CNN Based Emotion Regression Using EEG Signals

Firoozeh Shahrokni
MA, Department of Biomedical Engineering, Faculty of Engineering, Central Tehran Branch, Islamic Azad University ,Tehran, Iran.
Bentolhoda Ayati*
Assistant Professor, Department of Biomedical Engineering, Faculty of Engineering, Central Tehran Branch, Islamic Azad University ,Tehran, Iran.

Abstract- Emotions are considered to be a characteristic of humans and play an important role in their everyday interactions. A distinctive trait that separates humans from computers is their unique emotions. Emotion recognition of humans by computers offers a variety of applications in the area of human-computer interaction and brain-computer interface. Regarding the importance of this area, this study investigates two-dimensional (2D) convolutional neural network (CNN) based emotion recognition by using electroencephalogram (EEG) signals. In doing so, the signals are firstly transformed into 2D images through the use of continuous wavelet transforms (CWT) and three different CNN models are then compared for emotion regression. The study proposed method is evaluated by using the DEAP database and via the EEG signals recorded from 32 subjects who were watching 40 music video clips with a length of 60s. Findings from the evaluation of the three models indicate that the accuracy of 97/90%, was achieved by the first proposed model which has the best performance. It is also observed that the performance of this model has considerably improved as compared with the results of other studies using the same method.

Keywords- EEG signals . Emotion regression . Convolutional neural networks . Scalogram.

Introduction

A type of human response to life events is emotions which can influence different whole-body activities [1-2]. Correct recognition of emotions is essential as it plays an important role in human non-verbal communication and daily living [3-4]. In the past few decades, a great number of psychological studies regarding the analysis and identification of emotions were performed. Individuals experience feelings in response to a stimulus that is generated whenever they take a part in an activity. If the stimulus is considered a positive stimulus by an individual, it can be triggered to positive emotions. Regardless of the type of emotions, all can manifest themselves as mental physiology, facial expression, gesture, or biological reactions. There exists a significant number of papers with the purpose of proposing an automatic emotion recognition system and a few have addressed non-physiological signals, including speech [5], facial expressions [6], and body posture [7]. In fact, these manners are resulting from mental processes and depend upon age and culture. Thus, it is hard to believe that the methods for the classification of emotions based on such manners are accurate. In several other studies, other physiological signals, such as heartbeat [8], skin impedance [9], EEG brain signals, and functional magnetic resonance imaging [fMRI], have been addressed and improvements in the recognition of emotions have been documented. The results of such investigations show that emotion recognition through physiological and biological reactions can be measured and recorded using human body sensors. An example to illustrate this method is the use of the brain-computer interface (BCI) technique [8,9] for recording physiological signals generated by the brain and applying them for the detection of emotions. The electroencephalogram (EGG) signal is non-linear and non-stationary and has a considerable amount of artifacts and noise. Regarding this, recognizing emotions from EEG signal features is very challenging. Features are mainly extracted from EEG signals in the domains of time, frequency, or time-frequency. Some of the drawbacks while using these features as input to recognize emotions is a manual search for characteristic features and selecting the most appropriate one from EEG signals, both can manifest themselves as a problem for the recognition of emotions. Besides, researchers have widely focused on the classification of emotions but only a few have attempted to perform accurately the regression of each emotion in a user. Motivated by the mentioned drawbacks, the aim of this research is to propose an approach to solve today's challenges for investigators. In doing so, maps of new features from EEG signals have been produced and a high-quality model for the recognition of users' emotional states with a high level of accuracy has been designed in this study. In recent years, deep learning (DL) [12-13] has achieved numerous successes in various fields, including signal processing, artificial intelligence, and emotion recognition, through the use of new modern systems. The most commonly employed approaches to DL are deep belief network (DBN) [16], convolutional neural network (CNN) [14-15], and recurrent neural network (RNN) [17]. The present study aims to use the CNN approach to automatically obtain the most appropriate features for emotion regression. Related studies on the recognition of emotions based on EEG signals are reviewed below. The remained of the paper is structured as follows. In Section 2, a multi-channel EEG dataset used for the proposed approach is described. Section 3 suggests deep CNN models. In Section 4, the results achieved by the DL model assessment are analyzed. The discussion and conclusion are provided in the final section of the research.

1.2 Related works

Mico and Chika (2016) reported a convolution neural network-based method for EEG-based emotion recognition. In their study, the PSD+RFs low-depth architecture and a high-depth CNN model were assessed and compared. The results showed that the best segmentation accuracies achieved with low-depth architecture and high-depth architecture were 58.51% and 81.16%, respectively [19].

Lee et al. (2016) reported a method based on a recurrent convolutional neural network for multi-channel EEG-based emotion recognition. In this study, the signals were converted to network-like frames via scalograms and wavelet transform. The acquired data were used as input to the architectures of the SVM, RFs, and R-CNN for performing segmentation [20].

In a study by Elham et al. (2018), data augmentation techniques were utilized to improve the performance of the proposed CNN model. In their study, the 3D presentation of multichannel EEG signals was formalized and used as input to the proposed model.

A 2D-CNN was designed by Yehun Koan et al. (2018) for feature extraction using convolution filters. Before performing the convolution operation, EEG signals were preprocessed through wavelet transform and GSR signals via a zero-crossing rate [22].

Zhong Kegao et al. (2019) used EEG signals' features in the time, frequency, and time-frequency domains to perform emotion segmentation using low-depth models, including support vector machine (SVM), linear discriminant analysis (LDA), and bias linear discriminant analysis (BLDA), and deep learning models, including CNN, GSCNN, and GSLTCNN.

Zhong Min et al. (2019) investigated emotion recognition based on graph convolution neural networks and combined RASM features over five frequency bands to apply the proposed model (P-GCNN and G-CNN+PLV two models) [24].

Zhang et al. (2019) extracted the power spectral characteristics over four frequency bands (theta, alpha, beta, and gamma) and converted the obtained features to cortex-like frames while preserving the spatial information of the electrode locations. Next, a low-depth parallel CNN was utilized for segmentation [25].

Zhong Kegao et al. (2019) reported a GPSO-optimized convolutional neural network for EEG-based emotion recognition. They compared their proposed GSO-optimized CNN model with four prominent emotion recognition methods: PSD+SVM, DE+SVM, HCNN, and DE-DBN [26].

Mehmat Balal et al. (2020) used a spectrogram to convert EEG signals into 2D data and employed the AlexNet and VGG16 networks for emotion recognition [27].

Panayo Kilawat et al. (2021) performed a comparative study on window size and channel arrangement for EEG emotion recognition using Deep-CNN. After applying many changes to CNN parameters, such as window size, they found that CNN parameters significantly influenced emotion recognition operation [28].

Antetopic et al. (2021) exploited 9 characteristics (BP-DE-FDHA-HM-HC-PP-PSD-RMS) in time, frequency, and time-frequency domains for emotion recognition. Then, they proposed a new model by creating feature maps from topographic (TOPO-FM) and holographic (HOLO-FM) representations of EEG signal features [29].

2. Dataset

DEAP dataset [19] has been applied to assess the method proposed in this research. This dataset consists of signals recorded from 32 subjects (16 females and 16 males) who were aged between 19 and 37 years. Along with EEG signals, the dataset contains eight physiological signals including galvanic skin, breathing range, skin temperature, heartbeat, blood pressure, neck muscle activity, muscles involved in smiling, and EOG. The EEG signals were collected from 40 music video clips with a length of 60s, using 32 channels based on 10-20 standards at a sampling rate of 512 Hz. After watching the chosen 40 music video clips, the subjects rated each video in terms of the level of arousal, liking, valence, and dominance from 1-9. In this study, preprocessed multichannel EGG signals of the dataset were used for processing. During preprocessing, the DEAP dataset signals are firstly down-sampled to a 128 Hz sampling rate, and the ocular artifacts are then deleted through the use of independent component analysis (ICA) [19].

3. The proposed method

Figure 1 shows the flowchart of the proposed method for emotion regression. According to Figure 1, the proposed method is made up of two fundamental parts, 1) preprocessing of multi channel EEG signals, and 2) deep CNN with three distinct architectures for emotion regression of different subjects. The descriptions for each part of the model are presented below.

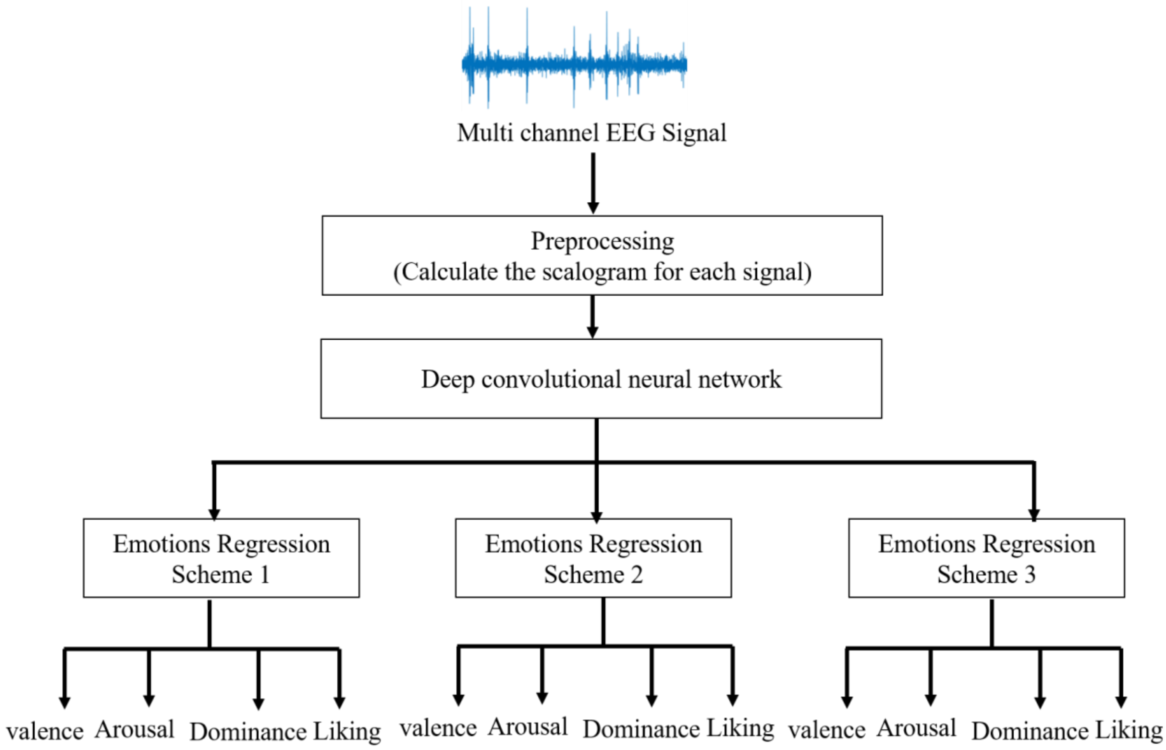

Fig. 1 The flowchart of the proposed method.

3.1. Preprocessing

To classify the time domain signals, one-dimensional CNN models are developed. However, two-dimensional CNN models show better performance as compared with one-dimensional CNN models. Since EGG time domain signals can be transformed into two-dimensional images via various transformers, such as continuous wavelet transforms (CWT), a two-dimensional CNN model has been used in this study to perform regression. In doing so, the 32-channel EEG signals were converted into 2D-scalogram images through the use of CWT. The EEG signals array shape of 32 (Subjects) × 40 (video) ×40 (channel). After performing preprocessing, 1280 scalogram data of the size of 45 × 8064 was obtained and considered as the input.

3.2 Deep CNN

This part is performed using free parameters of the network – i.e. weights and biases. In other words, a set of paired data generated by using the network input and a proper output were applied to the training system. Once the data are used as the input to the network, the output is compared with the proper output and the learning error is calculated. The purpose of using this procedure for regulating network parameters is to produce a more similar output to the proper result if the same input is given to the network. There exist two methods to train the network parameters. In the first method, known as ‘ordinal’ or ‘randomized’ training, a parameter is trained after training a sample. The second method, known as the ‘batch method’, involves updating each parameter once every sample is trained. While the memory requirement is low in using the first method, the consistency is not enough since each training sample can lead one network parameter to the other. Despite the higher memory requirement of the second method, the consistency is higher because changes and parameters are stored. Regarding this, the batch method of training is used in this study. Firstly, the scalogram produced from the preprocessing stage is randomly divided into a training group (which encompasses 80% of the dataset or 1024 train data) and a testing group (which encompasses 20% of the dataset or 256 test data). Then, the three architectures of the two-dimensional CNN are applied for emotion regression. The overall architectures of the three proposed models, consisting of three convolution layers, an activation layer, a normalization layer, and a pooling layer, are shown in Figure 2.

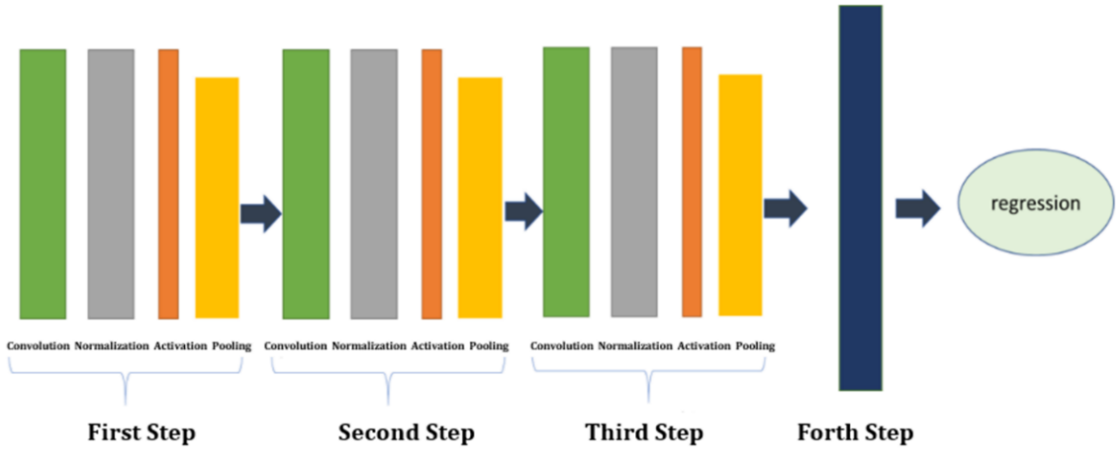

Fig. 2 Schematic illustration of the CNN used in this study.

The properties of the proposed models, as well as their input and output sizes, are presented in Table 1. The input size of 32×32 is considered for the first and second models, and 64×64 for the third architecture. The produced scalograms are firstly resized in accordance with the input size of each model, and then applied to the model to be trained.

Table 1 The properties of each proposed architecture used in this work

	Scheme 1	**Scheme 2**	**Scheme 3**
Input Layer	Scalogram Image (32×32×1)	Scalogram Image (32×32×1)	**Scalogram Image (64×64×1)**
Conv 1	Convolutional Layer (32×32×20)	Convolutional Layer (32×32×8)	**Convolutional Layer (64×64×16)**
Bachnorm1	Batch normalization Layer (32×32×20)	Batch normalization Layer (32×32×8)	**Batch normalization Layer (64×64×16)**
Relu1	Relu Activation Function (32×32×20)	Relu Activation Function (32×32×8)	**Relu Activation Function (64×64×16)**
Pooling 1	Max Pooling Layer, stride=2 (16×16×20)	Average Pooling Layer, stride=1 (31×31×8)	**Average Pooling Layer, stride=1 (63×63×16)**
Conv 2	Convolutional Layer (16×16×40)	Convolutional Layer (31×31×16)	**Convolutional Layer (63×63×32)**
Bachnorm2	Batch normalization Layer (16×16×40)	Batch normalization Layer (31×31×16)	**Batch normalization Layer (63×63×32)**
Relu2	Relu Activation Function (16×16×40)	Relu Activation Function (31×31×16)	**Relu Activation Function (63×63×32)**
Pooling 2	Max Pooling Layer, stride=2 (8×8×40)	Average Pooling Layer, stride=1 (30×30×16)	**Average Pooling Layer, stride=1 (62×62×32)**

Conv 3	Convolutional Layer (8×8×80)	Convolutional Layer (30×30×32)	**Convolutional Layer (62×62×64)**
Bachnorm3	Batch normalization Layer (8×8×80)	Batch normalization Layer (30×30×32)	**Batch normalization Layer (62×62×64)**
Relu3	Relu Activation Function (8×8×80)	Relu Activation Function (30×30×32)	**Relu Activation Function (62×62×64)**
Pooling 3	Max Pooling Layer, stride=2 (4×4×80)	Average Pooling Layer, stride=1 (29×29×32)	**Average Pooling Layer, stride=1 (61×61×64)**
Fully connected	Fully connected Layer 1×1280	Fully connected Layer 1×26912	**Fully connected Layer 1×238144**
Regression Layer	Regression Layer 1×1	Regression Layer 1×1	**Regression Layer 1×1**

3.3. Model assessment

Once the CNN models were designed and implemented based on the training data, the models were assessed through the use of testing data. In doing so, the testing data were given as the input to the trained models, and the output obtained by the models was calculated through an error index via Equation 1.

$$\mathbf{error}\ \boldsymbol{in\ percentage}: \frac{|\mathbf{actual\ value} - \mathbf{estimated\ value}|}{\mathbf{Max}\ (\mathbf{actual\ value})} \times \mathbf{100} \qquad \text{Equation 1}$$

4. Results

Through the use of EEG signals obtained from the DEAP database, a new algorithm for emotion regression is presented in this research. When the dataset preparation was carried out and the proposed model was trained, the model was evaluated by using the training data.

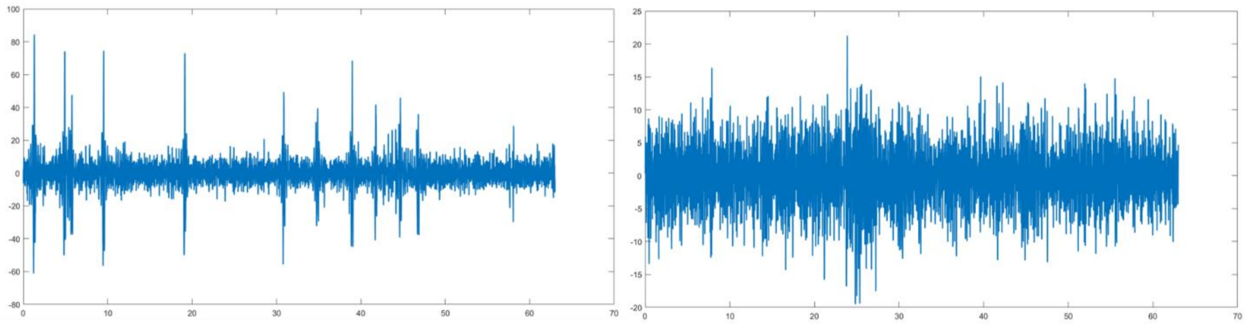

Fig. 2 Example of the signals obtained from the DEAP database.

Using the CWT, the scalogram was then obtained. A summary of the obtained scalogram of the DEAP database signals is presented in Figure 3. It can be seen that the 32-channel EEG signal was transformed into a unified 2D scalogram.

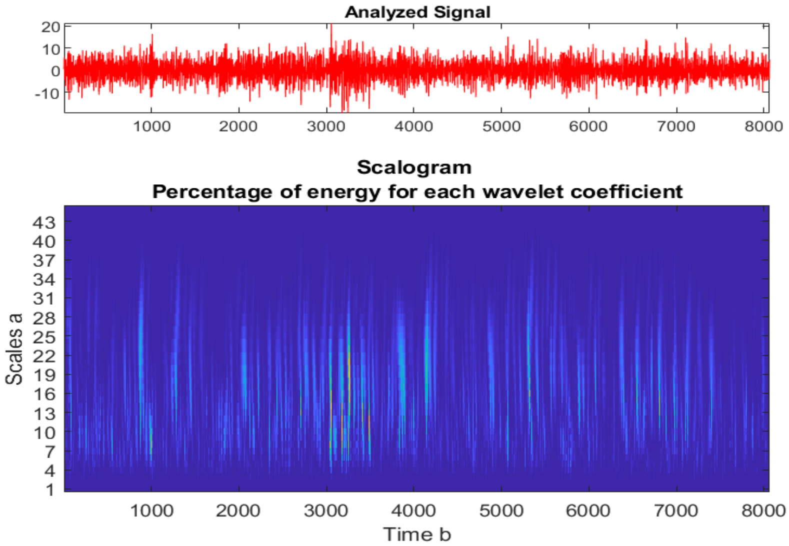

Fig. 3 Example of the extracted scalogram from the EEG signal

The scalogram extraction was performed and emotion regression was carried out based on three models.

Table 2 The results of the first proposed model (as a percentage)

emotion	Accuracy	Error
valence	97.47	2.52
Arousal	96.94	3.05
Dominance	97.68	2.13
Liking	99.33	0.66

Table 3 The results of the second proposed model (as a percentage)

emotion	Accuracy	Error
valence	94.05	5.94
Arousal	96.18	3.81
Dominance	92.65	7.34
Liking	96.24	3.75

Table 4 The results of the third proposed model (as a percentage)

emotion	Accuracy	Error
valence	84.64	15.35
Arousal	87.22	12.77
Dominance	85.80	14.19
Liking	53.53	46.46

According to Tables 2,3, and 4, it can be observed that the first proposed model shows the best performance for the regression of four types of emotions. A comparison between the performance results of each of the proposed models is illustrated in Figure 4.

Fig. 4 Comparison between the models' performance results

5. Discussion and conclusion

There exists a significant difference between convolutional neural networks (CNN) and other types of artificial neural networks since CNNs focus on the input properties than on the problem itself. Regarding this, setting network architecture would become far simpler. Considerable works in the field of deep learning have made available the optimal utilization of deep neural networks (DNN) for researchers and achieved successes and advances in other fields. This research aims to improve the accuracy of emotion regression based on EEG signals. Studies in this area are performed based on different fundamental parts. But, the first part is the preparation of an appropriate and applicable dataset. In doing so, the DEAP database was selected for validating the proposed algorithm of the present study. In the second part, the conversion of the EEG signals to the scalogram images was done by using continuous wavelet transforms (CWT). In the last part of this study, three different CNN architectures were employed to extract features and regression from the four above-mentioned emotions. The findings of the present experiment indicate that the first CNN model can exhibit acceptable results. An accuracy of 97/47%, 96/94%, 97/68%, and 99/33% was achieved for the emotions of valence, arousal, dominance and liking, respectively. And an average accuracy of 97/90% was achieved in this study. Accordingly, the proposed method and the EEG signals showed a proper performance in emotion regression. Table 5 demonstrates a comparison between the results of the present research with other studies carried out in the same area.

Table 5 Comparison between the results of this research with previous studies.

Previous studies	Accuracy (%)
Mico and Chica [20]	81/16
Elham et al. [22]	87/96
Proposed method	**97/90**
Panayokilawat et al. [30]	84/20

As is presented in Table 5, the proposed method has a higher performance than the approaches used in the other studies. Considering the comparison between the resources and the proposed method in Table 5, it can be noted that classification techniques were used in these studies for emotion recognition. According to the present study findings, it is reasonable to mention that regression and score calculation approaches are more effective than classification techniques since each emotion in the DEAP database can be scored from 1 to 9.

Refrence

[1] H. I. Dino and M. B. Abdulrazzaq, "Facial expression classification based on SVM, KNN and MLP classifiers," in 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 70-75.

[2] O. F. Mohammad, M. S. M. Rahim, S. R. M. Zeebaree, and F. Y. Ahmed, "A survey and analysis of the image encryption methods,"International Journal of Applied Engineering Research, vol. 12, pp.13265-13280, 2017.

[3] V. Shrivastava, V. Richhariya, and V. Richhariya, "Puzzling Out Emotions: A Deep-Learning Approach to Multimodal Sentiment Analysis," in 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), 2018, pp. 1-6.

[4] D. A. Zebari, H. Haron, S. R. Zeebaree, and D. Q. Zeebaree, "Enhance the Mammogram Images for Both Segmentation and Feature Extraction Using Wavelet Transform," in 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 100-105.

[5] R.A. Khalil, E. Jones, M.I. Babar, T. Jan, M.H. Zafar, T. Alhussain, Speech emotion recognition using deep learning techniques: A review, IEEE Access 7 (2019) 117327–117345.

[6] A. Alreshidi, M. Ullah, Facial Emotion Recognition Using Hybrid Features, Informatics 6 (2020) 7.

[7] F. Ahmed, A.S.M.H. Bari, M.L. Gavrilova, Emotion Recognition From Body Movement, IEEE Access 8 (2020) 11761–11781.

[8] L. Shu, Y. Yu, W. Chen, H. Hua, Q. Li, J. Jin, X. Xu, Wearable Emotion Recognition Using Heart Rate Data from a Smart Bracelet, Sensors 20 (2020) 718.

[9] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, X. Yang, A Review of Emotion Recognition Using Physiological Signals, Sensors 18 (2018) 2074.

[10] X. Chen, W. Zeng, Y. Shi, J. Deng, Y. Ma, Intrinsic Prior Knowledge Driven CICA fMRI Data Analysis for Emotion Recognition Classification, IEEE Access 7 (2019) 59944–59950.

[11] K.D. Singh, C. Fioravanti, A. Elshahabi, S. Ruiz, R. Sitaram, B. Braun, Involvement of top-down networks in the perception of facial emotions: A magnetoencephalographic investigation, NeuroImage 222 (2020) 117075.

[12] E. Ghaleb, M. Popa, and S. Asteriadis, "Multimodal and Temporal Perception of Audio-visual Cues for Emotion Recognition," in 2019 8th International Conference on Affective Computing and Intelligence Interaction (ACII), 2019, pp. 552-558.

[13] K. B. Obaid, S. Zeebaree, and O. M. Ahmed, "Deep Learning Models Based on Image Classification: A Review," International Journal of Science and Business, vol. 4, pp. 75-81, 2020.
[14] T. D. Nguyen, "Multimodal emotion recognition using deep learning techniques," Queensland University of Technology, 2020.
[15] Detection Using Convolutional Neural Networks," in 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 2020, pp. 136-140.
[16] T. D. Nguyen, "Multimodal emotion recognition using deep learning techniques," Queensland University of Technology, 2020.
[17] H. I. Dino and M. B. Abdulrazzaq, "A Comparison of Four Classification Algorithms for Facial Expression Recognition," Polytechnic Journal, vol. 10, pp. 74-80, 2020.
[18] H. Miao, Y. Zhang, W. Li, H. Zhang, D. Wang, and S. Feng, "Chinese Multimodal Emotion Recognition in Deep and Traditional Machine Leaming Approaches," in 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), 2018, pp. 1-6.
[19] Yanagimoto M, Sugimoto C. Recognition of persisting emotional valence from EEG using convolutional neural networks. In2016 IEEE 9th International Workshop on Computational Intelligence and Applications (IWCIA) 2016 Nov 5 (pp. 27-32). IEEE.
[20] Li X, Song D, Zhang P, Yu G, Hou Y, Hu B. Emotion recognition from multi-channel EEG data through convolutional recurrent neural network. In2016 IEEE international conference on bioinformatics and biomedicine (BIBM) 2016 Dec 15 (pp. 352-359). IEEE.
[21] Salama ES, El-Khoribi RA, Shoman ME, Shalaby MA. EEG-based emotion recognition using 3D convolutional neural networks. Int. J. Adv. Comput. Sci. Appl. 2018 Jan; 9(8):329-37.
[22] Kwon YH, Shin SB, Kim SD. Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system. Sensors. 2018 May;18(5):1383.
[23] Chen JX, Zhang PW, Mao ZJ, Huang YF, Jiang DM, Zhang YN. Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks. IEEE Access. 2019 Mar 29; 7:44317-28.
[24] Wang Z, Tong Y, Heng X. Phase-locking value based graph convolutional neural networks for emotion recognition. IEEE Access. 2019 Jul 10;7:93711-22.
[25] Zhan Y, Vai MI, Barma S, Pun SH, Li JW, Mak PU. A computation resource friendly convolutional neural network engine for EEG-based emotion recognition. In2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA) 2019 Jun 14 (pp. 1-6). IEEE.
[26] Gao Z, Li Y, Yang Y, Wang X, Dong N, Chiang HD. A GPSO-optimized convolutional neural networks for EEG-based emotion recognition. Neurocomputing. 2020 Mar 7; 380:225-35.
[27] Er MB, Çiğ H, Aydilek İB. A new approach to recognition of human emotions using brain signals and music stimuli. Applied Acoustics. 2021 Apr 1;175:107840.
[28] Keelawat P, Thammasan N, Numao M, Kijsirikul B. A Comparative Study of Window Size and Channel Arrangement on EEG-Emotion Recognition Using Deep CNN. Sensors. 2021 Jan; 21(5):1678.
[29] Topic A, Russo M. Emotion recognition based on EEG feature maps through deep learning network. Engineering Science and Technology, an International Journal. 2021 Apr 16.
[30] Hwang S, Hong K, Son G, Byun H. Learning CNN features from DE features for EEG-based emotion recognition. Pattern Analysis and Applications. 2020 Aug; 23(3):1323-35.

Biography

FIRST AUTHOR: Firoozeh Shahrokni received the B.S. degree in biomedical engineering from the Islamic Azad University of Dezful, in 2018. She is currently a M.S. student in the Biomedical Engineering Department at, Islamic Azad University of Central Tehran Branch, Tehran, Iran. Her research interests include signal and image processing.

SECOND AUTHOR: Bentolhoda Ayati received the Ph.D. degree in biomedical engineering from Loughborough University, U.K., in 2016. She is currently an Assistant Professor with the Biomedical Engineering Department, Islamic Azad University of Tehran-Central Branch, Tehran, Iran. She has experiences in biomedical data collection and image reconstruction. Her research interests include signal and image processing.