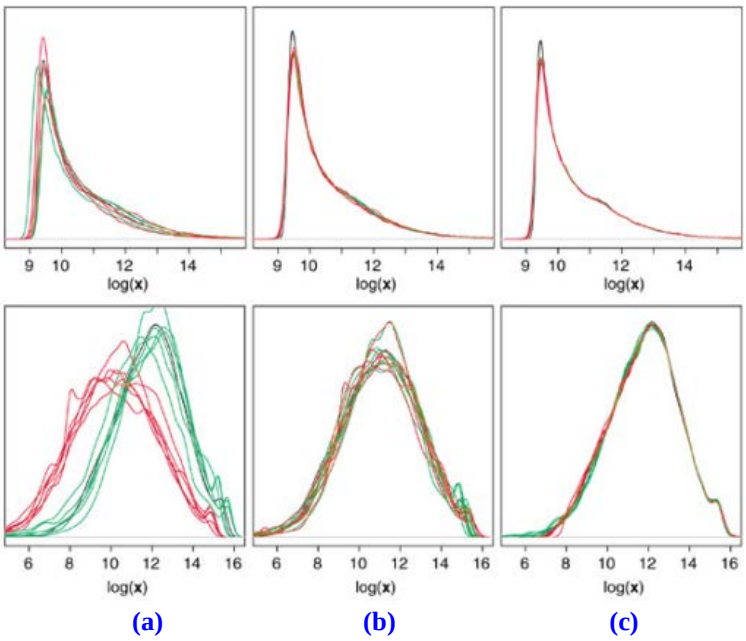

Fig. 6. Examples of physiological emotion signal normalization: (a) without normalization, (b) incorrect normalization, (c) correct normalization

3.4 Emotion Feature Extraction and Classification using Convolution Neural Network

As mentioned above, the CNN model can be freely designed depending on the application. The proposed emotion recognition model uses CNN to extract features representing emotional characteristics. However, feature map aspects learned in the upper and lower layers differ greatly. Thus, the CNN should be designed according to inputs to distinguish the lower and upper layers. Previous studies have highlighted that at least two CNN layers are required to extract physiological sensor characteristics reliably. **Fig. 7** shows the structure of the CNN network for physiological emotion data.

In **Fig. 7**, important CNN parameters include the number of convolution filters, size, and stride size. The larger number of filters will enable more diverse characteristics to be learned. Therefore, I employed ten filters for the first convolution layer and 20 for the second. The filter size was 1x3 for all layers, and the stride was set to 1. Relu layers were used for non-linearity with two subsequent max-pooling layers. Convolution and max-pooling filter sizes were set to the optimum value from training, assessed by comparing performance according to filter size, as shown in **Table 2 and 3** details regarding the proposed CNN model, respectively.

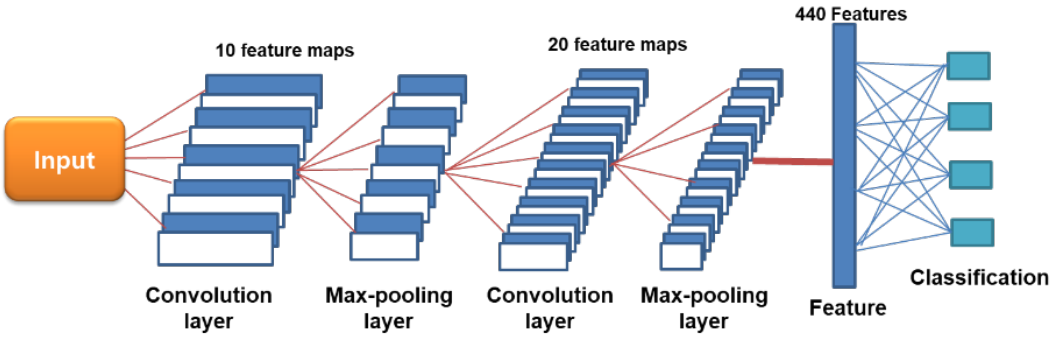

Fig. 7. Structure of CNN network for physiological emotion feature extraction and recognition

Table 2. Performance Comparison by Parameter

Layer	Parameter / Performance			
Convolution	Size	1x3	1x4	1x5
	Accuracy	88.2	83.6	80.7
MaxPooling	Size	1x2	1x3	1x4
	Accuracy	89.2	77.83	70.06

The designed CNN model was then trained using the training dataset. It is critical to avoid overwriting when training the model. Dropout and data augmentation techniques were employed to prevent overfitting. I used a dropout layer learned through neural networks, which has been reduced by omitting some neurons in the input or hidden layers. I set max epochs, initial learning rate, mini-batch size to 100, 0.001, 64, respectively, and used softmax as the activation function for training. **Table 3** presents the principal parameters used in our CNN network.

Table 3. Parameters in CNN Model

Layer	Conv. 1	Maxp. 1	Conv. 2	Maxp. 2	F.C.
Size	1x3	1x2	1x3	1x2	600
Stride	1	2	1	2	

I used the training data to train the CNN network and employed the validation dataset to determine model performance. The trained model performance was then tested using the test dataset. The 1D sensor data inputs were created as discussed in sections 3.2.1 and 3.2.2 for model inputs. Since the optimal parameters for training are unknown, i performed a preliminary test and selected appropriate values depending on the data and application. Once training is completed and the feature map is created, classify it through a fully connected layer.

4. Experimental Result and Analysis

4.1 Experimental Environments

4.1.1 Experimental System Configuration

The hardware platform comprised a Windows 10 operating system with an Intel Core i7-4770K processor running at 3.70 GHz with 8GB RAM. Sensor data was measured directly. Reference datasets for emotional studies are measured mainly by non-Korean researchers in Europe and the US, which is inconsistent with Korean sentiments. Data sets used in this paper were measured for Koreans, in a restricted environment, minimum of 24 hours prohibited the consumption of drugs that could affect the central nervous system, including cigarettes, coffee, and alcohol. Therefore, the dataset will allow emotion recognition appropriate to Korean emotions. **Fig. 8** shows sensor signals and the experimental environment employed to acquire sensor measurements.

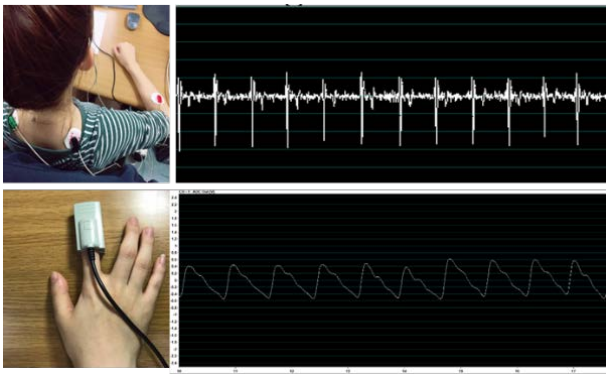

Fig. 8. The appearance of the sensor signal from the experimental system

I randomly divided the collected dataset into training, test, and validation datasets, with 112,000, 24,000, and 11,200 sensor data in each set, respectively. The validation dataset was employed to ensure training was moving in the right direction. **Fig. 8** shows typical dataset examples.

I used a P400 [19] physiologic recorder to record physiological signals. It can measure various signals, including bioelectrical and physiological signals. Up to six measurement modules can be measured simultaneously from four channels. Moreover, sensor measurements can be checked in real-time and recorded or analyzed by sending them to the PC. I used the base module to connect various physiological sensor outputs, including amps and the bio-Amp to measure blood flow were used to measure physiological signals such as PPG, ECG, electrocardiogram, and safety. **Table 4** shows the P400 base module characteristics and specifications, and **Fig. 9** shows the whole system. Additional sensors can be connected to the base module to measure other biological signals as required.

Table 4. Experimental System Specification

Category		Specification
Input Siganl	channels	4
	Input voltage range	Physiologic module: 2.5V external input signal: 5V
Output signal	Sampling rate	Maximum 2000 SPS
	ADC resolution	12 bit
Communication	method (speed)	USB 1.1 (12Mbps)
	Power supply	Input: 80–240VAC; Output: 12V DC
	Voltage / Current	12V / 2A

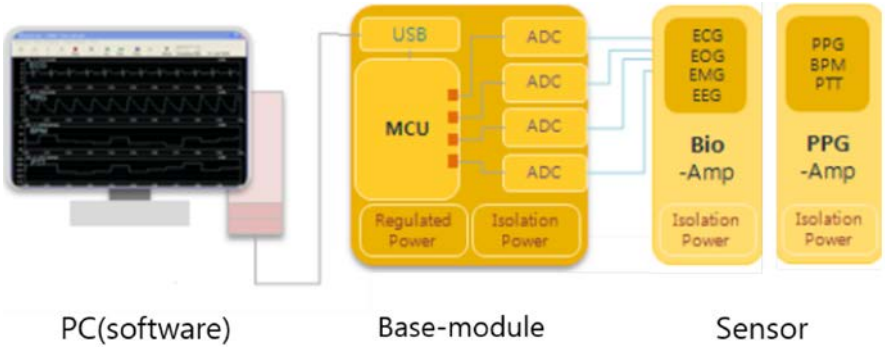

Fig. 9. Block diagram of the experimental system

4.1.2 Stimuli Selection for Experimental Dataset Construction

Generally, emotion classification requires a stimulus to trigger emotions. Previous studies have selected various stimuli, including photos, videos, and music. The DEAP study used photographs as stimuli to measure emotions [20].

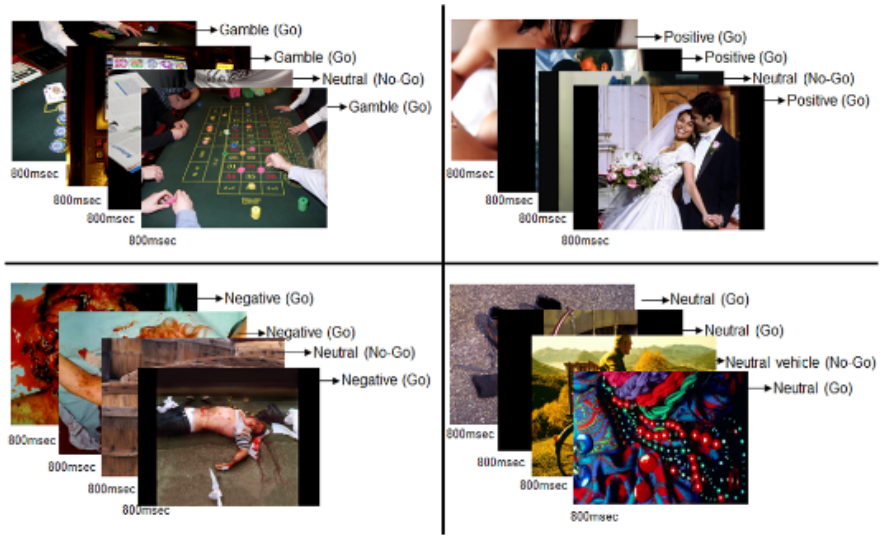

Fig. 10. International Affective Picture System (IAPS)

Fig. 10 shows a typical stimulus image used in the emotional recognition trial of the international affective picture system (IAPS). Videos were used as stimuli in DECAF studies [21]. Music and photography can evoke emotions through sight and hearing, respectively [22]. Creating emotions from images without sound or sounds without images is not as efficient as using both simultaneously [23].

I selected video to stimulate hearing and seeing as the stimulus for this trial. The video was selected subject to a survey on the same age group as the final participants (20-the 30s). A total of thirty people were surveyed and participated in an experiment. I used 80 image sequences for sixteen emotions on four-level arousal and valence axes. After viewing a 5-10 min video, participants selected the level of arousal and valence (both had four possible stages). As a result, I chose a total of 32 videos based on the highest scoring images for each area and edited them to 3-5 min duration. Physiological signals were measured for the final selected video.

4.1.3 Physiological Dataset Construction

Most physiological signal studies use a recognized dataset. However, the sensor type and measurement method vary depending on the study purpose(s). Researchers generally use the authorized dataset suitable for research to verify the performance. However, in the case of our study, there is no dataset suitable for sixteen-emotion conditions using EMG and PPG physiological signal data. Therefore, I measured sensors directly for each participant and created the required datasets by ourselves. Sensor measurements were based on the video selected in Section 4.1.1. Individual data sets were constructed and tested using the conventional 4-level arousal-valence model, and then all datasets were collated for the final analysis. **Table 5** shows the number of datasets created.

Table 5. Emotion dataset configuration using physiological emotion signals

	# of Training Data	# of Test Data
Arousal 1 / Valence 1	42000	9000
Arousal 2 / Valence 2	42000	9000
Arousal 3 / Valence 3	42000	9000
Arousal 4 / Valence 4	42000	9000

4.2 Experimental Results

This section describes experimental results obtained using the previously described dataset. The CNN as a feature extractor provided superior outcomes than hand-crafted feature sets. Conventional arousal-valence models categorize four emotions from the two-level arousal-Valence model, whereas the proposed approach offers sixteen emotions more detail. Despite the increased number of emotions, accuracy also improved using the proposed method.

4.2.1 Accuracy for Emotion Recognition by Participants

This section describes the outcomes using the dataset generated in this study. I experimented with assessing the accuracy of arousal and valence data selection. Once, thirty participants

watched the image sequence and voted the emotion for that image sequence. If the voted results matched the pre-labeled emotion, I regard that image sequence is correct. **Fig. 11** shows arousal and valence accuracy for data selection using EMG data by thirty persons, respectively. In **Fig. 11**, individual accuracy ranged from 92-98%, with 92% mean accuracy.

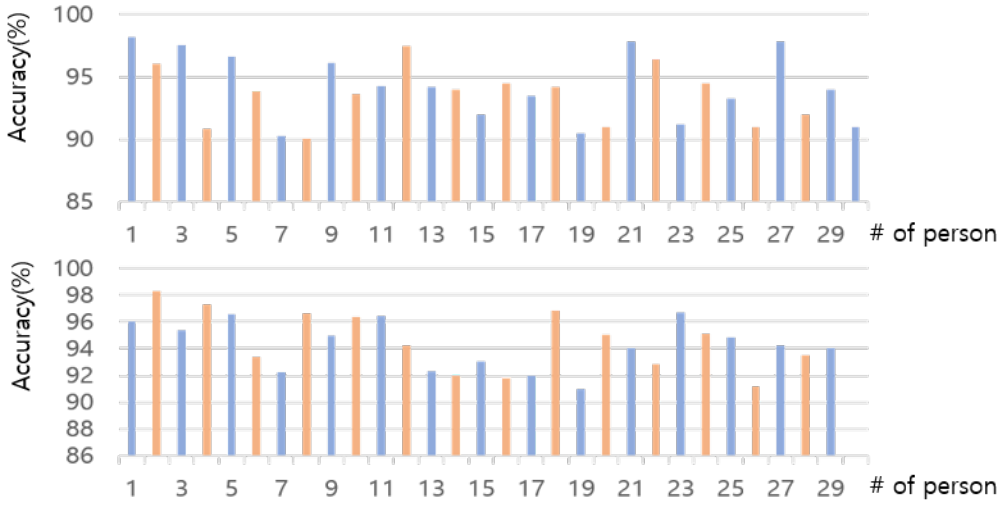

Fig. 11. Experimental results of data selection accuracy of EMG data by Participants

Fig. 12 also shows arousal and valence accuracy for data selection using PPG data. Similar to EMG, individual accuracy ranged from 90-94%, with an average of 92%. Overall accuracy for arousal and valence were 83.5% and 83.1%, respectively, with 1.43% and 1.23% precision, respectively.

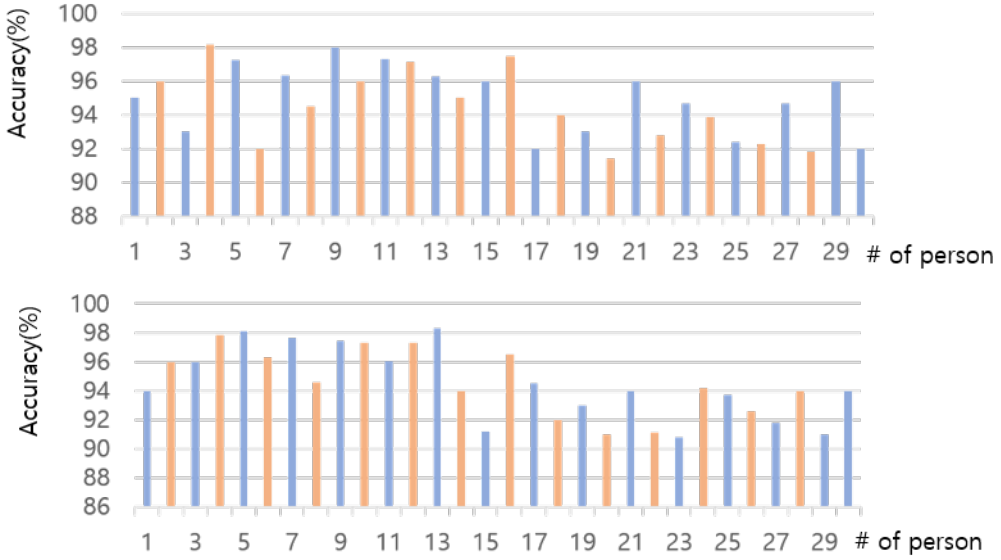

Fig. 12. Experimental results of data selection accuracy of PPG data by Participants

Table 6 shows classification results incorporating all data using the proposed CNN-based method, and **Tables 7 and 8** present a confusion matrix for classification accuracy for arousal and valence levels.

In **Table 6**, these experimental results represent average classification accuracy obtained for four-level arousal and valence over 100 trials. As shown in **Fig. 6**, EMG data classification accuracy for four-level arousal and valence is 89.2% and 88%, respectively. In addition, that of PPG data presents 87% and 86.5%, respectively.

Table 6. Accuracy results of emotion recognition using the proposed method

Data Type	Arousal Accuracy (%)	Valence Accuracy (%)
EMG	89.2	88
PPG	87	86.5
EMG+PPG (proposed)	89.25	88.9

In the case of the proposed method, the fusion of EMG and PPG data shows 89.25% for arousal accuracy and 88.9% for valence accuracy, which means that confusion of multi-physiological signals is efficient without conflict comparing each that of signal type. Comparing each Arousal and Valence four-levels is also shown in **Table 7 and Table 8**.

Table 7. Confusion matrix for four-level arousal by the proposed method

Accuracy	Arousal 1	Arousal 2	Arousal 3	Arousal 4
Arousal 1	91.6	6.6	0.8	1.0
Arousal 2	7.7	88.3	2.8	1.2
Arousal 3	1.8	2.3	87.5	8.4
Arousal 4	0.9	1.3	7.5	90.25

As shown in **Table 7** four-level arousal model showed superior recognition performance in terms of diversity and accuracy than two-level arousal since arousal 1, and 2 show similar feelings in the existing two-level emotion model. Moreover, Arousal 3 and 4 also have a higher probability of losing each other for the same reason as arousal 1 and 2.

Table 8. Confusion matrix for four-level valence by the proposed method

Accuracy	valence 1	valence 2	valence 3	valence 4
valence 1	87.3	8.1	2.1	2.5
valence 2	6.6	89.8	1.7	1.9
valence 3	1.4	2.3	90.1	6.2
valence 4	1.1	2.6	7.4	88.9

As shown in **Table 7** four-level arousal model showed superior recognition performance in terms of diversity and accuracy than two-level arousal since arousal 1, and 2 show similar feelings in the existing two-level emotion model. Moreover, Arousal 3 and 4 also have a higher probability of losing each other for the same reason as arousal 1 and 2. **Table 9** shows the experimental emotion recognition results for sixteen emotions using the EMG and PPG fusion datasets.

Table 9. Experimental results of emotion recognition for sixteen emotions by the proposed method

Emotions	**Distress**	**Disgust**	**Sad**	**Anxious**	**Annoyed**	**Impatient**	**Worried**	**Bored**
Accuracy	**92**	**87**	**84**	**90**	**87**	**86**	**82**	**86**
Emotions	**Astonished**	**Convinced**	**Confident**	**Serious**	**Excited**	**Delighted**	**Pleased**	**Calm**
Accuracy	**88**	**85**	**86**	**87**	**91**	**85**	**88**	**93**

As shown in **Fig. 9**, our proposed method offers an average of 87% recognition accuracy for sixteen emotions though the physiological emotion data is one-pulse. From the results, I can know that the proposed approach can be applied to real-world situations by the person because the only one-pulse physiological signal to recognize the emotions.

4.2.2 Accuracy comparison of emotion recognition with existing methods

In this section, I proposed a CNN-based method to extract features. However, previous studies have also considered artificial neural networks (ANNs) and support vector machines (SVM) using hand-crafted features [24]. Sensors were measured and used directly, similar to the current approach. **Table 10** details the proposed algorithms and compares outcomes with the current approach.

Table 10. Characteristics of comparison algorithms: (a)ANN-based method, (b)SVM-based method

Method	ANN algorithm	Proposed algorithm
Similarities	Based on Russell's emotional model	
	Measure physiological sensor directly	
Differences	Two-level arousal/valence two class	Four level arousal/valence
	EEG, ECG, PPG	EMG, PPG
	Stimuli: Picture	Stimuli: video
	ANN	CNN

(a)

Method	SVM algorithm	Proposed algorithm
Similarities	Using EMG and PPG sensor	
	Measure physiological sensor directly	
	Stimuli: Video	
Differences	Four basic emotion model	Four-level arousal/valence Emotion model
	EOG, ECG, PPG	EMG, PPG
	Hand-crafted feature and SVM	CNN

(b)

Because of the characteristics of the SVM algorithm, I carry out the experiment for recognition accuracy for emotion data. Moreover, I compare the recognition performance with the SVM-

based method. For the SVM-based reduction category experiment, 100 segments were extracted from each of the subjects considered highly emotional during the video viewing and used as experimental data. The results of recognizing emotions using 100 bio-sensor data are shown in **Table 11**.

Table 11. Experimental results for four representative emotions: (a) SVM-based method, (b) the proposed method

Emotion	Happy	Joy	Fear	Sad
Happy	79/100	5/100	15/100	1/100
Joy	8/100	75/100	4/100	13/100
Fear	12/100	2/100	82/100	4/100
Sad	0/100	18/100	6/100	76/100

(a)

Emotion	Happy	Joy	Fear	Sad
Happy	89/100	5/100	10/100	1/100
Joy	10/100	84/100	6/100	11/100
Fear	7/100	2/100	86/100	4/100
Sad	0/100	7/100	5/100	87/100

(b)

As shown in **Table 11**, experimental results have shown that the proposed system using characteristic information extracted from a bio-sensor has identified emotions with a probability of 87.5%. Whereas the existing method averagely presents 78% accuracy. Therefore, our approach using CNN, a deep learning algorithm, shows an improvement of more than 9.5% over the results of comparison algorithms using hand-made features and SVM algorithms. These results stem from the robust feature extraction by the CNN network and personal normalization for one-pulse physiological emotion data.

As seen in the experimental results from **Table 6 to Table 10**, the proposed method showed more accurate recognition performance for more emotions than the existing methods, which can be analyzed as synergy in accuracy because of the emotional data for PPG and the emotional data for EMG complement each other. In addition, the fact that the EEG and PPG individual recognition test results also showed better emotion recognition accuracy performance than the existing methods indicates that the proposed method is also superior to the existing method.

5. Conclusion

This paper proposed an emotion recognition system based on directly measured EMG and PPG bio-sensor data characterized and classified using a 1D CNN. The proposed algorithm redefined the conventional two-level Russel's model to the four-level models, providing significantly more detailed emotion separation. A total of 32 image sequences were selected and measured considering the genre and bits of images used as stimuli to classify the following emotions accurately. The proposed system identified emotions with 87% average accuracy for the four-level arousal-valence model, whereas classification using SVM and hand-crafted features produced 78% accuracy for the two-level e arousal-valence model. Increasing the number of stages would reduce accuracy, but the proposed model provided comparable

performance. Thus, the proposed model can accurately identify more complex emotions. Algorithms using physiological sensors to categorize emotions or user states have traditionally been used for medical purposes only. However, the proposed method is expected to be utilized in various fields of mobile-based healthcare as it can quickly and easily identify emotional or health status through the biosensor attached to the device in a smartphone or mobile device. The fields to which the specifically proposed technology can be applied are as follows.

- The user's emotion recognition technology can be used in robot artificial intelligence, home IoT, interactive home shopping, virtual reality, and architectural interior fields, and has various application fields such as computer interface and biosignal-based HCI/HRI.
- The PPG technology of this study can be applied to a driving assistance system by measuring the fatigue level of a car driver, a correction parameter for patterning when measuring an EMG, and a health care system such as fatigue or stress measurement.
- EMG-based technology can be used as a control system for drones or robots and as a steering system for cars for the disabled.
- The user's body, state recognition technology, can be used for unmanned material transport technology, construction automation equipment control technology, building defect detection and repair technology, driver driving environment information provision technology, and car door lock technology.

Moreover, human emotions can be expressed not as one at a time but as a combination of several psychological states. However, it can be said that this study is meaningful in recognizing representative and primary emotions among them. As a future study, I plan to study a method for identifying emotions based on more diverse biosignals. In addition, I plan to study a methodology to analyze a person's personality or characteristics by measuring the change in emotion based on the recognized emotion.

References

[1] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias; W. Fellenz, and J.G. Taylor, "Emotion recognition in human-computer interaction," *IEEE signal processing magazine*, vol. 18, no. 1, pp. 32-80, 2001. Article (CrossRef Link)

[2] E. Hudlicka, "To feel or not to feel: The role of act in human-computer interaction," *International journal of human-computer studies*, vol. 59, no. 1-2, pp. 1-32, 2003. Article (CrossRef Link)

[3] S. Kim, Y. Kim, and T. Lee, "Rendering of general paralyzed patient's emotion by using EEG," in *Proc. of the Korean Institute of Electrical Engineers (KIEE) summer conference*, pp. 343-344, 2007. Article (CrossRef Link)

[4] M. Kim, Y. Joo, and J. Park, "Development of facial image based emotion recognition system," in *Proc. of Korean Institute of Intelligent Systems (KFIS) spring conference*, vol. 15, no. 1, pp. 433-436, 2005. Article (CrossRef Link)

[5] M. Schroder, "Emotional speech synthesis: A review," in *Proc. of Seventh european conference on speech communication and technology*, 2001. Article (CrossRef Link)

[6] C. Anagnostopoulos, T. Iliou, and I. Giannoukos, "Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011," *Artificial intelligence review*, vol. 43, no. 2, pp. 155-177, 2015. Article (CrossRef Link)

[7] W. Kim, "Emotional Speaker Recognition using Emotional Adaptation," *The transactions of the Korean Institute of Electrical Engineers*, vol. 66, pp.1105-1110, 2017. Article (CrossRef Link)

[8] Y. Zhong, M. Zhao, W. Yongxiong, Y. Jingdong, and Z Jianhua, "Recognition of emotions using multi-modal physiological signals and an ensemble deep learning model," *Computer methods and programs in biomedicine*, vol. 140, pp. 93-110, 2017. Article (CrossRef Link)
[9] C. Hieda, T. HoriiI, and T. Nagai, "Emotion Differentiation based on Decision-Making in Emotion Model," in *Proc. of 27th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pp. 659-665, 2018. Article (CrossRef Link)
[10] U. Lundberg, R. Kadefors, B. Melin, G. Palmerud, P. Hassm_en, M. Engstrom, and I. E. Dohns, "Psychophysiological stress and EMG activity of the trapezius muscle," *International journal of behavioral medicine*, vol. 1, no. 4, pp. 354-370, 1994. Article (CrossRef Link)
[11] Y. Lee, O. Kwon, H. Shin, J. Jo, and Y. Lee, "Noise reduction of PPG signals using a particle filter for robust emotion recognition," in *Proc. of 2011 IEEE International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, 2011. Article (CrossRef Link)
[12] G. Wu, G. Liu, and M. Hao, "The analysis of emotion recognition from GSR based on PSO," in *Proc. of International Symposium on Intelligence Information Processing and Trusted Computing (IPTC)*, pp. 360-363, 2010. Article (CrossRef Link)
[13] F. Zhou, S. Kong, C. Fowlkes, T. Chen, and B. Lei, "Fine-Grained Facial Expression Analysis Using Dimensional Emotion Model," *Neurocomputing*, vol. 392, pp 38-49, 2020. Article (CrossRef Link)
[14] S. An, L. Ji, M. Marks, and Z. Zhang, "Two Sides of Emotion: Exploring Positivity and Negativity in Six Basic Emotions across Cultures" *Frontiers in Psychology*, vol. 8, no. 610, pp. 1-14, 2017. Article (CrossRef Link)
[15] S. Peng, L. Zhang, Y. Ban, M. Fang, and S. Winkler, "A Deep Network for Arousal-Valence Emotion Prediction with Acoustic-Visual Cues," *arXiv preprint arXiv: 1805.00638*, 2018. Article (CrossRef Link)
[16] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541-551, 1989. Article (CrossRef Link)
[17] T. Colibazzi, J. Posner, Z. Wang, D. Gorman, A. Gerber, S. Yu, H. Zhu, A. Kangarlu, Y. Duan, and J. A. Russell, "Neural systems subserving valence and arousal during the experience of induced emotions," *Emotion*, vol. 10, no. 3, pp. 377-289, 2010. Article (CrossRef Link)
[18] T. Barszcz, M. Bielecka, A. Bielecki, and M. Wojcik, "Wind turbines states classification by a fuzzy-ART neural network with a stereographic projection as a signal normalization," in *Proc. of International Conference on Adaptive and Natural Computing Algorithms*, pp. 225-234, 2011. Article (CrossRef Link)
[19] Physio Lab Co., Ltd.. [Online]. Available: www.physiolab.co.kr
[20] S. Koelstra, C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A Database for Emotion Analysis Using Physiological Signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18-31, 2012. Article (CrossRef Link)
[21] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition," in *Proc. of International conference on machine learning*, vol. 32, no.1, pp. 647-655, 2014. Article (CrossRef Link)
[22] J. Christensen, *Sound and the aesthetics of play*, 2018, pp. 39-65. [Online] Available: https://link.springer.com/book/10.1007/978-3-319-66899-4
[23] H. Wang and S. Huang, "Musical rhythms affect heartrate variability: algorithm and models," *Advances in Electrical Engineering*, vol. 2014, pp. 1-14, 2014. Article (CrossRef Link)
[24] G. Yoo, S. Seo, S. Hong, and H. Kim, "Emotion extraction based on multi bio-signal using back-propagation neural network," *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4925-4937, 2018. Article (CrossRef Link)
[25] Mustaqeem and S. Kwon, "CLSTM: deep feature-based speech emotion recognition using the hierarchical ConvLSTM network," *Mathematics*, vol. 8, no. 12, pp. 1–19, 2020. Article (CrossRef Link)

[26] S. Liu, S. Wang, X. Liu, A. H. Gandomi, M. Daneshmand, K. Muhammad, and V. C. Albuquerque, "Human memory update strategy: a multi-layer template update mechanism for remote visual monitoring," *IEEE Transactions on Multimedia*, vol. 23, pp. 2188-2198, 2021.
Article (CrossRef Link)
[27] Mustaqeem and S. Kwon, "Att-Net: enhanced emotion recognition system using lightweight self-attention module," *Applied Soft Computing*, vol. 102, no.4, pp. 1–11, 2021.
Article (CrossRef Link)

Tae-Koo Kang received his B.S. in Applied Electrical Engineering, M.S. in visual image processing, and Ph.D. in Electrical Engineering from Korea University, Seoul, Republic of Korea, in 2001, 2004, and 2012 respectively. He was a research professor at Korea University, Seoul, the Republic of Korea from 2012 to 2014 and an assistant professor in Information and Telecommunication Engineering, Cheonan, the Republic of Korea in 2015 and 2016. He is now an Assistant Professor in the Department of Human Intelligence and Robot Engineering, Sangmyung University, Cheonan, Republic of Korea. His research interests include computer vision, robotics, artificial intelligence, and machine learning.