ORIGINAL ARTICLE

Emotion Recognition Using Continuous Wavelet Transform and Ensemble of Convolutional Neural Networks through Transfer Learning from Electroencephalogram Signal

Sara Bagherzadeh [1], Keivan Maghooli [1*], Ahmad Shalbaf [2*], Arash Maghsoudi [1]

[1] Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

[2] Department of Biomedical Engineering and Medical Physics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

*Corresponding Authors: Keivan Maghooli, Ahmad Shalbaf
Email: keivanmaghooli91@gmail.com, Shalbaf@sbmu.ac.ir

Received: 18 January 2022 / Accepted: 04 April 2022

Abstract

Purpose: Emotions are integral brain states that can influence our behavior, decision-making, and functions. Electroencephalogram (EEG) is an appropriate modality for emotion recognition since it has high temporal resolution and is a non-invasive and cheap technique.

Materials and Methods: A novel approach based on Ensemble pre-trained Convolutional Neural Networks (ECNNs) is proposed to recognize four emotional classes from EEG channels of individuals watching music video clips. First, scalograms are built from one-dimensional EEG signals by applying the Continuous Wavelet Transform (CWT) method. Then, these images are used to re-train five CNNs: AlexNet, VGG-19, Inception-v1, ResNet-18, and Inception-v3. Then, the majority voting method is applied to make the final decision about emotional classes. The 10-fold cross-validation method is used to evaluate the performance of the proposed method on EEG signals of 32 subjects from the DEAP database.

Results: The experiments showed that applying the proposed ensemble approach in combinations of scalograms of frontal and parietal regions improved results. The best accuracy, sensitivity, precision, and F-score to recognize four emotional states achieved 96.90% ± 0.52, 97.30 ± 0.55, 96.97 ± 0.62, and 96.74 ± 0.56, respectively.

Conclusion: So, the newly proposed model from EEG signals improves recognition of the four emotional states in the DEAP database.

Keywords: Emotion Recognition; Electroencephalogram; Deep Learning; Transfer Learning; Ensemble Approach; Continuous Wavelet Transform.

DOI: https://doi.org/10.18502/fbt.v10i1.11512

Frontiers in BIOMEDICAL TECHNOLOGIES

1. Introduction

Emotion is a mental state that arouses when confronted with various kinds of stimuli such as watching videos [1]. Physiological responses such as increasing heart rate and sweating hands vary in each emotion. Also, facial expressions like eyebrow movement, eyes and lips expressions, etc. vary in each emotion and represent a unique emotional state. Humans experience a wide spectrum of emotions like happiness, sadness, anger, fear, surprise, and other emotions in daily life. Emotions affect human functions such as attention, working memory etc. [2-4]. For example, people while watching emotional faces an attentional process takes place during that stimulation and an Event Related Potential (ERP) arouses.

Nowadays, recognition and understanding of emotional classes is a great subject in many areas such as affective Brain-Computer Interface (aBCI) [1, 5], diagnosis of psychophysiological disorders [6, 7], e-learning, and entertainment [1]. The two-dimensional valence-arousal model is the most common emotional model in recognition studies [8]. Valence describes the amount of desirability during sensing of the emotion and arousal describes the amount of excitation from the emotion. Each person understands emotions in a different way and with different values of valence and arousal. Electroencephalogram (EEG) originates from the Central Nervous System (CNS) and is a good candidate to record brain activity during emotion. The other brain mapping technique is functional Magnetic Resonance Imaging (fMRI) which is more precise in spatial resolution than EEG, but has a lower temporal resolution in response to emotional stimulations. Also, it is more convenient and comfortable for individuals to be sitting in a room and watching emotional clips than laying in a gantry, it can make people with claustrophobia nervous. Moreover, EEG is cheaper than fMRI, therefore, making it a common technology to evaluate brain functions such as attention [2-4] or diagnose mental classes [9,10] and disorders such as schizophrenia [11].

Often, an emotion recognition system is divided into four basic steps: preprocessing EEG signals, extracting linear or nonlinear features [12-17], selecting discriminative features, and classifying emotional classes using machine learning techniques [5,12-14,16-18]. These studies faced some issues such as overfitting, lack of generality, non-flexibility, and high dependency on parameters. Recently, deep learning and especially the Convolutional Neural Network (CNN) becomes more widespread due to its advantages (generality, robustness, and flexibility) [19, 20]. CNNs are widely used in medical diagnosis from images [21] or physiological signals [15,16,22-25,27-32]. For example, Yang *et al.* [15] processed EEG signals using the Recurrence Quantification Analysis (RQA) and then used Channel-Frequency CNN (CFCNN) to classify emotional states. In a recent study, Yang *et al.* [22] used multi-column CNN to discriminate emotional states. In these studies, the EEG signal was processed by a feature extraction method [15, 16, 28], connectivity measure [31, 32], or arranged to form multiple matrixes to enter CNN [22]. In this study, we used time-frequency analysis to make a valuable representation of a single EEG signal as input of CNN.

Also, some studies presented Deep Neural Networks (DNN) [33, 34] or hybrid schemes of deep learning algorithms to discriminate emotional states. For example, the fusion of CNNs and Recurrent Neural Network (RNN) is presented [23-26, 30]. Cascade and parallel combinations are presented to discriminate emotional states [23]. A combination of graph CNN with Long Short-Term Memory (LSTM) (modified version of RNN) is presented to recognize emotional states [25]. A 4-D image consisting of time, frequency, spatial, and channel is presented as input of Convolutional Recurrent Neural Network (CRNN) [24]. These are worthy studies but are more complex due to the combination of two sophisticated deep structures (CNNs and RNNs). In this study, we used a popular time frequency transform which has good recognition performance, while being simple; then, preferred to use different pre-trained CNN models due to benefits such as acceleration in the learning process, ease of use, generalization ability and flexibility.

The aim of this paper is to apply the Ensemble of pre-trained CNNs (ECNNs) to the time-frequency representation of EEG signals to recognize four emotional states during watching music video clips. Indeed, through this method, we applied the effective Continuous Wavelet Transform (CWT) processing method to represent valuable time-frequency information from EEG signal, benefited from transfer the learning approach by deep learning pre-trained CNNs and finally make a decision by an ensemble approach while investigating single or combined brain regions.

2. Materials and Methods

2.1. Emotional Classes

In this paper, the four quarters of the two-dimensional valence-arousal model are considered as four emotional classes [8] (Figure 1). Excitement and happiness belong to the first quarter that has High Valence and Arousal (HVHA) values. Fear or anger belongs to the second quarter that has Low Valence and High Arousal (LVHA) values. Depression and sadness belong to the third quarter and have Low Valence and Arousal (LVLA) values. Calmness belongs to the fourth quarter with High Valence and Low Arousal (HVLA) values.

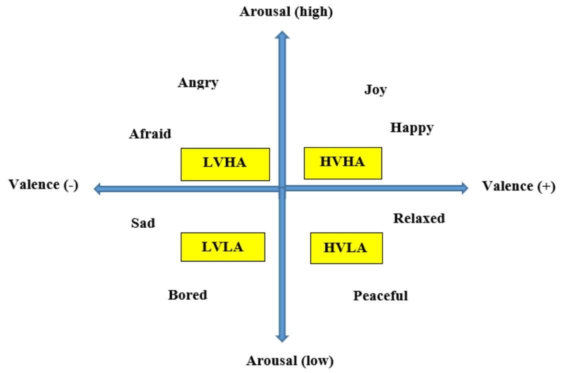


Figure 1. The Valence-arousal emotional model

2.2. DEAP Database

EEG signals from the well-known DEAP dataset were used [35]. This dataset contains 32 EEG channels from 16 men and 16 women in the age ranges from 17 to 37 in two separate locations in Twente (Netherland) and Geneva (Switzerland). EEG channels were recorded from 32 individuals while watching music clips. The length of the signals was 60 seconds, signals were down-sampled to 128 Hz at preprocessing step. The location of EEG channels was based on the international 10-20 electrode system. We considered nine anatomical brain regions (Table 1). Individuals rated each clip based on valence and arousal concepts by the Self-Assessment Manikin (SAM) from 1 to 9 scales (1 = low value, 9 = high value). Also, a group apart from individuals rated videos and labeled them in an online self-assessment. EEG signals from music video clips were chosen that had equal labels from subjects and online self-assessment. Therefore, EEG signals from 33 video clips (7, 8, 9, and 9 clips for HVHA, LVHA, LVLA, and HVLA, respectively) were chosen and others were omitted. Table 2 mentions a number of samples that were used for each emotional class. For example, 224 (32 subjects × 7 clips) samples were used.

Table 1. Brain regions of EEG signals in DEAP database

Region	Electrodes
Pre frontal	Fp1, Fp2, AF3, AF4
Frontal	F7, F3, Fz, F4, F8
Frontal-Central	FC5, FC1, FC2, FC6
Central	C3, C4, Cz
Central-Parietal	CP5, CP1, CP2, CP6
Parietal	P7, P3, Pz, P4, P8
Temporal	T7, T8
Parietal-Occipital	PO3, PO4
Occipital	O1, Oz, O2

Table 2. Number of samples per emotional class

HVHA	LVHA	LVLA	HVLA
224	256	288	288

2.3. Represent EEG Signal Using CWT

CNNs require images as input, therefore, the one-dimensional EEG signal must be converted to a two-dimensional image. CWT is a popular time-frequency method that decomposes a signal into its time (1/scale) and frequency components [11]. CWT was introduced to solve the resolution problem of Short-Time Fourier Transform (STFT) and has high-resolution scalogram images. This method is the convolution of a signal with a set of functions created by a continuous function called the mother wavelet. The CWT for the given signal, x(t), is calculated by Equation 1 as below [11]:

$$W_{(a.b)}[x(t)] = \frac{1}{|a|^{1/2}} \int_{-\infty}^{+\infty} x(t) \emptyset^{*}(\frac{t-b}{a})\, dt \qquad (1)$$

Where a is the scale factor (real and positive integer), b is the translational value (real integer), and Ø is the basis wavelet function.

2.4. Pre-Trained CNNs

CNN has a hierarchical robust structure with basic convolutional, pooling, and Fully Connected (FC) layers [19-21]. Features are extracted and reduced in convolutional and pooling layers, respectively. Then, classification is performed in the FC and the softmax layers. Also, the dropout method is used to avoid complex co-adaption for the training samples. This is done by omitting neurons or weights, randomly. Pre-trained CNNs are trained on huge image databases, ImageNet, previously [36]. These CNNs have a good performance on EEG signals [11, 31, 32, 37] therefore, we selected them as a processing tool. Four emotional classes were classified using the AlexNet, VGG-19, Inception-v1, Inception-v3, and ResNet-18.

2.4.1. AlexNet

AlexNet wins the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 [36]. This pre-trained CNN includes 5 convolutional layers, 5 pooling layers with the maximum operator and 3 fully connected layers. This network uses several 3 × 3, 5 × 5 and 11 × 11 filters in convolutional and max pool layers. Also, it uses the dropout technique to prevent the overfitting problem.

2.4.2. VGG-19

VGGNet is the runner-up of ILSVRC 2014; it is a uniform network with thirteen to fifteen convolutional layers, five max pool layers and three connected layers [38]. It has two versions, VGG-16 (with 16 weight layers) and VGG-19 (19 weight layers). In comparison with AlexNet and Inception-v1, VGGNets have more parameters (138 million for VGG-16 and 144 million for VGG-19). In this paper, we used the deeper network, VGG-19.

2.4.3. Inceptions

Inception-v1 won the ILSVRC 2014 on ImageNet dataset and was designed with twenty-one convolutional layers, five max pools, and one fully-connected layer with 1000 neurons [39]. This network has deeper layers than AlexNet but has very fewer parameters (7 million). Inception-v2 was introduced by adding the batch normalization [40] and Inception-v3 was introduced by adding additional factorization ideas to reduce the number of parameters [41]. Inception-V3 is the runner-up of ILSVC in 2015 on the ImageNet dataset. It has seventy-five convolutional layers, five max pool layers and one fully-connected layer with 1000 neurons. Finally, Inception-v3 provides 23.9 million parameters. Inception-v1 and v3 are implemented in MATLAB software, so we applied the two versions to recognize emotional states.

2.4.4. ResNet-18

Residual network (ResNet) wins the ILSVRC in 2015 [42]. The residual unit consisted of convolutional layers with a shortcut from input to output that help to solve the vanishing gradient problem of CNNs. ResNet has different versions, ResNet-18, ResNet-50, ResNet-110, etc. By deepening the architecture, computational costs increase, therefore, we selected the ResNet-50. To summarize this section, Table 3 reports these networks with their specific configurations.

Table 3. The used pre-trained CNNs information

Name	Weight layer	Total numbers of Layers	Image input size
AlexNet	8	25	227×227
Inception-v1	22	144	224×224
ResNet-18	18	77	224×224
VGG-19	19	48	224×224
Inception-v3	48	316	299×299

2.5. Fine-Tuning

The pre-trained CNNs are fine-tuned on newly constructed scalogram images. In fine-tune procedure, the fully connected layer is replaced by a new one for the classification of four emotional classes problem (the fully connected layer of pre-trained CNNs classified 1000 classes) and the parameters of the network are retrained. The Adaptive Moment estimation optimizer (Adam) algorithm was used to fine-tune the mentioned CNNs. Also, the cross-entropy was the loss function of pre-trained CNNs.

2.6. Ensemble Approach

The majority voting method [43] was used to make the final decision about emotional classes. Results of scalograms from five mentioned pre-trained CNNs are seen and the final decision generates, for example, if at most, three pre-trained CNNs predict the test image to be for HVHA class, the final decision is HVHA. If the two CNNs predict the first class (for example HVHA), and on the other hand the other two CNNs predict the second class (for example LVHA), and the last CNN predicts the third or fourth class, the class will be determined based on the accuracy of CNNs. In the other words, if the two CNNs predict the first class (for example HVHA) with the accuracy of 90% and 85%, and the other two CNNs predict the second class (for example LVHA) with the accuracies of 91% and 95%, and the last one predicts the third or fourth class and has the accuracy of 96%, then based on the CNN with the highest accuracy, we decide the sample belongs to the third or fourth class.

2.7. Performance Evaluation

The 10-fold cross-validation method was used to evaluate classification accuracy on the DEAP database. Then, in each fold, the overall accuracy, precision, sensitivity, and F-score measures were calculated according to Equation 2 to Equation 5 and finally, the values of mean and standard deviation were reported [44].

$$Accuracy = \frac{\sum_{i=1}^{l} \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i}}{l} \quad (2)$$

$$Precision = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i}}{l} \quad (3)$$

$$Sensitivity = \frac{\sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i}}{l} \quad (4)$$

$$F - score = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \quad (5)$$

Where, tp_i, tn_i, fp_i and fn_i are true positive, true negative, false positive and false negative elements for ith emotional class from the confusion matrix obtained from each method [31].

2.8. Overview

The flowchart of the proposed classification system for the recognition of four emotional states (HVHA, HVLA, LVLA, and LVHA) is shown in Figure 2. First, scalograms were generated from EEG channels by the CWT technique. Then, the mentioned CNNs were fine-tuned on scalograms and the last FC layer was changed by a layer with four neurons. Then, through the ensemble approach, the majority voting method was used to make the final decision about emotional classes. Finally, 10-fold cross-validation was used to examine pre-trained CNNs and the ensemble approach.

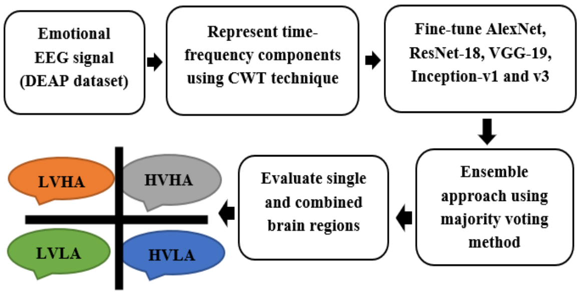


Figure 2. The flowchart of the proposed ensemble of CNNs for the recognition of four emotional states (HVHA, HVLA, LVLA, and LVHA) from EEG signal

3. Results

Multichannel preprocessed EEG signals from 32 subjects from the DEAP database were used in this study. Scalograms were built from the CWT method. The frequency component of each image was from 4 to 45 Hz. Harr mother wavelet and Hanning window were selected due to their effectiveness to build scalogram images. The number of scalogram images was 33792 (32 (channels) × 32 (subjects) × 33 (video clips)). Figure 3 shows a sample of the scalogram for four emotional classes. The

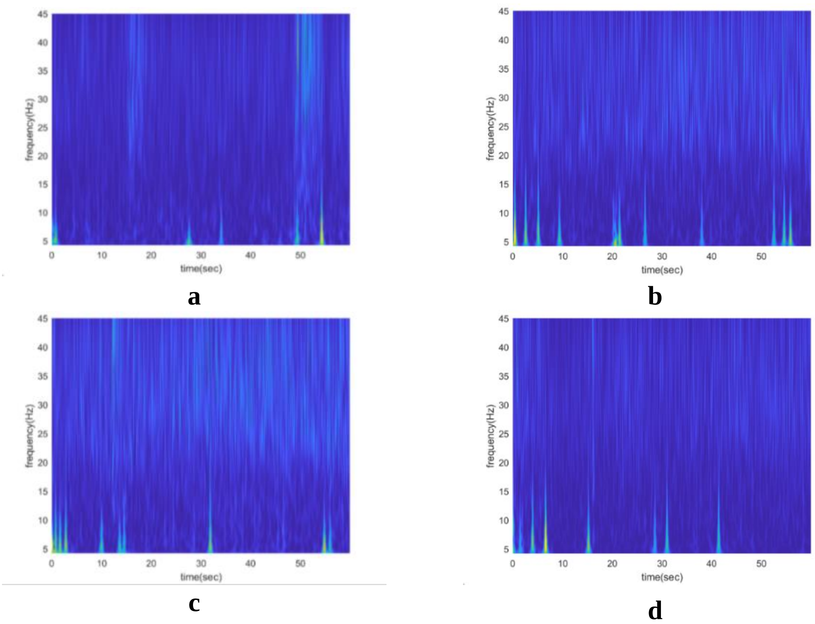


Figure 3. Scalogram of FP1 channel of an individual from four emotional classes of (a) HVHA, (b) LVHA, (c) LVLA, and (d) HVLA

horizontal ax represents time in second and the vertical represents the frequency contents in Hertz. Then, scalograms were resized to input each mentioned CNN model to re-train parameters of them independently. Then, the final decision was made using the ensemble approach, i.e., results of these pre-trained CNNs were seen and the final decision was made using the majority voting method. The initial learning rate was considered 0.0004 for all CNNs except AlexNet (0.0001). Also, the squared gradient decay factor, max epochs, and mini-batch size were selected by error and tried 0.99, 20, and 32, respectively. Codes were performed on an Intel (R) Core (TM) i7-6500U CPU @2.50 GHz 2.60 GHz. Also, CNNs were run on the MATLAB software (version 2021a).

Figure 4 shows the overall accuracy for scalograms of 9 separate brain regions using five mentioned CNN models and the ensemble approach for recognition of the four mentioned emotional classes, respectively. Maximum accuracy was achieved for the ensemble approach in approximately all brain regions, followed by ResNet-18. As it can be observed from this Figure 4, frontal and parietal regions achieved higher accuracies using majority voting with the accuracy of 79% and 76%, using the CWT method. Then, EEG channels of brain regions are combined to increase the recognition rate. All possible combinations of the two brain regions are considered here. Table 3 mentions the highest results for recognition of four emotional states using pre-trained CNNs and the ensemble approach for best combinations of two brain regions.

The optimization algorithm and fine-tune of pre-trained CNNs parameters were done like before. The highest accuracy achieved using the ensemble approach was 96.90 ± 0.52 for combining frontal and parietal regions for recognition of four emotional states. Finally, all combined items from three to more regions are examined. The combination of three and more possible brain regions has lower accuracy values compared with the combination of frontal and parietal regions and did not improve recognition performance, so we did not report them (the highest accuracy achieved at the combination of frontal, frontal-central, and parietal scalograms 87.75%). Therefore, we can differentiate between four emotional states with the accuracy of 96.90 ± 0.52 for the CWT scalogram of combined parietal and frontal regions using the ensemble approach. Figure 5a and 5b show accuracies and loss function in the training process of combined frontal and parietal regions, respectively using AlexNet, ResNet-18, VGG-19, Inception-v1 and v3. As observed in Figure 5, ResNet-18 and Inception-v3 converge more rapidly in the 7th epochs than others.

4. Discussion

An emotion recognition system based on the ensemble of deep learning CNNs through the transfer learning technique from time-frequency components via the CWT method was proposed to recognize four emotional classes from EEG signals of individuals while watching emotional music clips. An accuracy value of 96.90 ± 0.52 is achieved for the ensemble approach in scalogram of combined

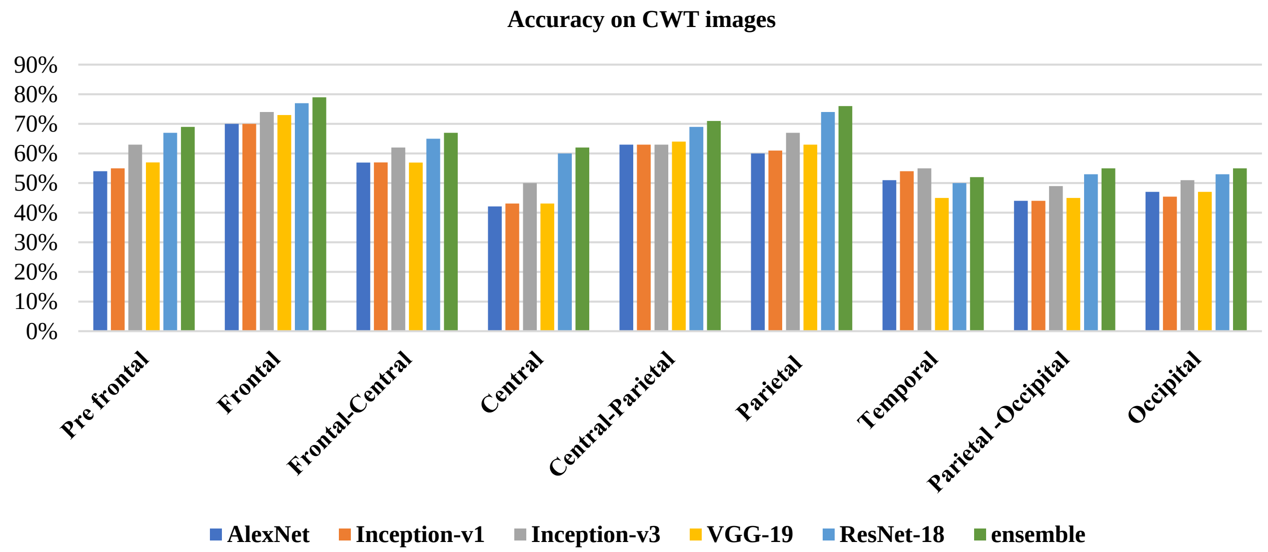
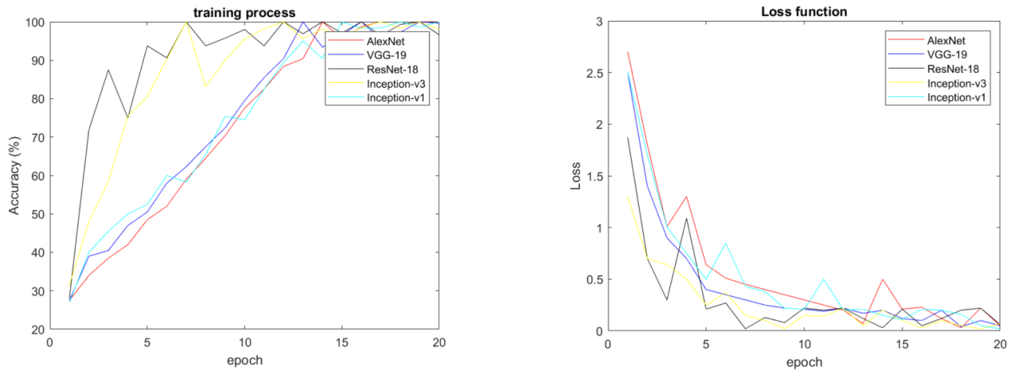


Figure 4. Overall accuracy for AlexNet, VGG-19, Inception-v1, Inception-v3, ResNet-18, and ensemble approach on scalograms from CWT method in 9 brain regions



a b

Figure 5. The accuracy of training process (a) and cross-entropy loss function (a) obtained from the pre-trained CNNs for the best combination of scalograms of parietal and frontal regions

frontal and parietal brain regions from EEGs. CWT was considered due to its effective processing. As observed in Table 3, the highest classification accuracy was achieved through the ensemble approach using the majority voting method than individual pre-trained CNNs. Besides, improving classification results, the ensemble approach causes the final decision more reliable, since it votes according to major pre-trained CNNs results. For example, when an EEG signal from HVHA emotional state is input to the proposed ECCNs method, we are more confident that it would be classified correctly than if we used a single CNN. Following ECNNs, ResNet-18 had better performance than other pre-trained networks (Figure 4). This network is designed from multiple residual units that have two convolutional layers connected by identity shortcuts. This unit helps to improve the accuracy and reduce computational costs. Also, Inception v1 and v3 or VGG-19 are deeper than ResNet-18, but their results showed no significant differences. Deepening networks like VGG does not increase the accuracy, surely; since the network saturates and its performance degrades. Moreover, ResNet-18 had low processing time (90 sec) on the test set than Inception v3 and VGG-19 while having higher accuracy. Also, ResNet-18 has fewer times to be fine-tuned, low-capacity size and requires low memory to be used on ordinary laptops. Finally, as can be observed in Figure 5, ResNet-18 and Inception-v3 are trained sooner than others. This is because these two networks have batch normalization techniques that speed up the convergence. Therefore, the specific and robust architecture of ResNet-18, the usefulness of residual unit information, and the rich information of scalograms lead to learning patterns of emotional classes more accurately and the results of this network are higher than other pre-trained networks.

According to Table 3, the combined parietal and frontal regions demonstrated as most related regions to emotion recognition since achieved the highest accuracy (96.90 ± 0.52). According to neurophysiology studies, the limbic systems are responsible for emotion processing procedures in humans [1, 45]. That limbic systems include the frontal cortex and temporal lobe [46]. Our findings of the best regions support these areas and are consistent with them [47].

Table 4 compares our results with related emotion recognition studies from the EEG DEAP dataset. According to Table 5, this study had achieved higher accuracy than others. Results show that high-level extracted features from deep layers of pre-trained CNNs could better represent four emotional states than the handy crafted non-linear features [17, 18]. Also, our result is higher than [15, 16, 23-27, 30, 34, 35] which used self-designed CNNs, LSTM, DNN, or hybrid systems of CNNs and RNNs. This proves the superiority of the ensemble of pre-trained CNNs and built scalograms. However, our result is a little bit different from [31] (98.16 vs. 96.90); they used effective connectivity maps and pre-trained CNNs.

This work had some difficulties, including, the long-time spent fine-tuning pre-trained CNNs, especially in VGG-19 and Inception-v3. It takes more time for fine-tuning and needs strong processors. For future studies, we will use connectivity measures to build images from EEG signals rather than the CWT method for the cooperation of multiple brain regions in the emotion processing procedure.

Table 4. Recent emotion recognition studies on EEG signals from the DEAP database

Ref	Processing method	Class numbers	Accuracy (%)
[14]	feedback artificial tree, shuffled shepherd optimization, deep maxout network	2	88.9
[16]	3-d feature maps, CNN	4	76.77
[17]	Differential entropy, Graph regularized Extreme Learning Machine	4	69.67
[18]	Histogram of oriented gradient, connectivity measures, support vector machine		95.21
[15]	RQA, parallel CRNN	2	90.8 (valence), 91.03 (arousal)
[23]	2-d mesh, PCRNN, CCRNN	2	93.17 (valence), 93.05 (arousal)
[24]	Differential entropy, 4-D-CRNN	2	94.22 (valence), 94.58 (arousal)
[25]	GCNN-LSTM	2	90.45 (valence), 90.60 (arousal),
[26]	multimodal residual LSTM	2	92.30 (valence), 92.87 (arousal)
[27]	Ensemble CNN	4	82.92
[29]	lightweight pyramidal 1-d CNN	2	98.43 (valence), 97.65 (arousal)
[30]	LSTM, attention, CNN	2	65.9 (valence), 69.5 (arousal)
[31]	Effective connectivity maps, pre-trained CNNs	5	98.16
[33]	CWT, pre-trained CNN-multiclass support vector machine	4	87.45
[34]	DNN, Sparse autoencoder	2	89.49 (valence), 92.86 (arousal)
[35]	2-d feature maps, DNN	2	97.69 (valence), 97.53 (arousal)
Proposed method	CWT, AlexNet, ResNet-18, VGG-19, Inception-v1, Inception-v3, majority voting	4	96.90

Table 5. Highest accuracy values for classifying four emotional classes using the pre-trained CNNs and ensemble of them on scalograms of the best combination of two brain regions (frontal and parietal)

method	**Accuracy (%)** (mean ± std)	**Precision (%)** (mean ± std)	**Sensitivity (%)** (mean ± std)	**F-score (%)** (mean ± std)
ResNet-18	94.87 ± 0.54	94.94 ± 0.55	95.27 ± 0.53	94.71 ± 0.55
Inception-v3	92.50 ± 0.51	92.45 ± 0.56	93.95 ± 0.56	92.25 ± 0.54
Inception-v1	87.50 ± 0.53	87.65 ± 0.60	88.28 ± 0.55	87.34 ± 0.55
VGG-19	85.60 ± 0.52	85.40 ± 0.55	86.56 ± 0.52	85.38 ± 0.52
AlexNet	83.50 ± 0.53	83.75 ± 0.50	83.90 ± 0.55	83.75 ± 0.53
ECNNs	**96.90 ± 0.52**	**96.97 ± 0.62**	**97.30 ± 0.55**	**96.74 ± 0.56**

5. Conclusion

Ensemble transfer learning and CWT method are applied successfully to recognize four emotional classes from individuals while watching emotional music video clips (EEG signals from DEAP database). The accuracy, sensitivity, precision, and F-score of the proposed method for combined parietal and frontal brain regions are 96.90% ± 0.52, 97.30 ± 0.55, 96.97 ± 0.62, and 96.74 ± 0.56, respectively. Based on the results, the proposed model successfully analyzed the emotional function of the brain and found related brain regions to emotions.

Acknowledgments

We would like to thank Science and Research Branch, Islamic Azad University for their support.

References

1- Soraia M Alarcao and Manuel J Fonseca, "Identifying emotions in images from valence and arousal ratings." *Multimedia Tools and Applications,* Vol. 77 (No. 13), pp. 17413-35, (2018).

2- Ross Gordon, Joseph Ciorciari, and Tom van Laer, "Using EEG to examine the role of attention, working memory, emotion, and imagination in narrative transportation." *European Journal of Marketing,* (2018).

3- Colleen A Brenner, Samuel P Rumak, Amy MN Burns, and Paul D Kieffaber, "The role of encoding and attention in facial emotion memory: an EEG investigation." *International journal of psychophysiology,* Vol. 93 (No. 3), pp. 398-410, (2014).

4- Sebastian Schindler and Florian Bublatzky, "Attention and emotion: An integrative review of emotional face processing as a function of attention." *Cortex,* Vol. 130pp. 362-86, (2020).

5- John Atkinson and Daniel Campos, "Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers." *Expert Systems with Applications,* Vol. 47pp. 35-41, (2016).

6- Saime Akdemir Akar, Sadık Kara, Sümeyra Agambayev, and Vedat Bilgiç, "Nonlinear analysis of EEGs of patients with major depression during different emotional states." *Computers in biology and medicine,* Vol. 67pp. 49-60, (2015).

7- Yingjie Li, Dan Cao, Ling Wei, Yingying Tang, and Jijun Wang, "Abnormal functional connectivity of EEG gamma band in patients with depression during emotional face processing." *Clinical neurophysiology,* Vol. 126 (No. 11), pp. 2078-89, (2015).

8- James A Russell, "A circumplex model of affect." *Journal of personality and social psychology,* Vol. 39 (No. 6), p. 1161, (1980).

9- Antonio Maria Chiarelli, Pierpaolo Croce, Arcangelo Merla, and Filippo Zappasodi, "Deep learning for hybrid EEG-fNIRS brain–computer interface: application to motor imagery classification." *Journal of neural engineering,* Vol. 15 (No. 3), p. 036028, (2018).

10- Fahimeh Afshani, Ahmad Shalbaf, Reza Shalbaf, and Jamie Sleigh, "Frontal–temporal functional connectivity of EEG signal by standardized permutation mutual information during anesthesia." *Cognitive neurodynamics,* Vol. 13 (No. 6), pp. 531-40, (2019).

11- Ahmad Shalbaf, Sara Bagherzadeh, and Arash Maghsoudi, "Transfer learning with deep convolutional neural network for automated detection of schizophrenia from EEG signals." *Physical and Engineering Sciences in Medicine,* Vol. 43 (No. 4), pp. 1229-39, (2020).

12- Abdulhamit Subasi, Turker Tuncer, Sengul Dogan, Dahiru Tanko, and Unal Sakoglu, "EEG-based emotion recognition using tunable Q wavelet transform and rotation forest ensemble classifier." *Biomedical Signal Processing and Control,* Vol. 68p. 102648, (2021).

13- Tian Chen, Sihang Ju, Fuji Ren, Mingyan Fan, and Yu Gu, "EEG emotion recognition model based on the LIBSVM classifier." *Measurement,* Vol. 164p. 108047, (2020).

14- KS Bhanumathi, D Jayadevappa, and Satish Tunga, "Feedback Artificial Shuffled Shepherd Optimization-Based Deep Maxout Network for Human Emotion Recognition Using EEG Signals." *International Journal of Telemedicine and Applications,* Vol. 2022(2022).

15- Yu-Xuan Yang *et al.*, "A recurrence quantification analysis-based channel-frequency convolutional neural network for emotion recognition from EEG." *Chaos: An Interdisciplinary Journal of Nonlinear Science,* Vol. 28 (No. 8), p. 085724, (2018).

16- Hao Chao and Liang Dong, "Emotion recognition using three-dimensional feature and convolutional neural network from multichannel EEG signals." *IEEE sensors journal,* Vol. 21 (No. 2), pp. 2024-34, (2020).

17- Wei-Long Zheng, Jia-Yi Zhu, and Bao-Liang Lu, "Identifying stable patterns over time for emotion recognition from EEG." *IEEE Transactions on Affective Computing,* Vol. 10 (No. 3), pp. 417-29, (2017).

18- Yunyuan Gao, Xiangkun Wang, Thomas Potter, Jianhai Zhang, and Yingchun Zhang, "Single-trial EEG emotion recognition using Granger Causality/Transfer Entropy analysis." *Journal of Neuroscience Methods,* Vol. 346p. 108904, (2020).

19- Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert, "Deep learning-based electroencephalography analysis: a systematic review." *Journal of neural engineering,* Vol. 16 (No. 5), p. 051001, (2019).

20- Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S Lew, "Deep learning for visual understanding: A review." *Neurocomputing,* Vol. 187pp. 27-48, (2016).

21- Geert Litjens *et al.*, "A survey on deep learning in medical image analysis." *Medical image analysis,* Vol. 42pp. 60-88, (2017).

22- Heekyung Yang, Jongdae Han, and Kyungha Min, "A multi-column CNN model for emotion recognition from EEG signals." *Sensors,* Vol. 19 (No. 21), p. 4736, (2019).

23- Jingxia Chen, Dongmei Jiang, Yanning Zhang, and Pengwei Zhang, "Emotion recognition from spatiotemporal EEG representations with hybrid convolutional recurrent neural networks via wearable multi-channel headset." *Computer Communications,* Vol. 154pp. 58-65, (2020).

24- Fangyao Shen, Guojun Dai, Guang Lin, Jianhai Zhang, Wanzeng Kong, and Hong Zeng, "EEG-based emotion recognition using 4D convolutional recurrent neural network." *Cognitive Neurodynamics,* Vol. 14 (No. 6), pp. 815-28, (2020).

25- Yongqiang Yin, Xiangwei Zheng, Bin Hu, Yuang Zhang, and Xinchun Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM." *Applied Soft Computing,* Vol. 100p. 106954, (2021).

26- Haiping Huang, Zhenchao Hu, Wenming Wang, and Min Wu, "Multimodal emotion recognition based on ensemble convolutional neural network." *IEEE Access,* Vol. 8pp. 3265-71, (2019).

27- Shuaiqi Liu, Xu Wang, Ling Zhao, Jie Zhao, Qi Xin, and Shui-Hua Wang, "Subject-independent emotion recognition of EEG signals based on dynamic empirical convolutional neural network." *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* Vol. 18 (No. 5), pp. 1710-21, (2020).

28- Emad-ul-Haq Qazi, Muhammad Hussain, Hatim AboAlsamh, and Ihsan Ullah, "Automatic Emotion Recognition (AER) System based on Two-Level Ensemble of Lightweight Deep CNN Models." *arXiv preprint arXiv:1904.13234,* (2019).

29- Aniket Singh Rajpoot and Mahesh Raveendranatha Panicker, "Subject independent emotion recognition using EEG signals employing attention driven neural networks." *Biomedical Signal Processing and Control,* Vol. 75p. 103547, (2022).

30- Sara Bagherzadeh, Keivan Maghooli, Ahmad Shalbaf, and Arash Maghsoudi, "Emotion recognition using effective connectivity and pre-trained convolutional neural networks in EEG signals." *Cognitive Neurodynamics,* pp. 1-20, (2022).

31- Sara Bagherzadeh, Keivan Maghooli, Ahmad Shalbaf, and Arash Maghsoudi, "Recognition of emotional states using frequency effective connectivity maps through transfer learning approach from electroencephalogram signals." *Biomedical Signal Processing and Control,* Vol. 75p. 103544, (2022).

32- Junxiu Liu *et al.*, "EEG-based emotion classification using a deep neural network and sparse autoencoder." *Frontiers in Systems Neuroscience,* p. 43, (2020).

33- Juan Cheng *et al.*, "Emotion recognition from multi-channel eeg via deep forest." *IEEE Journal of Biomedical and Health Informatics,* Vol. 25 (No. 2), pp. 453-64, (2020).

34- Jiaxin Ma, Hao Tang, Wei-Long Zheng, and Bao-Liang Lu, "Emotion recognition using multimodal residual LSTM network." in *Proceedings of the 27th ACM international conference on multimedia,* (2019), pp. 176-83.

35- Sander Koelstra *et al.*, "Deap: A database for emotion analysis; using physiological signals." *IEEE transactions on affective computing,* Vol. 3 (No. 1), pp. 18-31, (2011).

36- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems,* Vol. 25(2012).

37- Yashar Taghizadegan, Nader Jafarnia Dabanloo, Keivan Maghooli, and Ali Sheikhani, "Obstructive sleep apnea event prediction using recurrence plots and convolutional neural networks (RP-CNNs) from polysomnographic signals." *Biomedical Signal Processing and Control,* Vol. 69p. 102928, (2021).

38- Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556,* (2014).

39- Christian Szegedy *et al.*, "Going deeper with convolutions." in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), pp. 1-9.

40- Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift." in *International conference on machine learning*, (2015): *PMLR*, pp. 448-56.

41- Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna, "Rethinking the inception architecture for computer vision." in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), pp. 2818-26.

42- Shaoqing Ren, Jian Sun, K He, and X Zhang, "Deep residual learning for image recognition." in *CVPR*, (2016), Vol. 2, p. 4.

43- Ludmila I Kuncheva, Combining pattern classifiers: methods and algorithms. *John Wiley & Sons*, (2014).

44- Marina Sokolova and Guy Lapalme, "A systematic analysis of performance measures for classification tasks." *Information processing & management,* Vol. 45 (No. 4), pp. 427-37, (2009).

45- Edmund T Rolls, "Limbic systems for emotion and for memory, but no single limbic system." *Cortex,* Vol. 62pp. 119-57, (2015).

46- Peter J Morgane, Janina R Galler, and David J Mokler, "A review of systems and networks of the limbic forebrain/limbic midbrain." *Progress in neurobiology,* Vol. 75 (No. 2), pp. 143-60, (2005).

47- Wei-Long Zheng and Bao-Liang Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks." *IEEE Transactions on Autonomous Mental Development,* Vol. 7 (No. 3), pp. 162-75, (2015).