EEG-based Emotion Recognition Under Convolutional Neural Network with Differential Entropy Feature Maps

Yifan Li[1], Chi Man Wong[1], Yudian Zheng[2], Feng Wan[1,*], Peng Un Mak[1], Sio Hang Pun[1,3], Mang I Vai[1,3]

[1]*Department of Electrical and Computer Engineering*, [2]*Department of Computer and Information Science*, [3]*State Key Laboratory of Analog and Mixed-Signal VLSI*

University of Macau, Macau S.A.R, China

fwan@um.edu.mo

***Abstract*—In recent electroencephalograph (EEG)-based emotion recognition, the differential entropy (DE) features extracted from multiple electrodes are organized as a 2D feature map for convolutional neural network (CNN) in order to utilize the information hidden in the electrodes. In this study, we attempt to investigate the influence of different feature maps on the recognition performance. Six different 2D feature maps (M1-M4: baseline feature maps without sparsity and location relationship, M5-M6: pre-defined feature maps with sparsity and location relationship) are used to organize the DE features for the traditional CNN model. Evaluation study on the DEAP dataset finds that the 2D feature map configuration exhibits statistically significant effect on the classification performance of the traditional CNN model in classifying the high/low arousal and high/low valence, respectively. However, the differences are rather limited, e.g., only 1% improvement can be resulted from selecting the optimal 2D feature map among 6 feature maps. This implies that the feature map may not be a critical issue when applying the DE features to classifying the emotion states in a CNN.**

***Index Terms*—Emotion recognition, electroencephalograph, convolutional neural network, differential entropy, feature map**

I. Introduction

More and more researchers recognize that human emotion analysis should play an important role in the human-machine interaction [1], [2]. Emotions are fundamental in human daily life [3]. Basically, the human's emotion can be recognized by various types of methods. For example, the methods based on the human's facial expressions [4] and speech [5] have been introduced for many years. However, their recognition performance could be susceptible to the human's subjective disguise. In order to evaluate the human's emotion in a more objective and reliable way, the methods based on the human's physiological signals, such as electrocardiogram (ECG) [6], [7], blood volume pressure (BVP) [6], [7], and electroencephalograph (EEG) [2], have been proposed. Among them, the human's EEG signal may be closely related to the human's emotion. In addition, EEG signal can provide the high temporal resolution and the use of EEG signal is noninvasive, convenience, and inexpensive. Therefore, EEG signal have become attractive in the emotion recognition recently [2].

In the existing EEG-based emotion recognition studies, the well-known features are the power spectrum of the EEG signals within different frequency bands, i.e., δ (1-4 Hz), θ (4-8 Hz), α (8-13 Hz), β (13-30 Hz), and γ (>30 Hz). Specifically, δ band is associated with the unconscious mind or the depth of sleep. θ band is associated with the subconscious mind or light meditation/sleeping. α band is typically linked with relaxation and happiness. β band is related to an active state of mind, namely, is closely related to thinking process. Finally, γ band is associated with bursts of insight and high-level information processing. As the human's emotion is a complex psychological state, which may involve with the human's subjective experience, physiological status, and behavioral or expressive response, most of the EEG-based emotion recognition studies consider different frequency bands simultaneously, such as $\theta + \alpha + \beta + \gamma$ [2], [8].

To extract the EEG features within different frequency bands, several time-frequency analysis approaches, e.g., short time Fourier transform, wavelet transform, and Hilbert-Huang spectrum, are utilized to calculate their power spectrum, respectively. In [9], [10], they propose a new EEG feature, so-called differential entropy (DE), for EEG-based emotion recognition and show that the corresponding performance is better than the ones using the traditional power spectrum. In particular, the DE is equivalent to logarithm power spectrum [9]. Nowadays, most of the researchers use the DE features in the EEG-based emotion recognition [11]. For example, in [12] Zheng et al. report that using the DE features can obtain the highest performance in their studies.

Convolutional neural network (CNN), which is one of the hot topics in deep learning, has been demonstrated as an effective model for image classification [13]. Due to its strong ability of feature extraction or detection, CNN has been also applied successfully in natural language processing [14], medical image classification [15], as well as emotion

This work is supported in part by the Macau Science and Technology Development Fund (FDCT) under projects 142/2014/SB, 055/2015/A2, and 088/2016/A2, the University of Macau Research Committee under MYRG projects (Y1-L2)-FST13-WF, MYRG2014-00174-FST, MYRG2016-00240-FST, MYRG2016-00157-AMSV, and MYRG2017-00207-FST.

recognition [4], [16]. In [16], Tripathi et al. show that the CNN can provide the good performance of EEG-based emotion recognition based on the statistical features, e.g., the mean, median, maximum, minimum, standard deviation, variance, range, skewness, and kurtosis values of the EEG data. As a matter of fact, the traditional CNN is known to be effective in learning the local feature from the images, and thus this CNN may not provide the optimal performance when the data is discrete and discontinuous in the spatial domain, such as the raw EEG signal recorded from multiple electrodes. For this reason, Song et al. apply the graph CNN can improve the performance by using the graph representation of the EEG features in [17]. In [18], [19], the DE features is deliberately arranged as a 2D structure according to the location information since using this 2D feature map can maintain the spatial information among electrodes in the CNN. However, how different feature maps effect the final classification performance is still unclear. We provide a preliminary study to investigate the effect of different DE feature maps on the performance.

II. Methods and Materials

In this study, we attempt to explore the effect of the CNN with different DE feature maps on the recognition performance.

A. EEG Dataset

DEAP dataset [20], which was widely used in many previous EEG-based emotion recognition researches, is applied for the proposed investigation. The DEAP dataset includes 32 subjects' physiological data (i.e., 32-channel EEG, 4-channel EOG, 4-channel EMG, and so on) that are elicited by the affective stimuli. Each subject was presented 40 one-minute long music video (i.e., affective stimuli) while his/her physiological data was recorded. After each affective stimulus, the subject was asked to rate his/her feeling based on five aspects: valence, arousal, dominance, liking, and familiarity. For each subject there were 40-trial physiological data and 40-trial subjective scores corresponding to 40 affective stimuli. Hence, the dimension of all the recorded EEG data should be $40 \times 8064 \times 32$ (video×sample×channel). Main details of the DEAP dataset can be referred to [20].

In this work, at first each trial EEG data is filtered by four bandpass filters (or Butterworth filters) with different frequency bands (i.e., θ (4-8 Hz), α (8-13 Hz), β (13-30 Hz), and γ (>30 Hz)). The dimension of all the data should become $40 \times 8064 \times 4 \times 32$ (stimulus×sample×band×channel). Second, the bandpass filtered EEG data is segmented into 60 short periods with 128 samples (or 1 sec). Therefore, the emotion recognition is based on the segments. After segmentation, the EEG data is transformed into $40 \times 60 \times 128 \times 4 \times 32$ (stimulus×segment×sample×band×channel).

Finally, we classify the subject's emotion states into two classes according to the subjective rates on valence or arousal. Specifically, there are two difference cases. First, two classes based on valence rating are positive (i.e., high valence, valence rating>5) and negative (i.e.,low valence, valence rating≤5). Second, two classes based on arousal rating are active (i.e.,high arousal, arousal rating>5) and de-active (i.e.,low arousal, arousal rating≤5).

B. DE Feature

According to [10], the details of the DE calculation are shown as below. Given that x_{ij} is the EEG signal from the i-th channel through the j-th bandpass filter. Under normal circumstances, x_{ij} obeys the Gaussian distribution, i.e., $x_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}^2)$, the DE of x_{ij} calculated by

$$\begin{aligned} \mathrm{DE}(x_{ij}) &= -\int_{\infty}^{-\infty} c_{ij} e^{\frac{(x-\mu_{ij})^2}{2\sigma_{ij}^2}} \log\big(c_{ij} e^{\frac{(x-\mu_{ij})^2}{2\sigma_{ij}^2}}\big) dx \\ &= \frac{1}{2}\log(2\pi e \sigma_{ij}^2), \end{aligned} \tag{1}$$

where $i = 1, 2, \cdots, 32$, $j = 1, 2, 3, 4$, $c_{ij} = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}}$, μ_{ij} and σ_{ij} are the mean and standard deviation of x_{ij}. Note that four bandpass filters are corresponding to four frequency bands, i.e., θ (4-8 Hz), α (8-13 Hz), β (13-30 Hz), and γ (>30 Hz). Hence, the dimension of the DE features is $40 \times 60 \times 4 \times 32$ (stimulus×segment×band×channel). Namely, the dimension of the DE features in one specified segment, frequency band, and stimulus is 1×32.

C. 2D Feature Mapping

In the DEAP data, the 32 EEG channels are AF3, AF4, C3, C4, CP1, CP2, CP5, CP6, Cz, F3, F4, F7, F8, FC1, FC2, FC5, FC6, Fp1, Fp2, Fz, O1, O2, Oz, P3, P4, P7, P8, PO3, PO4, Pz, T7, and T8. We define that their indices are $1, 2, \cdots, 32$. Then, the corresponding 32 DE features can be arranged as a 2-D feature map by six different mappings, as shown in Fig. 1. Different colors denote their indices. Note that the feature map of M5 and M6 are referenced from two previous work [18], [19] and the other feature maps (M1–M4) can be considered as the baseline feature maps (i.e., without sparsity and location relationship). According to [18], [19], the advantage of transforming the DE features into a 2D feature map is to maintain the spatial information among electrodes for CNN. In the feature maps of M5 and M6, the zeros (or white blocks) are inserted between rows and columns, which can generate the sparse DE maps. In this way, the unrelated noise could be reduced in the following convolution operations. On the contrary, in the feature maps of M1, M2, and M3 the DE features from 32 electrodes are randomly relocated into a 6×6 matrix without considering their electrode location. Then the remaining four locations are assigned as zeros. Specially, in the feature map M4 the DE features are placed into the 6×6 matrix according to the following orders 1, 3, 6, 5, 10, 11, 15, 14, 19, 20, 24, 23, 28, 30, 31, 25, 2, 4, 7, 8, 9, 13, 12, 16, 17, 18, 22, 21, 26, 27, 29, 32. Such an order is the channel-order in the DEAP dataset after preprocessing. Therefore, the feature map M4 includes some local spatial information.

After the mapping, the dimension of the DE features for each frequency band is changed from 1×32 to either 6×6,

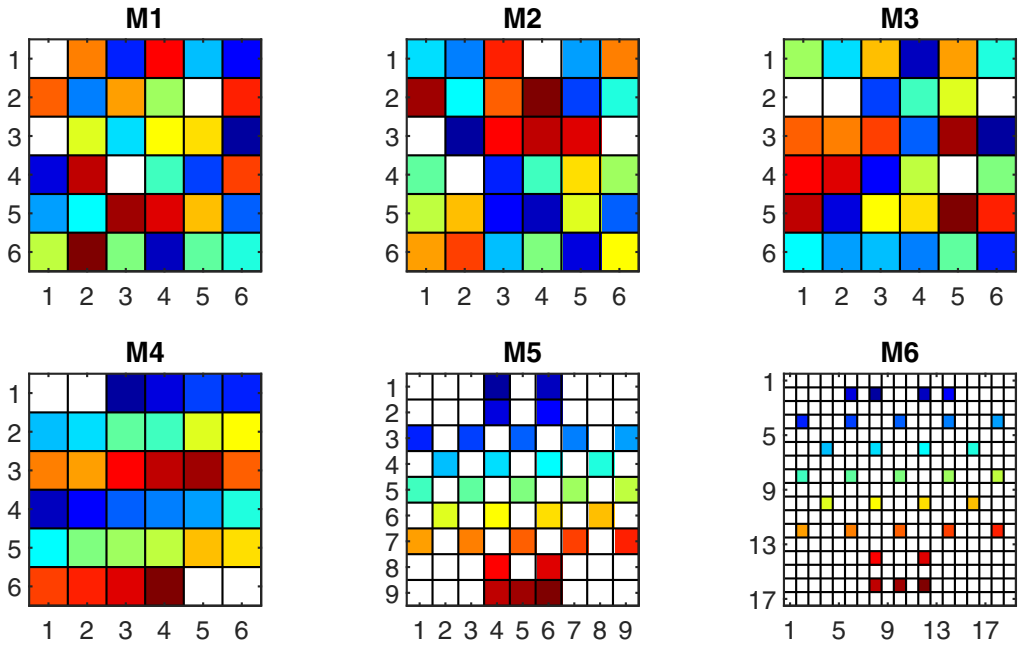

Fig. 1. Six different feature maps for the DE features. The feature maps of M1 – M3 present the randomized locations. The feature map of M5 is proposed in [18] and the feature map of M6 is proposed in [19]. Different colors denote the indices of 32 channels

9×9, or 19×17. For four different bands, the corresponding 3D feature maps are $6 \times 6 \times 4$, $9 \times 9 \times 4$, and $19 \times 17 \times 4$, respectively.

D. Convolution Neural Network

In this paper, we design a CNN with four convolutional layers to extract the features from the 3D feature maps, and a fully connected layer with dropout operation for final emotion classification. An example can be found in Fig. 2. It should be noted that this CNN model does not have the pooling layer since the size of the 2D feature maps is relatively small when compared with the images in the computer vision field. Specifically, the first convolutional layer with 64 2D feature maps is designed and the following two convolutional layers with 128 and 256 2D feature maps are designed, respectively. The fourth convolutional layer with 64 2D feature maps is designed to merge the features and reduce computational cost. In the convolution layers, the kernel size is defined as 4×4 and and the stride size is defined as 1. The RELU (Rectified Linear Unit) activation function is used after the convolution. Finally, a fully connected layer is designed to map the 64 2D feature maps into a a feature vector and the softmax function is used to output the human emotional state.

To evaluate the classification performance, we perform a 10-fold cross validation scheme. Simply speaking, we randomly select 90 % of data for training and the remaining 10 % data for testing in each fold. This procedure is repeated for 10 times. The average of the 10 times classification performance is used for evaluation.

III. Results

Table I lists all subjects' classification performance of the CNN with six different DE feature maps (or M1–M6) in classifying two different emotions, i.e., high/low arousal and high/low valence. It can be observed that all subjects' classification performance can be higher than 80% except Subject 22. In addition, the average classification performance in classifying high/low arousal and high/low valence is around 89%, which is consistent with the previous study in [18]. Interestingly, it can be found that the average accuracies under six different feature maps are similar. It seems that their differences are only 1%.

One-way ANOVA is conducted to investigate whether the factor 'feature map' effects the classification accuracy or not. Results show that the factor 'feature map' significantly effects the classification accuracy, i.e., $p < 0.05$ in arousal classification and $p < 0.001$ in valence classification. Besides, their comparison results are listed in Table II. In particular, in classifying high and low arousal, the CNN with M6 feature map can achieve the highest accuracy. The CNNs with the baseline feature maps do not have significant performance difference. On the other hand, in classifying high and low valence the classification performance is susceptible to different feature maps. The baseline feature map M1 could lead to the worst performance. The two feature maps with sparsity and location relationship could not lead to the better classification performance than the baseline feature map (M3).

Based on the above results, we know that the CNN with using the suitable 2D feature map can be improved in classifying the human's emotion state, especially in arousal classification. For example, the CNN with using the M6 feature map in clas-

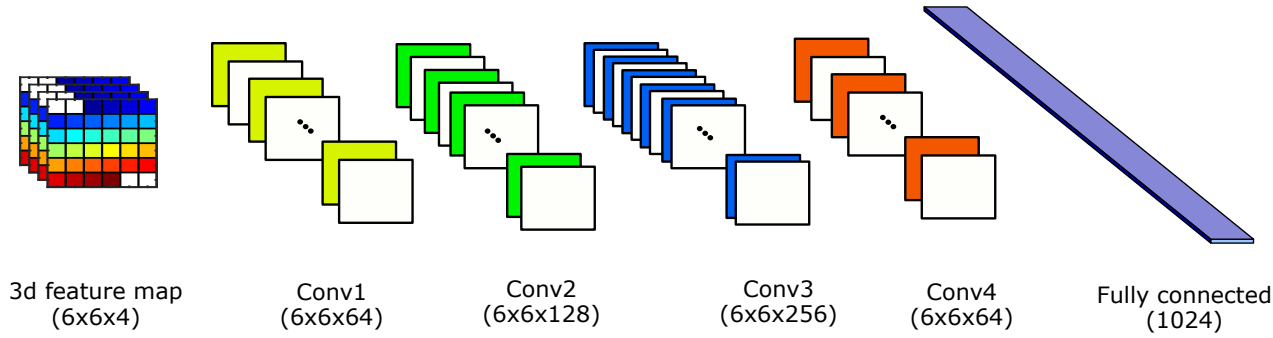

Fig. 2. An example of the block diagram of the proposed CNN with the designed $6 \times 6 \times 4$ DE feature map

TABLE I
ALL SUBJECTS' CLASSIFICATION ACCURACIES (%) UNDER DIFFERENT DE FEATURE MAPS (M1 – M6)

Subject	Category of Emotion Classification											
	Arousal						Valence					
	M1	M2	M3	M4	M5	M6	M1	M2	M3	M4	M5	M6
S1	93.42	92.47	92.36	92.33	93.23	93.20	92.86	93.32	93.31	92.50	93.48	93.68
S2	85.56	86.08	85.72	85.77	85.25	86.85	83.78	84.92	84.21	84.15	86.05	85.19
S3	92.27	92.80	91.60	92.48	92.95	92.97	85.67	86.14	87.72	86.17	87.69	87.73
S4	87.48	87.52	86.41	86.62	87.12	86.70	88.77	90.07	89.38	89.28	89.26	89.56
S5	83.31	83.33	83.87	84.55	83.05	81.80	81.70	82.44	82.15	82.14	81.61	82.07
S6	90.40	89.91	90.12	89.79	89.36	91.65	94.75	93.93	94.55	94.10	95.41	94.46
S7	92.81	93.15	93.10	92.73	93.53	92.99	92.96	92.03	93.69	93.61	93.66	93.35
S8	90.80	91.05	90.70	89.68	90.69	90.97	89.71	90.80	89.78	89.88	90.51	91.09
S9	88.07	88.28	89.50	89.38	88.84	89.30	89.38	89.53	90.57	90.26	89.73	90.35
S10	91.60	91.77	91.21	92.78	91.51	91.97	90.45	90.78	90.25	89.61	91.54	91.86
S11	85.91	87.26	84.75	85.54	86.16	85.70	82.82	83.28	84.94	82.67	83.84	82.82
S12	91.47	90.90	91.75	91.84	92.18	92.44	85.42	86.90	85.89	86.16	87.24	86.93
S13	95.57	94.76	94.29	94.36	94.77	94.91	93.05	93.03	92.15	93.21	92.96	93.82
S14	90.05	90.27	91.04	89.75	90.13	91.45	87.71	87.76	88.41	87.68	88.78	88.87
S15	91.43	91.72	90.52	91.85	91.91	91.21	93.52	94.50	94.10	94.06	94.05	94.81
S16	92.87	93.67	92.71	92.57	93.51	94.46	94.04	94.59	94.25	94.82	94.42	94.92
S17	87.90	89.21	88.42	89.71	88.87	88.66	87.41	87.16	87.74	87.59	88.52	88.36
S18	92.29	92.90	93.48	92.30	93.15	93.86	92.90	92.06	93.34	93.06	93.20	93.50
S19	91.99	91.96	92.62	91.17	91.75	92.68	90.98	91.93	90.72	91.28	91.43	91.44
S20	92.89	91.65	92.11	91.74	92.67	92.52	88.92	90.09	88.77	89.76	90.85	89.80
S21	93.01	93.88	93.92	93.60	92.85	94.26	89.00	88.69	88.62	88.92	88.88	89.53
S22	74.14	74.96	76.60	76.22	74.22	76.56	70.74	73.51	73.05	71.41	70.75	70.16
S23	93.99	95.08	94.76	94.26	94.83	95.23	93.11	91.67	92.84	92.89	92.89	92.38
S24	91.34	91.74	92.03	92.17	92.42	92.91	87.20	86.89	87.00	88.38	88.19	88.18
S25	93.50	93.85	93.76	94.61	94.10	92.54	88.82	89.89	90.17	89.11	89.15	87.97
S26	88.10	87.56	88.06	86.08	88.03	87.23	87.34	88.71	88.57	88.30	88.69	88.64
S27	92.53	92.22	92.89	92.63	93.57	93.32	94.67	93.99	94.69	94.10	94.61	93.93
S28	88.77	88.96	89.14	89.61	88.95	88.93	88.52	89.15	90.26	89.31	89.52	89.08
S29	91.10	90.47	91.68	90.68	91.24	91.02	90.74	90.02	91.10	90.22	91.56	92.56
S30	84.01	84.74	83.47	82.99	84.19	85.85	85.29	85.67	85.59	84.72	85.49	84.92
S31	90.29	90.03	90.88	90.25	90.90	91.23	89.77	89.04	90.00	90.55	91.37	90.41
S32	86.39	87.20	86.66	86.86	87.63	87.40	86.14	85.32	86.23	85.89	86.54	87.42
Mean	89.85	90.04	90.00	89.90	90.11	90.40	88.69	88.99	89.19	88.93	89.43	89.37
S.D.	4.16	3.99	3.93	3.92	4.24	4.05	4.77	4.29	4.37	4.67	4.72	4.86

TABLE II
PAIRWISE PERFORMANCE COMPARISON BETWEEN DIFFERENT MAPS

Arousal						
P1 \ P2	M1	M2	M3	M4	M5	M6
M1	-	=	=	=	$>^{**}$	$>^{**}$
M2	=	-	=	=	=	$>^{*}$
M3	=	=	-	=	=	$>^{**}$
M4	=	=	=	-	=	$>^{**}$
M5	$<^{**}$	=	=	=	-	$>^{*}$
M6	$<^{**}$	$<^{*}$	$<^{**}$	$<^{**}$	$<^{*}$	-
Valence						
P1 \ P2	M1	M2	M3	M4	M5	M6
M1	-	$>^{*}$	$>^{***}$	$>^{**}$	$>^{***}$	$>^{***}$
M2	$<^{*}$	-	=	=	$>^{**}$	$>^{*}$
M3	$<^{***}$	=	-	$<^{*}$	=	=
M4	$<^{**}$	=	$>^{*}$	-	$>^{***}$	$>^{**}$
M5	$<^{***}$	$<^{**}$	=	$<^{***}$	-	=
M6	$<^{***}$	$<^{*}$	=	$<^{**}$	=	-

[1] $>$ (or $<$): P1 is significantly higher (or lower) than P2.
[2] $=$: No significant difference between P1 and P2.
[3] $*$, $**$, and $***$: $p < 0.05$, $p < 0.01$, and $p < 0.001$.

sifying the high/low arousal. Nevertheless, the improvement is rather limited, only 1%.

IV. DISCUSSIONS

In this study, we investigate six different DE feature maps for the traditional CNN model. Although using the suitable DE feature map the classification performance of the CNN can be improved, the corresponding improvement is limited. Our statistical results show that different DE feature maps do effect the classification accuracy. One possible reason could be explained. In this study, the size of the 2D feature map is relatively small so that the CNN has ability in extracting the meaningful information from the small feature map even though this feature map does not have the spatial information among the electrodes. Therefore, how to design the feature map for the DE features may be not an important issue in the proposed CNN model. We suspect that the existing feature map could not extract the useful spatial structure of the EEG signals.

In the future, we would study how to combine the DE feature and the graph CNN model (in [17]). Besides, the existing work could be further extended to investigate the other parameters, e.g., kernel size, stride size, the number of convolutional layers, and so on.

V. CONCLUSIONS

In this paper, we provide a preliminary study to investigate the effect of different DE feature maps on the classification performance using the CNN model. Experimental results show that the DE feature map can significantly affect the classification performance of the traditional CNN in classifying the high/low arousal and the high/low valence, respectively. However, using the optimal feature map could only lead to a limited improvement among the six different feature maps. This may imply that designing an optimal 2D feature map is not very important for the CNN to classify the emotion states.

REFERENCES

[1] C. Mühl, B. Allison, A. Nijholt, and G. Chanel, "A Survey of Affective Brain Computer Interfaces: Principles, State-of-the-art, and Challenges," *Brain-Computer Interfaces*, vol. 1, no. 2, pp. 66–84, 2014.

[2] S. M. Alarcao and M. J. Fonseca, "Emotions Recognition using EEG signals: a Survey," *IEEE Transactions on Affective Computing*, 2017.

[3] J. Panksepp, *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford university press, 2004.

[4] S. E. Kahou, X. Bouthillier, P. Lamblin, C. Gulcehre, V. Michalski, K. Konda, S. Jean, P. Froumenty, Y. Dauphin, N. Boulanger-Lewandowski *et al.*, "Emonets: Multimodal Deep Learning Approaches for Emotion Recognition in Video," *Journal on Multimodal User Interfaces*, vol. 10, no. 2, pp. 99–111, 2016.

[5] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011.

[6] M. Soleymani, F. Villaro-Dixon, T. Pun, and G. Chanel, "Toolbox for Emotional Feature Extraction from Physiological Signals (teap)," *Frontiers in ICT*, vol. 4, p. 1, 2017.

[7] J. Kim and E. André, "Emotion Recognition Based on Physiological Changes in Music Listening," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 12, pp. 2067–2083, 2008.

[8] R. Jenke, A. Peer, and M. Buss, "Feature Extraction and Selection for Emotion Recognition from EEG," *IEEE Transactions on Affective Computing*, vol. 5, no. 3, pp. 327–339, 2014.

[9] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential Entropy Feature for EEG-Based Vigilance Estimation," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2013, pp. 6627–6630.

[10] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential Entropy Feature for EEG-Based Emotion Classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013, pp. 81–84.

[11] B. García-Martínez, A. Martinez-Rodrigo, R. Alcaraz, and A. Fernández-Caballero, "A Review on Nonlinear Methods Using Electroencephalographic Recordings for Emotion Recognition," *IEEE Transactions on Affective Computing*, 2019.

[12] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying Stable Patterns over Time for Emotion Recognition from EEG," *IEEE Transactions on Affective Computing*, 2017.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[14] Y. Kim, "Convolutional Neural Networks for Sentence Classification," *arXiv preprint arXiv:1408.5882*, 2014.

[15] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[16] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset." in *Twenty-Ninth IAAI Conference*, 2017.

[17] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks," *IEEE Transactions on Affective Computing*, 2018.

[18] Y. Yang, Q. Wu, Y. Fu, and X. Chen, "Continuous Convolutional Neural Network with 3D Input for EEG-Based Emotion Recognition," in *International Conference on Neural Information Processing*. Springer, 2018, pp. 433–443.

[19] J. Li, Z. Zhang, and H. He, "Hierarchical Convolutional Neural Networks for EEG-Based Emotion Recognition," *Cognitive Computation*, pp. 1–13, 2018.

[20] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A Database for Emotion Analysis using Physiological Signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2012.