2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Real-Time Affective Computing Platform Integrated with AI System-on-Chip Design and Multimodal Signal Processing System

Wei-Chih Li, Cheng-Jie Yang, Bo-Ting Liu and Wai-Chi Fang
Department of Electronics Engineering & Institute of Electronics
National Yang Ming Chiao Tung University
Hsinchu 30010, Taiwan R.O.C.
Pervasive Artificial Intelligence Research (PAIR) Labs, Taiwan
Corresponding Author: Professor Wai-Chi Fang (dr.wfang@gmail.com)

Abstract— **Recently, deep learning algorithms have been used widely in emotion recognition applications. However, it is difficult to detect human emotions in real-time due to constraints imposed by computing power and convergence latency. This paper proposes a real-time affective computing platform that integrates an AI System-on-Chip (SoC) design and multimodal signal processing systems composed of electroencephalogram (EEG), electrocardiogram (ECG), and photoplethysmogram (PPG) signals. To extract the emotional features of the EEG, ECG, and PPG signals, we used a short-time Fourier transform (STFT) for the EEG signal and direct extraction using the raw signals for the ECG and PPG signals. The long-term recurrent convolution networks (LRCN) classifier was implemented in an AI SoC design and divided emotions into three classes: happy, angry, and sad. The proposed LRCN classifier reached an average accuracy of 77.41% for cross-subject validation. The platform consists of wearable physiological sensors and multimodal signal processors integrated with the LRCN SoC design. The area of the core and total power consumption of the LRCN chip was 1.13 x 1.14 mm^2 and 48.24 mW, respectively. The on-chip training processing time and real-time classification processing time are 5.5 μs and 1.9 μs per sample. The proposed platform displays the classification results of emotion calculation on the graphical user interface (GUI) every one second for real-time emotion monitoring.**

Clinical relevance— **The on-chip training processing time and real-time emotion classification processing time are 5.5 μs and 1.9 μs per sample with EEG, ECG, and PPG signal based on the LRCN model.**

I. INTRODUCTION

Applications involving the learning of changes in human emotion have been a popular subject for many researchers in interdisciplinary fields such as computer science, psychology, engineering, and cognitive science [1]. The main sources of obtaining data related to emotional states involve completing questionnaires, intuitive changes in facial expressions, and physiological signals such as the electrocardiogram (ECG), electroencephalogram (EEG), and photoplethysmogram (PPG) signals [2]. Compared to the methods mentioned, physiological signals are the most objective and true responses [2]. Thus, in the past two years research on emotion recognition has mainly been focused on using ECG and EEG signals. From shallow models to deep models [3]–[6], many studies have used machine learning as algorithms to discover the correlation between physiological signals and human emotional responses. In the research of EEG-related emotions, many EEG studies point out that EEG has a high degree of correlation with human emotions [7], [8]. Similarly, PPG signal is also widely used for emotion recognition [9], [10]. Handouzi et al. [11] developed an automatic emotion recognition system that used only one physiological signal called "blood volume pulse" (BVP). Among healthy people, this system can identify short-term emotions with 95% accuracy.

However, these studies all face the same problem: The generalization of a model trained by a physiological signal from only one subject is not quite suitable for other subjects due to the unique nature and responses of each emotional-cognitive system in each subject. This paper presents a real-time affective computing platform integrated with real-time emotion recognition and multi-modal physiological signals systems based on a subject-independent artificial intelligence (AI) algorithm. By applying real-time emotion calculation methods and real-time physiological signals display, subjects can observe their emotional changes and physical health information.

II. SYSTEM ARCHITECTURE

The system architecture of the proposed affective computing platform is shown in Fig. 1. It consists of three wearable front-end sensors: EEG, ECG, and PPG sensors, a signal processing system, an AI system-on-chip (SoC) platform, and graphical user interfaces on a laptop. The EEG, ECG, and PPG front-end sensors send raw data to the signal processing system through Bluetooth slaves, which are connected to the Bluetooth master of the signal processing system. Within the signal processing system, signal pre-processing and feature extraction were performed to suit the emotion classifier input. After the processor sent the formatted input, the Field-Programmable-Gate-Array (FPGA) controller in the AI SoC platform sent emotion features to the emotion classifier implemented in the Long-term Recurrent Convolutional Networks (LRCN) chip and received emotion

This research is partially supported by the Ministry of Science and Technology through Pervasive Artificial Intelligence Research (PAIR) Labs, Taiwan. Wei-Chih Li, Cheng-Jie Yang, Bo-Ting Liu, and Wai-Chi Fang are with the Department of Electronics Engineering and Institute of Electronics, National Yang Ming Chiao Tung University, Taiwan, and PAIR Labs, Taiwan.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

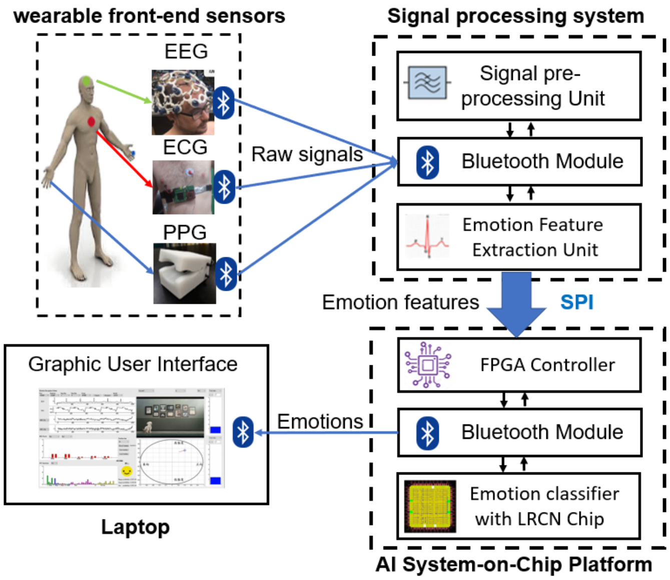


Fig. 1. The integrated affective computing platform with an AI system-on-chip design and a multimodal signal processing system

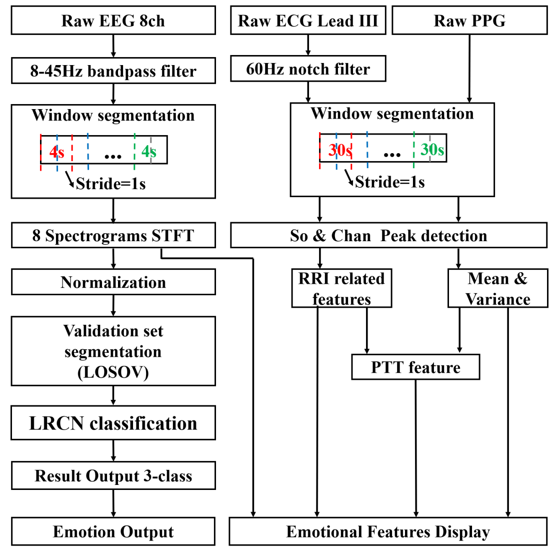


Fig. 2. Signals processing flow of the entire affective computing platform

classification results from the chip. Finally, the FPGA controller in the AI SoC platform sent the emotion classification result to the graphical user interface on the laptop via a Bluetooth point-to-point connection.

III. Emotion Recognition Algorithm

A. Processing Flow

Fig. 2 describes the processing flow of EEG, ECG, and PPG signals respectively. Considering the computation power limitation of multiple signal processing, we chose an 8-channel EEG signal, a 3-lead ECG signal, and the fingertip PPG signal. According to the study of Yang et al. [12], EEG channels such as Fp1, Fp2, F3, F4, F7, F8, T3, and T4 are recommended selections for the EEG-based emotion recognition algorithm. Each EEG sample was 24 bits for a total of 192 bits sample size, collected at a rate of 250 samples per second. An 8-45 Hz bandpass filter was deployed to extract the most emotion-related frequency parts that should improve the classification. The spectrograms were extracted by hardware-friendly Short-Time Fourier Transform (STFT) on each 4-second data sequence with a 1-second stride. The baseline normalization was then applied to eliminate the background differences between each subject. Here we proposed the LRCN classifier to replace the traditional CNN model which incorporated emotion changes over time into analysis without changing other procedures. Finally, the 3-class emotional output would be produced by LRCN with leave-one-subject-out validation (LOSOV).

The ECG Lead III was measured with an ADS1298 sensor with a gain of 12 for 250 samples per second and a sample size of 24 bits. A 60 Hz notch filter was deployed to pre-process the ECG data. However, the PPG data were directly windowed due to the stable sources. To synchronize the data segmentation of the ECG and PPG signals with the EEG signals, both ECG and PPG data were segmented by a 30 seconds window with the same 1-second stride as EEG data. The So and Chan algorithm [13] was employed to real-time and accurately locate the peaks of the ECG and PPG signals. The R-R intervals (RRI) features and pulse transmission time (PTT) features were then calculated. The mean and variance of PPG features were also derived. Eventually, the spectrograms of EEG and all the derived features of ECG and PPG signals could be applied to the real-time physiological monitoring display.

B. LRCN Training Model

The LRCN architecture for real-time emotion classification is shown in Fig. 3. The Long Short-Term Memory (LSTM) layers were appended behind three convolution-pooling layers to extract time-sequence features of the EEG signal. The many-to-one architecture was applied due to the inherent time-dependent characteristic of the LSTM layers The feature map inputs of 10 seconds were split into 10 timesteps. which consisted of 38 Hz x 8 channels. In every second, one timestep would go through three convolution-pooling layers for the training process until 10 timesteps have all been processed. The LSTM layer then sends the output to the fully connected layer and Softmax layer for emotion classification output calculated with 10 time-dependence data.

IV. System Hardware Implementation

The proposed affective computing platform was developed and integrated as shown Fig. 4. This platform consists of three wearable front-end sensors, a RISC-V processor for multiple signal processing, and an AI SoC platform integrated with an FPGA controller and the implemented LRCN chip.

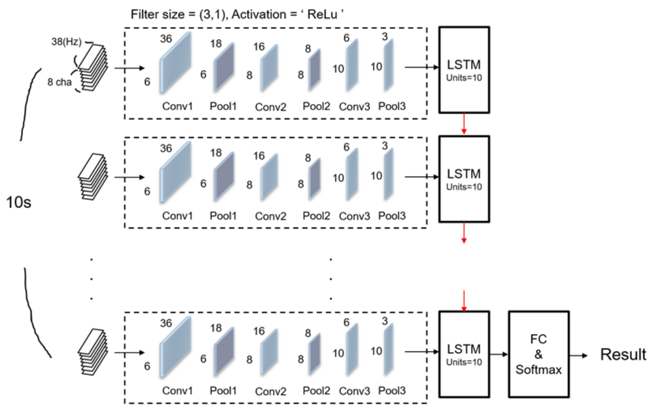


Fig. 3. The LRCN architecture for real-time emotion classification

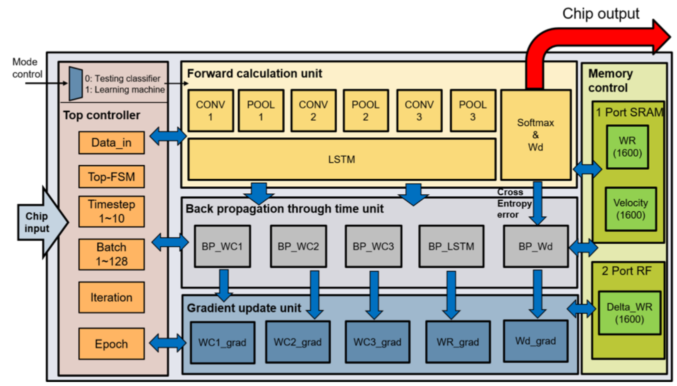


Fig. 5. Chip function block of the LRCN structure

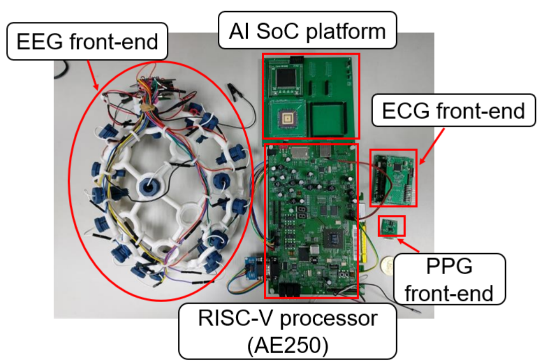


Fig. 4. The hardware design of the affective computing platform integrated with an AI SoC system and wearable front-end sensors

A. Wearable Sensors Development

The wearable sensors of the three signals are integrated with the Bluetooth connection device of the signal processing system. To facilitate wearing, we use a dry electrode brainwave cap as the front-end circuit design for collecting EEG signals. The signals of the three wearable sensors are integrated with the Bluetooth device connected to the signal processing system. For comfortable wearing, we used a dry electrode brainwave cap as the front-end circuit design for collecting EEG signals. The front-end circuit of ECG is designed as a miniaturized device that is portable to connect to the measurement electrode patch signal. In the PPG front-end circuit, the MAX30102 sensor is used to collect the PPG signal from the fingertips. All front-end circuits will then transmit the signal to the signal processing system via Bluetooth for pre-processing and emotional feature extraction.

B. Signal Processing Flow of the RISC-V Processor

After receiving the signal transmitted by the front-end circuit, the RISC-V processor performs the pre-processing and emotional feature extraction of all signals. To send emotion-related features into the emotion classifier for emotion recognition and calculation in real-time, the RISC-V processor will give priority to the pre-processing algorithm of the EEG signal as shown on the left of Fig. 2. After the EEG emotion characteristics are calculated, the processor will send the emotion characteristics to the emotion recognition chip through the FPGA controller to perform the calculation of the LRCN algorithm. ECG and PPG features are also processed by the RISC-V processor at this time and then sent to the FPGA controller. After the FPGA controller completes the emotion recognition on the chip, it transmits the recognition results together with the emotional characteristics of all physiological signals to the graphical user interface of the laptop via Bluetooth.

C. AI System-on-Chip Platform integrated with the LRCN Chip Design

For real-time emotion classification, we implemented the proposed LRCN model into a 16nm technology chip. The EEG signal processing in the Spartan6 FPGA was first converted into a 38*8 matrix with time, frequency, and initial weight information and sent into the LRCN chip for affective computing acceleration (the function block of the chip is shown in Fig. 5).

The Top control center is responsible for arranging and transmitting the 24-bit parallel input data to each computing, controlling, and operating unit of the system, and also compiling the output results after the arrangement.

The Forward calculation unit includes three layers of convolution, three layers of average pooling, one layer of LSTM, one layer of full connection, and one layer of Softmax operation. The main function of the Forward calculation unit is to perform convolution and recursion with the weights after training. The structure would then calculate, complete the classification effect, and retain the calculation results for each layer in the process. Consequently, the calculated results would be the output of Back-propagation through the time unit for weight update calculation.

In the Back-propagation through the time unit and the Gradient update unit, the classification errors were first calculated by the Forward calculation unit and then applied with Back-propagation through the time (BPTP) method to calculate the error contribution of each time-step of each layer. The action of updating weights stopped until the calculation error converged.

V. System Evaluation and Performance

A. Dataset Description

The physiological data used for the algorithm evaluation was recorded by the professional clinical psychologists in Kaohsiung Medical University (KMU) [14]. In the KMU study, sixty subjects were asked to induce four emotions (happy, angry, sad, and neutral) through the recalling of the subject's personal experiences. Before the formal collection of physiological data, the researcher would first ask the participants to recall their emotional events. The main purpose is to clarify that the participants recalled a single emotional feeling (such as neutral, angry, happy, sad) to the corresponding event, including the people, times, places, and things at that moment. If the statement of the participant contains complex emotions (for example, sad and happy, love-hate), the researcher would guide them to avoid inducing complex emotions in the following formal data measurement. For each emotional state, signals were collected for 15 minutes: 5 minutes for the statement period (pre-training phase), 5 minutes for the recall period (mind-training phase), and 5 minutes for the recovery period (post-training phase).

B. LRCN Algorithm Validation

The LRCN classifier of the affective computing system was conducted through leave-one-subject out validation (LOSOV). Due to the time limitation of the real-time training process, we selected 40 subjects from the KMU dataset for validation. In LOSOV validation, the data of each subject is treated as the validation set only and the data of the other subjects are the training set. After the validation process of all subjects is completed, the mean and standard deviation of overall accuracies were considered as the validation performance. The validation results are shown in Fig. 6. The mean and the standard deviation of the accuracies are 77.41% and 15.14% respectively for classifying 3 emotional states: happiness, anger, and sadness. In the KMU dataset, there was no certainty that neutral emotions would not be mixed with other emotions during the experiment. To ensure the integrity of the emotion classification model, we decided to remove neutral emotion classification.

C. Comparison with other works

The comparison with other state-of-the-art works is shown in Table.I. In our work, we achieve the mean accuracy of 77.41% for 3-class classification on the KMU dataset. Kwon et al. [15] achieve the mean accuracy of 76.56% for binary classification with CNN as the classifier. [16] selects the multi-phase CNN as the classifier which achieves the mean accuracy of 83.36% for binary classification. However, the subject-dependent method was adopted rather than the rigorous cross-subject (subject-independent) validation. Song et al. [4] proposes the DGCNN which uses a graph to model the multi-channel EEG features. It achieves a mean accuracy of 79.95%. The proposed LRCN model achieved better classification and multi-class recognition rather than only binary classes to further apply to more fields. To develop the applicable affective computing system, fewer EEG channels and fewer neural network layers should be considered which was also realized in our work.

Subject	Accuracy (Happy-Angry-Sad)	Subject	Accuracy (Happy-Angry-Sad)	Subject	Accuracy (Happy-Angry-Sad)
Sub1	48.69%	Sub15	60.32%	Sub29	84.49%
Sub2	78.05%	Sub16	77.14%	Sub30	78.75%
Sub3	89.52%	Sub17	93.33%	Sub31	69.88%
Sub4	91.43%	Sub18	84.76%	Sub32	98.52%
Sub5	79.4%	Sub19	100%	Sub33	70.44%
Sub6	56.18%	Sub20	76.19%	Sub34	93.33%
Sub7	54.86%	Sub21	73.52%	Sub35	72.3%
Sub8	93.48%	Sub22	63.81%	Sub36	83.81%
Sub9	56.33%	Sub23	48.27%	Sub37	72.38%
Sub10	56.19%	Sub24	53.85%	Sub38	90.54%
Sub11	100%	Sub25	93.33%	Sub39	65.71%
Sub12	68.57%	Sub26	88.57%	Sub40	72.38%
Sub13	81.9%	Sub27	95.24%	Mean	77.41%
Sub14	92.5%	Sub28	88.57%	Std	15.14%

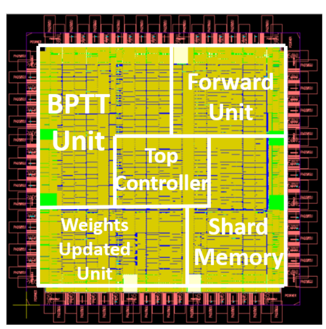
Fig. 6. The emotion classification accuracy of the proposed LRCN structure for subject-independent validation

TABLE I
Comparison of our approach with some latest works

	[15]	[16]	[4]	**This work**
Signals	EEG	EEG	EEG	**EEG**
No. classes	2	2	2	**3**
No. subjects	32	32	15	**52**
Classifier	CNN	CNN	DGCNN	**LRCN**
Subject Dependency	independent	dependent	independent	**independent**
Accuracy	76.56%	83.36%	79.95%	**77.41%**

D. System-on-Chip Performance

The proposed LRCN chip layout and specification are shown in Fig. 7. This chip has been implemented using TSMC 16nm Fin-FET technology. The core size is 1.13 x 1.14 mm^2 and the on-chip SRAM is 221.3 Kbits. The total power consumption of the chip is 48.24 mW. The chip has training mode and testing mode and the on-chip training processing time and testing processing time are 5.5 μs and 1.9 μs per sample respectively.



Technology Process	TSMC 16nm
Operating Freq.	125MHz
Core Size	1.13 x 1.14 mm2
Core Voltage	0.8 V
On-Chip SRAM	221.3 Kbits
Power Consumption	48.24 mw
Training FPS	5.5 us
Inference FPS	1.9 us

Fig. 7. Chip layout of innovative LRCN structure and specification table

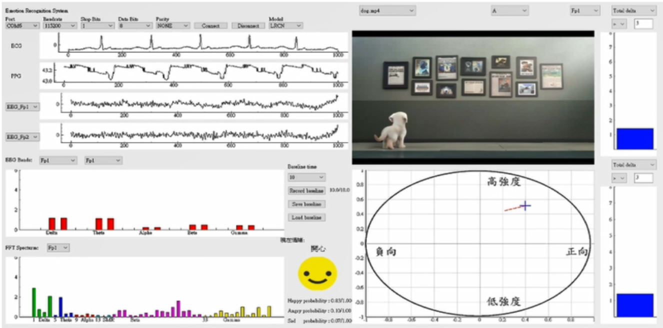


Fig. 8. A real-time displayed graphic user interface for emotion tracking and physical status monitoring

E. Real-time Emotion Monitoring with the Graphic User Interface

The GUI platform of the laptop displayed the measured raw data, the pre-processed data from the RISC-V processor, and the 3-class emotion outputs given by the LRCN chip. For real-time monitoring of human emotion variation, the graphic user interface is designated as shown in Fig. 8. The Python GUI runs on the end-user laptop and displays 4 of the 8 EEG channels. The probabilities of three emotions except neutral emotions are calculated with the output result of the LRCN chip. The final emotion is determined with calculated probabilities and displayed through an emotional smiley icon. For EEG analysis, the spectrum intensity of each EEG band and the difference between two selected EEG channels are both displayed for reference. As for emotion recognition, we also applied the valence-arousal circumplex model for real-time emotion tracking. The blue cross symbol location displays the current emotion recognition result. Additionally, the red line shows the emotion recognition result from the previous to the present second.

VI. Conclusion

This study implemented a real-time affective computing platform with an LRCN system-on-chip design and multiple physiological signal processing systems, including wearable EEG, ECG, and PPG sensors. The emotional features of the ECG and PPG signals were extracted directly, while the EEG emotional features were extracted by STFT for meaningful data. The LRCN algorithm was used as the emotion classifier due to the time-dependent characteristic, which improves the emotion classification accuracy of the previous research by using a CNN classifier. The classification accuracy of the proposed LRCN model was evaluated by LOSOV with a mean accuracy of 77.41%, improving the previous research by using CNN classifier. The LRCN chip was implemented using TSMC 16nm Fin-FET technology with the core area of 1.13 x 1.14 mm^2 and the total power consumption of 48.24 mW. The training speed of the chip reaches 5.5 μs per sample, which allows the overall system to display emotion classification results in 1 second. In conclusion, the proposed platform is suitable for emotional and physical monitoring applications that require real-time emotional state monitoring.

VII. Acknowledgement

This research is partially supported by the Ministry of Science and Technology under Grant Number MOST 108-2634- F-037-001- through Pervasive Artificial Intelligence Research (PAIR) Labs, Taiwan.

References

[1] P. J. Bota, C. Wang, A. L. N. Fred and H. Plácido Da Silva, "A Review, current challenges, and future possibilities on emotion recognition using machine learning and physiological signals", IEEE Access, vol. 7, pp. 140990-141020, Sep. 2019.

[2] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao and X. Xu, "A review of emotion recognition using physiological signals", Sensors, vol. 18, pp. 2074, 2018.

[3] S. M. S. A. Abdullah, S. Y. A. Ameen, M. A. Sadeeq and S. Zeebaree, "Multimodal emotion recognition using deep learning", Journal of Applied Science and Technology Trends, vol. 2, no. 2, pp. 52-58, Apr. 2021.

[4] T. Song, W. Zheng,P. Song and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks", IEEE Trans. on Affective Computing, vol. 11, no. 3, pp. 532-541, Mar. 2018.

[5] J. Zhang, Z. Yin, P. Chen and S. Nichele, "Emotion recognition using multi-modal data and machine learning techniques: a tutorial and review", Inf. Fusion, vol. 59, pp. 103-126, Jul. 2020.

[6] A. Patane and M. Kwiatkowska, "Calibrating the classifier: siamese neural network architecture for end-to-end arousal recognition from ECG", In Int. Conf. on Machine Learning, Optimization, and Data Science, pp. 1-13, Feb. 2019.

[7] Y. Yin, X. Zheng, B. Hu, Y. Zhang and X. Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM", Applied Soft Computing, vol. 100, Mar. 2021.

[8] J. X. Chen, P. W. Zhang, Z. J. Mao, Y. F. Huang, D. M. Jiang and Y. N. Zhang, "Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks", IEEE Access, vol. 7, pp. 44317-44328, Mar. 2019.

[9] M. S. Lee, Y. K. Lee, D. S. Pae, M. T. Lim, D. W. Kim and T. K. Kang, "Fast emotion recognition based on single pulse PPG signal with convolutional neural network", Appl. Sci., vol. 9, Aug. 2019.

[10] M. Lee, Y. K. Lee, M. T. Lim and T. K. Kang, "Emotion recognition using convolutional neural network with selected statistical photoplethysmogram features", Appl. Sci., vol. 10, May. 2020.

[11] W. Handouzi, C. Maaoui, A. Pruski and A. Moussaoui, "Objective model assessment for short-term anxiety recognition from blood volume pulse signal", Biomedical Signal Processing and Control, vol. 14, pp. 217-227, Nov. 2014.

[12] C. J. Yang, N. Fahier, C. Y. He, W. C. Li and W. C. Fang, "An AI-edge platform with multimodal wearable physiological signals monitoring sensors for affective computing applications", In 2020 IEEE Int. Symp. on Circuits and Systems (ISCAS), pp. 1-5, Oct. 2020.

[13] H. H. So and K. L. Chan, "Development of QRS detection method for real-time ambulatory cardiac monitor", Proc. 19th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., vol. 1, pp. 289-292, Oct. 1997.

[14] I. M. Lin, S. Y. Wang, S. Y. Fan, E. Peper, S. P. Chen and C. Y. Huang, "A single session of heart rate variability biofeedback produced greater increases in heart rate variability than autogenic training", Applied Psychophysiology and Biofeedback, vol. 45, no. 4, pp. 343-350, Aug. 2020.

[15] Y. H. Kwon, S. B. Shin, and S. D. Kim, "Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system", Sensors, vol. 18, no. 5, Apr. 2018.

[16] W. C. Fang, K. Y. Wang, N. Fahier, Y. L. Ho and Y. D. Huang, "Development and validation of an EEG-based real-time emotion recognition system using edge AI computing platform with convolutional neural network system-on-chip design", IEEE J. on Emerging and Selected Topics in Circuits and Systems, vol. 9, no. 4, pp. 645-657, Dec. 2019.