

Received January 18, 2022, accepted February 11, 2022, date of publication February 28, 2022, date of current version March 9, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3155647

# Deep Learning Model With Adaptive Regularization for EEG-Based Emotion Recognition Using Temporal and Frequency Features

ALIREZA SAMAVAT<sup>1</sup>, EBRAHIM KHALILI<sup>2</sup>, BENTOLHODA AYATI<sup>1</sup>, AND MARZIEH AYATI<sup>3</sup>

<sup>1</sup>Department of Biomedical Engineering, Islamic Azad University of Tehran-Central Branch, Tehran 13185/768, Iran

<sup>2</sup>Department of Biomedical Engineering, University of Tarbiat Modares, Tehran, 14115/111, Iran

<sup>3</sup>Department of Computer Science, University of Texas Rio Grande Valley, Edinburg, TX 78539, USA

Corresponding author: Bentolhoda Ayati (tehranmarkaz92@yahoo.com)

**ABSTRACT** Since EEG signal acquisition is non-invasive and portable, it is convenient to be used for different applications. Recognizing emotions based on Brain-Computer Interface (BCI) is an important active BCI paradigm for recognizing the inner state of persons. There are extensive studies about emotion recognition, most of which heavily rely on staged complex handcrafted EEG feature extraction and classifier design. In this paper, we propose a hybrid multi-input deep model with convolution neural networks (CNNs) and bidirectional Long Short-term Memory (Bi-LSTM). CNNs extract time-invariant features from raw EEG data, and Bi-LSTM allows long-range lateral interactions between features. First, we propose a novel hybrid multi-input deep learning approach for emotion recognition from raw EEG signals. Second, in the first layers, we use two CNNs with small and large filter sizes to extract temporal and frequency features from each raw EEG epoch of 62-channel 2-s and merge with differential entropy of EEG band. Third, we apply the adaptive regularization method over each parallel CNN's layer to consider the spatial information of EEG acquisition electrodes. The proposed method is evaluated on two public datasets, SEED and DEAP. Our results show that our technique can significantly improve the accuracy in comparison with the baseline where no adaptive regularization techniques are used.

**INDEX TERMS** EEG, emotion recognition, deep learning.

## I. INTRODUCTION

EEG is a complex time-series data used by researchers to recognize brain activation and implement the acquisition and storage of EEG signals in the clinical and psychiatric fields. In particular, in the last ten years, EEG signals are used to reflect the human brain's activity in the field of Brain-computer interface (BCI). Automatic recognition of human emotion is an important research area in human-computer interaction [1]. A perfect ideal BCI can detect arousal or valence of emotion state through spontaneous EEG signals without explicit user feedback [2].

A variety of machine learning techniques and models are used for EEG-based emotion recognition problems. Generally, the methods can be described by three pro-

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

cedures: pre-processing, feature extraction, and classification. Obviously, the pre-processing procedures prepare data prior to any further analysis. There are two primary technical approaches in feature extraction: traditional machine learning and deep learning methods. The traditional approach relies on handcrafted characteristics [3]. Due to the high non-stationarity of EEG signals, the extraction of such features is a difficult task and needs an expert's knowledge. The more recent studies used multivariate statistical analysis techniques in the frequency, time-frequency, and nonlinear domain, and capture handcrafted feature that can represent EEG characters [4]–[8]. For example, Zheng *et al.* [4] evaluated the performance of different common methods for feature extraction. They investigated Discriminative Graph regularized Extreme Learning Machine using differential entropy features to achieve the best average accuracy. Yang *et al.* [5] proposed multiple features for the formation

of high-dimensional features. Generally, traditional recognition methods usually combined handcrafted features and shallow models like support vector machines (SVM) and ensemble learning [9], [10].

The recent development of deep learning-based methods is becoming dominant to boost the efficiency and generalizability of EEG applications. It is claimed that deep learning method addresses the shortcomings of recent approaches, such as the requirement of a large amount of data [11]. Therefore, deep learning networks have gained substantial interest from researchers in the literature [12], [13]. Convolutional Neural Network is one of most important deep networks because of its excellent performance in the signal processing [14]. CNNs allow the extraction of higher level features from the EEG signal. Recent articles demonstrate CNN may have promising performance due to its advantages of automatic feature detection. Also, Recurrent Neural Networks (RNN) as a family of neural networks is used for processing variable-length sequential data [15]. However, using deep RNN architectures could lead to the problem of gradients vanishing or exploding. Recent study showed that long short-term memory (LSTM) outperform traditional RNNs substantially [16].

In this paper, we introduce a multi-input deep model with CNNs and Bidirectional Long Short-term Memory (Bi-LSTM). Bi-LSTM is a two-way LSTM that incorporates the LSTM forwarding direction and the reverse path. In our model, each individual parallel input, at the first layers, consists of two CNNs with small and large filters. Also, for each segment, we extract five frequency bands of EEG signals. Since the experimental results indicate that the gamma and beta band (roughly beta: 14-30 Hz, gamma: 31-50 Hz) is suitable for EEG-based emotion classification, we calculated differential entropy for two bands of frequency. Finally, CNN's features are integrated with differential entropy feature. We applied the adaptive regularization method over each parallel CNN's layer to consider the spatial information of EEG acquisition electrodes.

In summary, the major contributions of this paper are:

- A multi-input deep learning CNN+BiLSTM architecture to perform the classification of raw signals from EEG channels with differential entropy for the two band of frequencies.
- An approach with the ability to extract temporal EEG patterns as well as frequency components simultaneously, which leads to improvement of the performance of our model.
- Adjusting the regularization parameter of the model adaptively to utilize the unnecessary channels and prevent overfitting.

The rest of this paper is organized as following: In section II, we describe the research methodology by elucidating our presented deep learning model. In Section III, we explain the model training and prediction algorithm. In section IV, we present the results of our analysis. Finally,

Section V and VI draws the discussion of this work and conclusion, respectively.

## II. MATERIALS AND METHODS

In this section, we first describe the SEED and DEAP dataset and the pre-processing procedures. Then, we introduce our model in details. Finally, we explain the training parameters used in our model.

### A. DATASET AND DATA PREPARATION

#### 1) SEED

SEED is a electroencephalogram signal data measured during the emotional experiment. The dataset was collected from 15 subjects (7 males and 8 females) while showing them 15 chinese movie clip. The focus of the dataset is on three specific emotions: positive, neutral and negative. Therefore, there are 45 trials in the database. Each scalp EEG signal was collected with the standard 10-20 system by 62- channel and downsampled to 200Hz sampling rate. Also, a bandpass frequency filter of 0-75 Hz was applied to remove physiological noises [17].

#### 2) DEAP

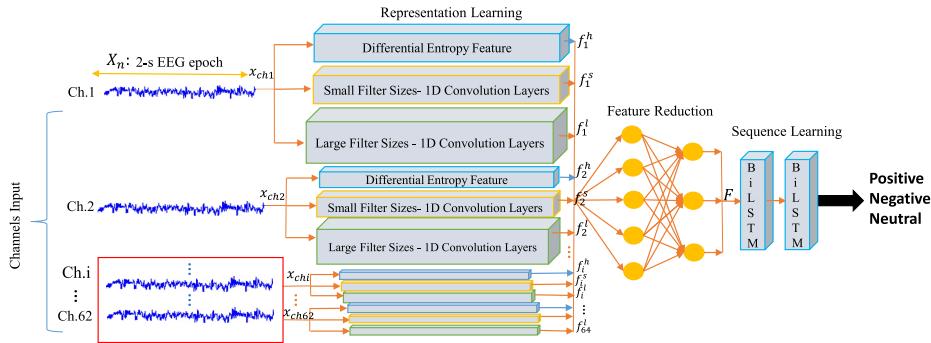
The DEAP dataset includes 40 channels of peripheral physiological signals from 32 participants (16 males and 16 females) when watching 40 one-minute music videos. Among the 40 trials for each subject, various signals were recorded as 40-channel data. The first 32 channels are EEG signals, and the last 8 channels are autonomous physiological signals. After watching the video, participants marked each video in terms of the levels of arousal, valence, like/dislike, dominance, and familiarity. Valence is the scale used for this work. It ranges from one (low) to nine (high), and scales are divided into three parts to construct positive, neutral, and negative labels in accordance with the SEED dataset. Emotions are categorized negative if the valence rating is smaller than 3; neutral if valence rating is smaller than 7 and greater than 3; and positive if valence rating is greater than 7 [18].

### B. PREPROCESSING

To remove noise and the artifacts, the EEG data is pre-processed with a bandpass filter between 0.3 to 50 Hz. After Noise removal, we separate the EEG segments corresponding to *Movi-clip* and remove excess parts including *Hint* of start, *Self-assessment*, and *Rest*. The length of EEG signals collected after removing extra fragments are 3300 clean epochs for one experiment as explained in [19]. Finally, EEG data in each channel is divided into 2s segments without overlap. Each segment is standardized to unit variance and zero mean.

### C. ARCHITECTURE DESIGN

As shown in Figure 1, our new architecture is made up of four sequential main parts, including representation learning, feature reduction, sequence learning and classification.



**FIGURE 1.** Architecture of the proposed model for emotion recognition.

In representation learning part, the model is trained to extract features from each of raw single-channel EEG epochs, both manually and automatically. Second part is reducing the number of features to prevent heavy computation. Third part is sequence learning, which provides temporal information with the Bi-LSTM block. The last part is classification unit which labels the emotions to Positive, Negative and Neutral. The details of each part are given below.

### 1) REPRESENTATION LEARNING

This structure uses all the spatial, temporal and frequency information of EEG signals. For each raw single channel 2-s EEG epochs, we apply two CNNs with small and large filter sizes at the first layers. The idea of having filters with different sizes comes from the context of signal processing that offers trade-off between temporal and frequency precision and resolution. Namely, the frequency resolution is defined by the number of samples in the time series, and the temporal resolution is always defined by the data sampling rate, and does not depend on the length of the time windows. Thus, small and large filter sizes are best to extract temporal and frequency features respectively [20], [21].

In our proposed model, each two 1D-CNN layers followed by a max-pooling layer to down-sample the input representation. 1D-CNNs consists of three main parameter: a filter size, the number of filters, and a stride size. We adjusted the parameters to extract temporal and frequency information from the EEG. Namely, since we want to analyze the signal between the ranges of 2-50 Hz, the window must be at least 500-20ms long. For two cycles, the length of time segments need to have at least 8 points(EEG Samplingrate = 200, Minimum window size for two cycles of capturing =  $2 \times 20\text{ms}$ , therefore  $200 \times 2 \times 20\text{ms} = 8$  points are required). We can't use three or more cycles, because it fails to catch higher-frequency transient behavior. In our dataset, there are 2-s EEG epochs  $\{x_{ch1}, x_{ch2}, \dots, x_{ch62}\}$  from 62-channel EEG. For each EEG channel, we use two large-filter and small-filter CNNs to extract  $f_n^l$  and  $f_n^s$  features from the  $x_n$ .

### 2) DIFFERENTIAL ENTROPY FEATURE

Entropy is a quantity that measures the disorder of a system. The entropy features has been successfully applied to expand

research of EEG signals. We use an efficient frequency domain feature with the concept of entropy called differential entropy. Differential entropy performs signal decomposition on the original signal to extract useful information. The previous research [5] indicated that the gamma and beta band (beta: 14-30 Hz, gamma: 31-50 Hz) is suitable for EEG-based emotion classification. We calculate the differential entropy of two frequency band of 62 channels, and create features with 124 dimensions for each segment. Differential entropy is defined as follows:

$$h(X) = - \int_X f(x) \log(f(x)) dx \quad (1)$$

$$h(X) = 1/2 \log(2e\pi\sigma^2) \quad (2)$$

$$h_{gamma} = 1/2 \log(2e\pi\sigma_{gamma}^2) \quad (3)$$

$$h_{beta} = 1/2 \log(2e\pi\sigma_{beta}^2) \quad (4)$$

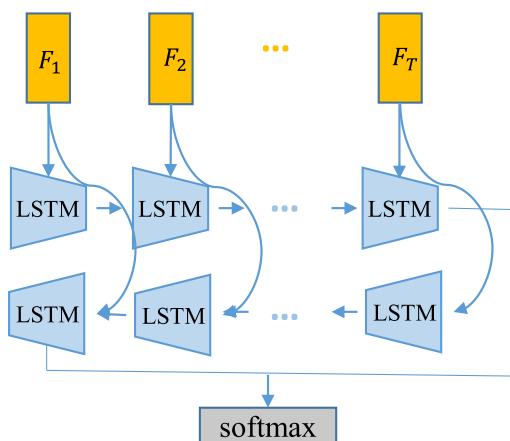
where the time series X is the Gaussian distribution  $N(\mu, \sigma^2)$ . Finally, the moving average filter with window length 20s were applied to smooth the feature sequence [17].

### 3) FEATURE REDUCTION

This unit is a neural network that could help to improve the speed and reliability of computing. At the end of feature extraction, the outputs of CNN's parts are concatenated with Differential Entropy feature and followed by feature reduction networks in order to select and reduce the space of feature.

### 4) SEQUENCE LEARNING

Since EEG signals contain temporal dynamic information, we apply two layers Bi-LSTM cells to extract temporal information. We denote  $F = \{f_n\}_{n=1}^N$  be the set of extracted feature corresponding to the n-th epoch of emotion data with label  $Y = \{y_n\}_{n=1}^N$ . As shown in Fig. 2, Bi-LSTM can be regarded as a Forward LSTM (Fw-LSTM) and a Backward LSTM (Bw-LSTM). Assume  $F_f = F$  represents Forward target feature sequence and  $y_n$  represents the final prediction value, we calculate the prediction sequence as follows:  $y_n = LSTM(F_f)$  where LSTM represents a function that processes sequences of features by using the input gates, forget gates and output gates in the memory block. The output of the



**FIGURE 2.** Structure of Bi-LSTM in the sequence learning Unit of the proposed model.

LSTM layer can be calculated as follows:

$$i_t = \sigma(W_{qi}q_t + W_{hi}h_{t-1} + W_{ci}C_{t-1} + b_i) \quad (5)$$

$$f_t = \sigma(W_{qf}q_t + W_{hf}h_{t-1} + W_{cf}C_{t-1} + b_f) \quad (6)$$

$$o_t = f_t C_{t-1} + i_t \tanh(W_{hc}h_{t-1} + b_c) \quad (7)$$

$$o_t = \sigma(W_{qo}q_t + W_{ho}h_{t-1} + W_{co}C_{t-1} + b_o) \quad (8)$$

The LSTM function is computed as follows:

$$h_t = o_t \tanh(c_t) = \text{LSTM}(h_{t-1}, x_t) \quad (9)$$

where  $\sigma$  is the logistic sigmoid function, and  $t, i, f, o$ , and  $c$  are time step, the input gate, the forget gate, the output gate, and cell activation vectors, respectively. The  $W$  terms are weight matrices and the  $b$  terms are bias vectors. Similarly, we calculate the backward sequence by input  $F_b = \{f_n\}_{n=N}^1$ . We then calculate the hidden vector  $h$  by concatenating  $h_n$  in forward and backward path. The output of the Sequence Learning unit is based on the hidden vector  $h$  as follows:

$$y^n = W_h h + b_h \quad (10)$$

where  $W_h$  and  $b_h$  are the weight matrix and bias vector, respectively.

## 5) CLASSIFICATION

Finally, we predict the label of each emotion EEG segment by softmax. The softmax layer is used as the output layer and give the probabilities of the each classes is computed as follows:

$$Y'_n = P(c | f(x_n, c)) = \frac{\exp(f_i(x_n, c))}{\sum_{i=1}^c \exp(f_i(x_n, c))} \quad (11)$$

where  $P(c | f(x_n, c))$  is the probability of the EEG segment  $x_n$  labeled as class  $c$ , given the real value calculated by our model. We train the model by minimizing the cross-entropy error [22]:

$$w = \text{argmin} \left\{ \frac{1}{N} \sum_{j=1}^N Y_j \log(Y'_j) + (1 - Y_j) \log(1 - Y'_j) \right\} \quad (12)$$

## 6) ADAPTIVE REGULARIZATION

Adaptive regularization method aims to find variant features and reduce weight distribution in layers. In other words, adaptation regularization strategy is actually a trade-off between information of different electrodes with variant spatial resolution that preserved both unnecessary and necessary electrodes in emotion recognition problem. This algorithm consists of three alternating steps: training model, calculating the variance of extracted features from the individual convolution layers, and update regularization parameters. Therefore, the adaptive model iteratively changes the hyperparameters between the adaptive layer and the former layer. In this section, we investigate the adaptive regularization approach in the general formulation. The goal is to adjust the regularization parameters of CNN's layer adaptively. Specifically, the outputs of represents learning unit denoted with  $\{f_i\}_{i=1}^{N_{ch}}$ , where  $f_i = f_i^l || f_i^s$ , with the length of  $R = 64 (2 \times 32 = \text{CNNs feature})$ . To minimize the distance, we uses the standard deviation:

$$\sigma_{fi} = \sqrt{\frac{\sum_{r=1}^R (f_{r,i} - \bar{f}_r)^2}{R}} \quad (13)$$

$$\Sigma_i = \sum_{l=1}^N \sigma_{f_{i,l}} \quad (14)$$

where  $\{f_i\}_{i=1}^{N_{ch}} = \{f_i^l || f_i^s\}_{i=1}^{N_{ch}}$  is the output of represents learning unit (with the length of  $R = 64$ ).

We define  $\Sigma = \{\Sigma_i\}_{i=1}^{N_{ch}}$  where  $\Sigma_i$  is total distribution of pairs large and small CNNs unit. We apply  $L1$  penalty and  $L2$  penalty terms to the kernel weights for each parallel convolutional layers. A general form of the regularization methods is as follows:

$$\min \left\{ \frac{1}{N} \sum_{n=1}^N l(Y_n, f(X_n)) \right\} + \lambda_i \|f\|_k \quad (15)$$

where  $l(\cdot, \cdot)$  is a loss function, and  $\lambda_i$  is the regularization parameter of pairs large and small CNNs unit. When  $f$  is in the linear form and the loss function is square loss,  $f_k$  is the norm of the coefficient of linear model. In this paper  $\lambda$  is the tuning parameter which adapted as follows:

$$\lambda_i = \lambda_0 \frac{\Sigma}{\max(\Sigma)} \quad (16)$$

where  $\lambda_i$  in the  $[0, 1]$ , then  $\lambda_0$  is a constant value smaller than one.

## D. TRAINING ALGORITHM

Our training algorithm is developed based on iterative back-propagation to adjust the network weights (see Algorithm), while applying the adaptive regularization method over each parallel CNNs layer to consider the spatial information of EEG acquisition electrodes. The algorithm trains the model iteratively until performance does not improve for a fixed

**Algorithm 1** Transfer Learning With Adaptive Regularization Parameters

```

Input: train_data, model = multi_input_CNN_LSTM,
batch_size = 128, learning_rate = 0.0005, decay =
0.5
Model Training
3: while accuracy is improved, do training:
   for subject in train_data do:
      for batch in subject do:
         model = model(batch, decay, learning_rate)
         save model weights
4: for channel in sum (variance(model( train_data))):
   L1_norm(two last layer) =  $\lambda_{channel}$ 
5: Update model
6: load model weights
7: Continue training
Output : model

```

number of training epochs. Then, the variance of CNNs representation of all available channels is estimated on the training data and is used to adjust the regularization penalty,  $\lambda$ , adaptively. In line 5 of the Algorithm, the model weights gets updated and fine-tuned on the whole model weights, and learning rates decreases after each fine-tune.

**E. TRAINING PARAMETERS**

We use the ADAM optimizer with default parameters of framework (learning rate set on  $lr=0.0005$ , learning  $decay=0.5$  and decay rate of the first and the second moments ADAM:  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ ) to model training. The model is trained using the training set with the mini-batch size of 120. For the 1D-CNNs in representation learning part, we consider small and large windows size according to the guideline provided in relevant section. For instance, the size of the large windows in the first layer was set to  $kernel\_size = 25$  (because it can capture several cycles from higher-frequency i.e. 50Hz), and its stride size is set to 5 (because, sampling rate is  $F_s = 200$  and importance higher-frequency transient behavior band belongs to emotion is less than 50, then we choose minimum step of  $F_s/4 = 50$  to capture cycles of higher-frequency). The use of stride may minimize the number of parameters, and it can reduce the cost of computing, and can also obtain a comparable complexity of model. Also, the small window size set to  $kernel\_size = 10$  to detect temporal EEG patterns. In each block, we use several sequence layer to allow the extraction of higher level features from the EEG signal. At the end of feature extraction, the outputs of CNN's parts are concatenated with Differential Entropy Feature and obtain 4092 features ( $62 \times 2 \times 32 = electrodes \times small\&large\_blocks \times 1DCNNs\_features + 124 =$  differential entropy of two frequency band extract from 62 channels) and follows by feature reduction in order for selecting by two layers neural networks. In the sequence learning part, the unit of the Bi-LSTM is set near size of the feature reduction output part, which is 256.

**III. RESULTS**

The performance results of the proposed method using SEED and DEAP datasets are presented in this section. The classification confusion matrices obtained from the test set of SEED dataset are illustrated in Figure 3. Each row and column represent the number of 2-s EEG emotion segments classified by the target emotion and our model predicted label, respectively. The diagonal of the table represents the percentage of epochs that our model classified the emotion correctly.

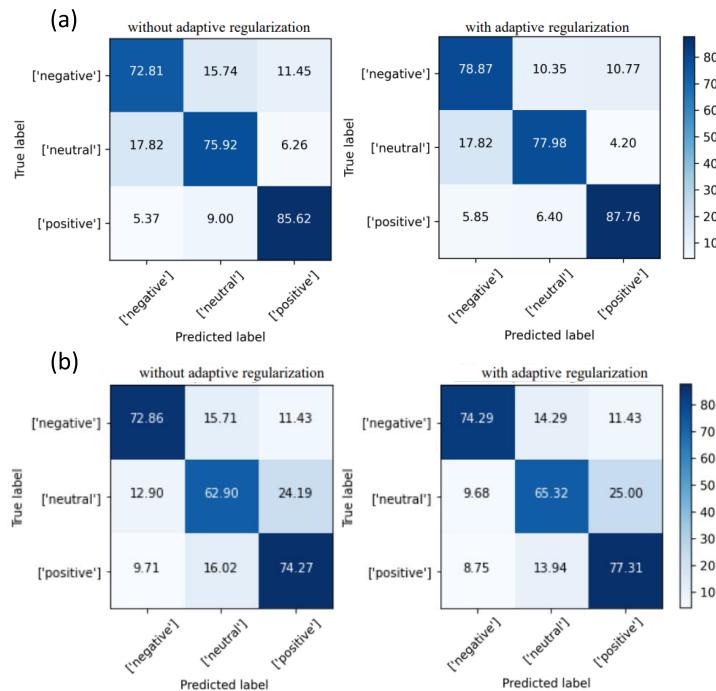
In Figure 4, we show the effect of the adaptive regularization on the final contribution of each electrode. In this figure, each bar represents  $\Sigma$  of each electrode as described in section 2.3.6. The comparison of the two charts clarifies that a set of electrodes like  $\{T7, T8, TP7, FC5, FPZ, F7, F3, FC3, O1, \dots\}$  have more influence on the output of the activation layer. Adaptive regularization strategy optimizes the usage of information of different electrodes with a variant spatial resolution that preserved both unnecessary and necessary electrodes in the emotion recognition problem. From these results, we can see that our method has the ability to smooth the weight of each parallel CNN's layer by considering the spatial information of EEG acquisition electrodes. In order to evaluate the static functional connectivity between the channels, and track their changes over time [23], we calculate the spatial correlation between all the features extracted from a multi-channel signal  $f = \{f_1, f_2, \dots, f_c\}$  of parallel network, where  $c$  is the number of channels and  $f_c = f_{sc} \| f_{lc}$ . For this purpose, we randomly select two samples from SEED dataset, and we calculate the correlation between features with and without regularization techniques. The result of this analysis is presented in Fig 5.

We also compare the accuracy of our proposed methods with state-of-the-art methods that are applied on 10-fold cross validation of SEED and DEAP database. The result of this analysis is presented in Table 1.

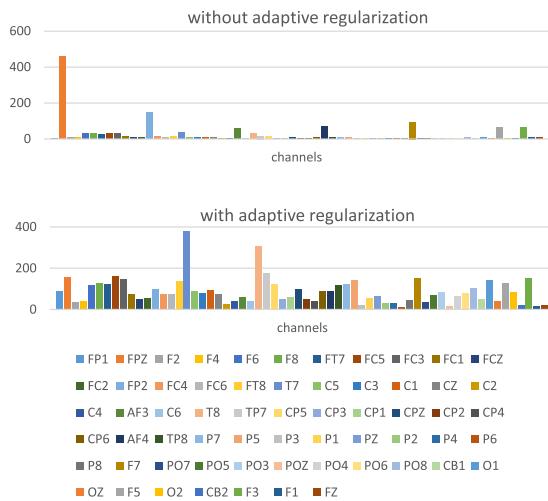
**IV. DISCUSSION**

We propose the deep learning model that utilizes a combination of differential entropy features and automatically extracted from CNNs to recognize the emotion from raw EEGs. The proposed model makes the features of each class more conspicuous and efficient. The proposed structure for CNN's have the ability to extract temporal EEG patterns and frequency components, simultaneously. Also, according the recent literature in emotion recognition, beta and gamma bands are more related to emotion recognition. So we selected just the differential entropy features of two relevant bands. Also we added an adaptive regularization technique to use all the unnecessary channels and prevent overfitting of the model.

Based on specification of the input, we design a hybrid parallel neural network of 1D-CNN and handcrafted features. The parallel architecture allows the identification of the most statistically significant features and automates the extraction of spatial information related to the channel position in EEG signals. Consequently, we apply an adaptive regularization



**FIGURE 3.** The confusion matrices of the proposed models applied on the test set of (a) SEED and (b) DEAP datasets. The left table represents the result from the model without adaptive regularization, and right table represents the result from the model with adaptive regularization.

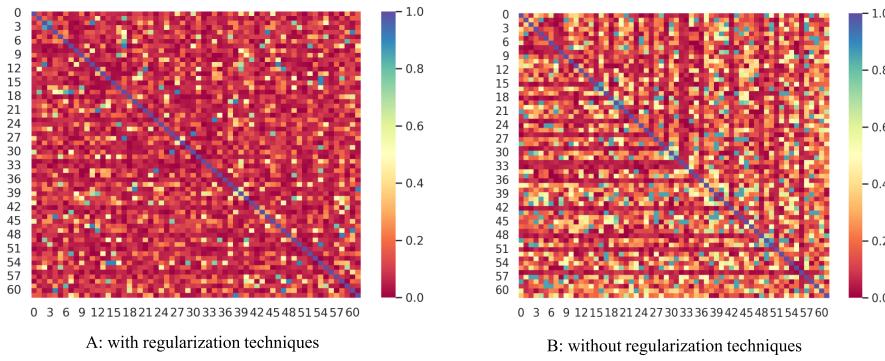


**FIGURE 4.** Effect of adaptive regularization technique on the contribution of each channel. Each bar represents a channel and the y-axis shows  $\Sigma_i$  for each channel.

method to exploited the time-frequency information involved in each channel, and remove the bias related to the correlation of the channel position. Therefore, the method extracts a wide range of features related to EEG signal patterns and overcomes the influence of channel correlation. According to Figure 5, the correlation between some electrodes is significantly higher than other electrodes, which means some brain parts may be more synchronized with other brain parts. Also, it shows that some areas may be responsible for the

generation of particular emotion. The correlation values are between 0 and 1, in which 0 means that  $f_i$  and  $f_j$  are completely uncorrelated while 1 means completely correlated. Comparing two correlation matrix in Figure 5 shows that the pair-wise correlation between the features are reduced after the regularization techniques. It shows that the adaptation regularization strategy establishes a trade-off between the information from functional connectivity and regions, and it focuses on finding silent features for emotion recognition.

Table 1 shows a brief summary and performance comparison of our method and other emotion recognition methods. These methods are deep learning models and used the same evaluation method. Zheng and Lu [24] extracted differential entropy features from multichannel EEG data and applied a Deep Belief Network (DBN) for emotion recognition on the SEED dataset. They identified the weight distributions of the trained DBN models, and they selected more important frequency bands and channels selected according to the tendency of the network. At last, they trained DBN with the profile of just 12 channels and evaluated the results. Recently, Chen Wei et al. [25] utilized simple recurrent units network and ensemble learning and obtained accuracy of 78.8% on the SEED dataset which is 2.74% low than our results. They applied highly complex systems for classification and feature engineering. Wang et al. [26] proposed a multichannel EEG emotion recognition method based on phase-locking value (PLV) graph convolutional neural networks (P-GCNN). Their P-GCNN model consists of the PLV-based graph signals construction, graph convolutional



**FIGURE 5.** Correlation matrix between feature of two randomly selected subjects of SEED data set under the same emotional EEG signal. This figure shows the correlation matrix (A) with the regularization technique and (B) without regularization technique.

**TABLE 1.** Summary of performance of state-of-the-art methods on SEED and DEAP datasets.

Method	year	Features	Method Description	Performance (Accuracy)
Zheng et al [24]	2015	Differential entropy features of five frequency bands	Deep Belief Network (DBN)	SEED: 86.08%
Chen Wei [25]	2020	frequency and nonlinear analysis	Using simple recurrent units network and ensemble learning	SEED: 78.8%
Zhongmin et al [26]	2019	five kinds of EEG features in different frequency bands	Phase-Locking Value Based Graph Convolutional Neural Networks	SEED: 84.3% DEAP: 73.31%
Chen Wei et al [29]	2020	Dual-tree Complex Wavelet Transform (DT-CWT)	Using deep Simple Recurrent Units (SRU)	SEED: 79.62 % DEAP: 76.67%
<b>Our approach (without adaptive regularization)</b>	-	CNNs features + two frequency bands	A hybrid multi-input deep model with convolution neural networks (CNNs) and bidirectional Long Short-term Memory (bi-LSTM)	SEED: 78.11% DEAP: 70.13%
<b>Our approach (with adaptive regularization)</b>	-	CNNs features + two frequency bands	A hybrid multi-input deep model with convolution neural networks (CNNs) and bidirectional Long Short-term Memory (bi-LSTM) with adaptive regularization technique	SEED: 81.54% DEAP: 72.38%

operation, graph pooling operation, Relu activation, and the full connection. This method led to an average accuracy of 84.35% on SEED and 73.31% on DEAP dataset. Although their performance is better than our proposed method, they need the large time series of EEG signal before calculating the PLV. Therefore, their model have computational complexity and also unexpected ability for online calculation. Moreover, they are using handicraft features that may be reduce the generalizability of the model.

The methods presented in [27], [28] has certain limitations including the number of emotion states considered, which result in higher accuracy. In general, the accuracy for the binary class prediction is higher than the multi-class prediction due to removing the complexity of the outcome. In [29]–[31] features were computed in multi-domain (time, wavelet, and frequency) from EEG signals and set of stable features are identified for emotion recognition. Although some studies have reported higher performance, it is important to note that they extract a large number of features and perform extra pre-processing, such as reducing number of

classes or removing subjects from the datasets. Also, selecting the hand engineering features and electrodes could not be efficient for developing an emotion recognition system, and decreases model generalizability.

In our model, all components of feature extraction, feature reduction, and classification are merged and compute in the parallel processes, which enhances computational efficiency. In comparison to other models, our model achieves comparable classification performance. Moreover, the result shows by using proposed adaptive regularization technique, all channels are utilized which may prevent overfitting of the model, and therefore more generalizable with better performance.

## V. CONCLUSION

The proposed architecture in this paper is based on combination of deep learning and differential entropy. We used small and large features to capture heterogeneous features representing temporal and frequency resolution of signals. Also, we used adaptive regularization to remove the bias from the model, and select more effective electrodes to classify

emotions. The proposed method is parallelizable, hence computationally efficient. For future studies, we suggest using this model on independent datasets or leave-one-out cross validation to assess the generalizability of the model.

## ACKNOWLEDGMENT

(Alireza Samavat, Ebrahim Khalili, and Bentolhoda Ayati contributed equally to this work.)

## REFERENCES

- [1] H. Chao and Y. Liu, "Emotion recognition from multi-channel EEG signals by exploiting the deep belief-conditional random field framework," *IEEE Access*, vol. 8, pp. 33002–33012, 2020.
- [2] H. Li, Y.-M. Jin, W.-L. Zheng, and B.-L. Lu, "Cross-subject emotion recognition using deep adaptation networks," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2018, pp. 403–413.
- [3] F. Wang, S. Wu, W. Zhang, Z. Xu, Y. Zhang, C. Wu, and S. Coleman, "Emotion recognition with convolutional neural network and EEG-based EFDMs," *Neuropsychologia*, vol. 146, Sep. 2020, Art. no. 107506.
- [4] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 417–429, Jul. 2019.
- [5] F. Yang, X. Zhao, W. Jiang, P. Gao, and G. Liu, "Multi-method fusion of cross-subject emotion recognition based on high-dimensional EEG features," *Frontiers Comput. Neurosci.*, vol. 13, p. 53, Aug. 2019.
- [6] S. Jirayuchoaroenksa, S. Pan-Ngum, and P. Israsena, "EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *Sci. World J.*, vol. 2014, pp. 1–10, Sep. 2014.
- [7] K. Guo, H. Candra, H. Yu, H. Li, H. T. Nguyen, and S. W. Su, "EEG-based emotion classification using innovative features and combined SVM and HMM classifier," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 489–492.
- [8] E. Khalili and B. M. Asl, "Automatic sleep stage classification using temporal convolutional neural network and new data augmentation technique from raw single-channel EEG," *Comput. Methods Programs Biomed.*, vol. 204, Jun. 2021, Art. no. 106063.
- [9] X. W. Wang, D. Nie, and B. L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, Apr. 2014.
- [10] H. Ullah, M. Uzair, A. Mahmood, M. Ullah, S. D. Khan, and F. A. Cheikh, "Internal emotion classification using EEG signal with sparse discriminative ensemble," *IEEE Access*, vol. 7, pp. 40144–40153, 2019.
- [11] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.
- [12] Y. Li, W. Zheng, Y. Zong, Z. Cui, T. Zhang, and X. Zhou, "A bi-hemisphere domain adversarial neural network model for EEG emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 494–504, Apr. 2021.
- [13] P. Zhong, D. Wang, and C. Miao, "EEG-based emotion recognition using regularized graph neural networks," *IEEE Trans. Affect. Comput.*, early access, May 11, 2020, doi: [10.1109/TAFFC.2020.2994159](https://doi.org/10.1109/TAFFC.2020.2994159).
- [14] M. S. Hossain and G. Muhammad, "Emotion recognition using deep learning approach from audio-visual emotional big data," *Inf. Fusion*, vol. 49, pp. 69–78, Sep. 2019.
- [15] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *Emotion*, vol. 8, no. 10, pp. 355–358, 2017.
- [16] B. Nakisa, M. Rastgoo, A. Rakotonirainy, F. Maire, and V. Chandran, "Long short term memory hyperparameter optimization for a neural network based emotion recognition framework," *IEEE Access*, vol. 6, pp. 49325–49338, 2018.
- [17] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Nov. 2013, pp. 81–84.
- [18] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Müller-Putz, "Domain adaptation techniques for EEG-based emotion recognition: A comparative study on two public datasets," *IEEE Trans. Cogn. Devel. Syst.*, vol. 11, no. 1, pp. 85–94, Mar. 2019.
- [19] Z. Gao, X. Wang, Y. Yang, Y. Li, K. Ma, and G. Chen, "A channel-fused dense convolutional network for EEG-based emotion recognition," *IEEE Trans. Cognit. Develop. Syst.*, vol. 13, no. 4, pp. 945–954, Dec. 2021.
- [20] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, Nov. 2017.
- [21] M. X. Cohen, *Analyzing Neural Time Series Data: Theory and Practice*. Cambridge, MA, USA: MIT Press, 2014.
- [22] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
- [23] D. Iacoviello, A. Petracca, M. Spezialetti, and G. Placidi, "A real-time classification algorithm for EEG-based BCI driven by self-induced emotions," *Comput. Methods Programs Biomed.*, vol. 122, no. 3, pp. 293–303, Dec. 2015.
- [24] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.
- [25] C. Wei, L.-L. Chen, Z.-Z. Song, X.-G. Lou, and D.-D. Li, "EEG-based emotion recognition using simple recurrent units network and ensemble learning," *Biomed. Signal Process. Control*, vol. 58, Apr. 2020, Art. no. 101756.
- [26] Z. Wang, Y. Tong, and X. Heng, "Phase-locking value based graph convolutional neural networks for emotion recognition," *IEEE Access*, vol. 7, pp. 93711–93722, 2019.
- [27] F. Wang, S. Wu, W. Zhang, Z. Xu, Y. Zhang, C. Wu, and S. Coleman, "Emotion recognition with convolutional neural network and EEG-based EFDMs," *Neuropsychologia*, vol. 146, Sep. 2020, Art. no. 107506.
- [28] Y. Luo, Q. Fu, J. Xie, Y. Qin, G. Wu, J. Liu, F. Jiang, Y. Cao, and X. Ding, "EEG-based emotion classification using spiking neural networks," *IEEE Access*, vol. 8, pp. 46007–46016, 2020.
- [29] C. Wei, L.-L. Chen, Z.-Z. Song, X.-G. Lou, and D.-D. Li, "EEG-based emotion recognition using simple recurrent units network and ensemble learning," *Biomed. Signal Process. Control*, vol. 58, Apr. 2020, Art. no. 101756.
- [30] Z. Wang, S. Hu, and H. Song, "Channel selection method for EEG emotion recognition using normalized mutual information," *IEEE Access*, vol. 7, pp. 143303–143311, 2019.
- [31] E. Yıldırım, Y. Kaya, and F. Kılıç, "A channel selection method for emotion recognition from EEG based on swarm-intelligence algorithms," *IEEE Access*, vol. 9, pp. 109889–109902, 2021.

**ALIREZA SAMAVAT** is currently a Senior Undergraduate Student in electrical engineering with the Islamic Azad University of Tehran-Central Branch, Tehran, Iran.

**EBRAHIM KHALILI** is currently a Graduate Student in electrical engineering at Trabiat Modares University, Tehran, Iran.

**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.

**MARZIEH AYATI** received the Ph.D. degree in computer science from Case Western Reserve University, Cleveland, OH, USA, in 2018. She is currently an Assistant Professor with the Computer Science Department, The University of Texas Rio Grande Valley, Edinburg, TX, USA. She is also working on development of algorithms to integrate various types of data to improve understanding of diseases. Her research is also applying machine learning data on biological data. Through her research, she gained significant experience in analyzing noisy, incomplete, high dimensional, and heterogeneous datasets, which enabled me to develop broad range of theoretical and practical skills in algorithm development and implementation, data integration, data management, data mining, machine learning, and statistical considerations.