A 3D Convolutional Neural Network for Emotion Recognition based on EEG Signals

Yuxuan Zhao [1,2], Jin Yang[1], Jinlong Lin[1], Dunshan Yu[1], Xixin Cao[1,*]
[1] *School of Software and Microelectronics, Peking University, Beijing, China*
[2] *Institute of Automation, Chinese Academy of Sciences, Beijing, China*
yuxuan.zhao@ia.ac.cn, (jin.yang, linjl, cxx)@ss.pku.edu.cn, yuds@pku.edu.cn

***Abstract*—As an important field of research in Human-Machine Interactions, emotion recognition based on the electroencephalography (EEG) signals has become common research. The traditional machine learning approaches use well-designed classifiers with hand-crafted features which may be limited to domain knowledge. Motivated by the outstanding performance of deep learning approaches in recognition tasks, we proposed a 3D convolutional neural network model to extract the spatial-temporal features automatically in the EEG signals. By the pre-processing method with baseline signals and the electrode topological structure relocated, the proposed model achieves a high accuracy rate of 96.61%, 96.43% in the Two class classification task (low/high arousal, low/high valence) and 93.53% in the Four class classification task (low arousal and low valence/high arousal and low valence/low arousal and high valence/high arousal and high valence) in the DEAP dataset, and 97.52%, 96.96% in the Two class classification task and 95.86% in the Four class classification task in the AMIGOS dataset.**

***Index Terms*—Emotion Recognition, Electroencephalography (EEG), 3D Convolutional Neural Network (3D CNN), Spatio-temporal Features, Deep Learning**

I. Introduction

Emotion recognition plays a crucial role in human-machine interaction and health care. The recognition method based on physiological signals, especially the electroencephalography (EEG) signals has become a research hotspot because the signals could represent the inner emotional states and cannot be controlled subjectively compare with other signals such as facial expressions or speech.

The traditional machine learning approaches which use well-designed classifiers with hand-crafted features have been studied for many years. The most common features [1], [2] contains Time Domain Features: Event Related Potentials (ERP) [3], Statistics of Signal [4] (Power, Mean, Standard deviation, 1st difference, 2nd difference et al.), Higher Order Crossings (HOC) [4] et al; Frequency Domain Features: Power Spectra Density (PSD) [2], Higher Order Spectra (HOS) et al; Time Frequency Domain Features: Hilbert-Huang Spectrum (HHS), Magnitude Squared Coherence Estimate (MSCE) et al. A traditional approach which could achieve a high emotion recognition accuracy mostly depended on the well-designed hand-crafted features. So the new and effective feature extraction methods which based on phase space reconstruction [5] and flexible analytic wavelet transform (FAWT) [6] make a good performance in the emotion recognition task. Soroush [7] made the EEG signals reconstructed in phase space and then in angle space, then extracted features from angle variability and length variability, and used Dempster-Shafer theory for emotion recognition, finally, used ten-fold cross-validation to evaluate their model. To our best knowledge, their method achieved the best performance in traditional machine learning approaches - classification accuracy was about 90% on average classified into four classes.

Motivated by the outstanding performance of deep learning approaches in pattern recognition tasks, many researchers had used these approaches in the emotion recognition task [8], [9]. Lin [10] and Liu [11] converted the EEG signals into 2D image format and used the pre-trained deep learning model AlexNet and ResNets to extract depth level features respectively, then combine the extract hand-crafted features for classification. Mei [12] and Kwon [13] used 2D Convolutional Neural Network Model for feature extraction and classification. These 2D conventional methods ignore the spatial characteristics of EEG signals, so the spatial-temporal features extraction methods had been proposed. Salama [14] proposed a 3D Convolutional Neural Network Model for spatial-temporal features extraction and classification in EEG signals, and stated that their model is outperforming the state of the art methods. Wang [15] converted the EEG channels into 2D electrode topological plate which could include topological position information and used the 3D CNN Model for spatial-temporal features extraction and classification. They found that compared with the 2D CNN Model used unconverted data, the 3D CNN Model made a better performance. Yang [16] implemented the 2D CNN module and LSTM module extract spatial and temporal features respectively and combined the features for classification, and achieved a high accuracy rate in the emotion recognition task.

Compared with the deep learning approaches, the performance of traditional machine learning approaches is little poor which may be limited to domain knowledge. In this paper, we proposed a 3D convolutional neural network model to extract the spatial-temporal features automatically in the EEG signals. By the pre-processing method with baseline signals and the electrode topological structure relocated, the proposed model achieves a high accuracy rate in emotion recognition tasks.

The remainder of this paper is organized as follows: Section 2 overview an open dataset DEAP [17] and AMIGOS [18]

*Corresponding author

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

which are used for researchers to validate the performance of their models; In Section 3, the method of data pre-processing and the architecture of the 3D convolutional neural network model are described; In Section 4, we present the result in DEAP and AMIGOS to evaluate the proposed model and comparing it with previous studies. In Section 5, we conclude our work.

II. DATABASE

The DEAP [17] is an open dataset for researchers to validate their model. This dataset contains 32 channels EEG signals and 8 channels peripheral physiological signals which be collected when 32 participants watched 40 videos each with one-minute duration. Each trial contains 63s signals and the first 3s is the baseline signals. The baseline signals are recorded when the participant under no stimulus. After watching a minute video, the participants rated a self-assessment of arousal, valence, liking, and dominance on a scale from 1 to 9. A preprocessed version had been provided: the data was down-sampled from 512Hz to 128Hz, and a bandpass frequency filter from 4.0-45.0Hz was applied. The EEG data size of DEAP is 32(participants)x40(videos)x32(EEG channels)x8064(signals), and the 8064 signals contain 384 baseline signals.

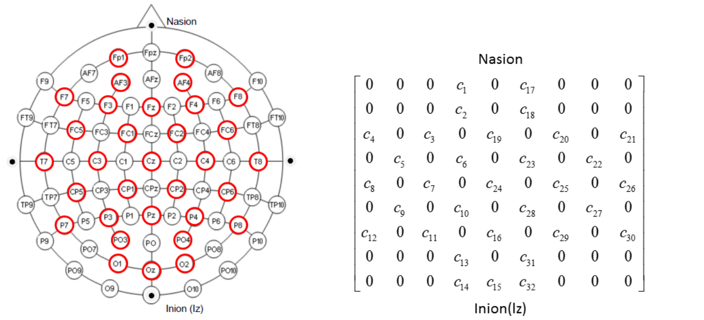
The AMIGOS [18] is a new open dataset. This dataset contains 14 channels EEG signals and 3 channels peripheral physiological signals which be collected when 40 participants watched 20 videos (16 short videos + 4 long videos). Each trial contains 5s baseline signals in first and the length of other signals depend on the duration of the video. The participants also rated a self-assessment of arousal, valence, liking, and dominance on a scale from 1 to 9 after watching the video. A preprocessed version had been provided: the data was down-sampled to 128Hz, and a bandpass frequency filter from 4.0-45.0Hz was applied.

III. MODEL

A. Pre-processing

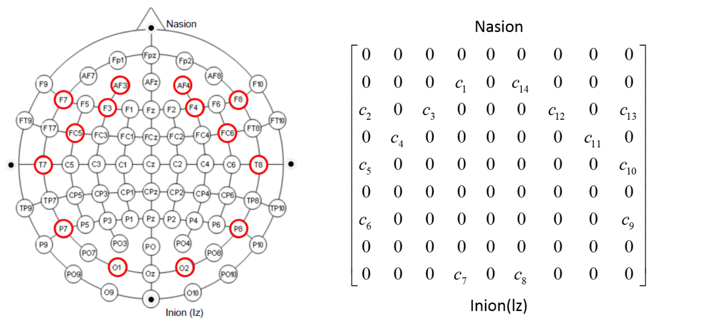
A pre-processing method with baseline signals which first elaborated by Yang [16] is an effective way to improve recognition accuracy. They reported that the pre-processing method can increase recognition accuracy by 32% approximately in the emotion recognition task. The pre-processing method contains: extract the baseline signals from all channels C and cut it in N segments with fixed length L, get N segments C x L matrixes; calculate the mean value of the baseline signals with segmented data, get the baseline signals mean value M, a C x L matrixes; cut the EEG signals which without baseline signals with length L and minus the baseline signals mean value M, get the preprocessed signals.

The international 10-20 system describes the location of scalp electrodes, and widely used in EEG experiments. The red nodes on the left side of Fig.1 and Fig.2 show the electrodes contained in the DEAP and AMIGOS dataset respectively. The raw EEG signals in DEAP and AMIGOS lost the topological position information of the electrodes. To solve this problem, the 32 electrodes used in DEAP and 17 electrodes used



(a) 10-20 System (b) 2D Electrode Topological Structure

Fig. 1. (a) The international 10-20 system describes the location of scalp electrodes, and the red nodes show the 32 electrodes used in DEAP dataset. (b) The 32 channels EEG signals are mapped into a 9x9 matrixes.



(a) 10-20 System (b) 2D Electrode Topological Structure

Fig. 2. (a) The international 10-20 system describes the location of scalp electrodes, and the red nodes show the 14 electrodes used in AMIGOS dataset. (b) The 14 channels EEG signals are mapped into a 9x9 matrixes.

in AMIGOS are relocated to the 2D electrode topological structure respectively based on the 10-20 system positioning. For each time sample point, the 32 channels EEG signals in DEAP are mapped into a 9x9 matrixes as shown in the right side of Fig.1 and the 17 channels EEG signals in AMIGOS are mapped into a 9x9 matrixes as shown in the right side of Fig.2. The unused electrodes are filled with zero. Z-score normalization is used in each transformation.

B. 3D Convolutional Neural Network Model

The architecture of the 3D convolutional neural network model contains two convolution layers, each followed by a max-pooling layer, and a fully-connected layer. A detailed illustration of the architecture is shown in Fig.3. The input size is 9x9x128, the 9x9 is the 2D electrode topological structure and the 128 is the number of the consecutive time sample point processed at once. The kernel size of the convolution layer is 3x3x4, which means the spatial-temporal features are generated based on a local topology of 3x3 and a time period of 4-time sample points. To prevent missing information of input data, the zero-padding be used in each convolutional layers. The RELU activation function is used after the convolution operation. The pooling size of a max-pooling layer is 1x1x2 which used to reduce the data size in the temporal dimension and improve the robustness of extracted features. The numbers of feature maps in the first and second convolutional layers are 32 and 64 respectively. Before passing the 64 resulting feature maps to the fully-connected layer, the output feature maps are reshaped in a vector. The fully-connected layer maps the feature maps into a final feature vector of 1024. And a dropout regularization after fully connected layers used to avoid overfitting. The N in the output layer means the numbers of the label in the task.

IV. RESULT

We use 2 classification tasks based on EEG signals to evaluate the proposed model: Two class classification task and Four class classification task. The proposed model is implemented by using Tensorflow framework [19] and deployed on NVIDIA Tesla K40c. The learning rate is set to 1E-3 with Adam Optimizer, and the keep probability of dropout operation is 0.5. The batch size for training and testing is set to 240. We use 10-fold cross-validation to evaluate the performance of our model.

A. Result in DEAP Dataset

In the process of data pre-processing for one trial signals (32x8064), the baseline signals (32x384) have been cut in 3 segments (3 32x128), and calculate the mean value of the baseline signals (1 32x128). And the EEG data without baseline signals cut in 60 segments (60 32x128) then minus the baseline signals mean value, get the preprocessed signals (32x7680). For each time sample point, the 32 channels EEG signals are mapped into a 9x9 matrixes, get the 2D electrode topological structure (7680 9x9) with Z-score normalization. Finally, the signals are cut into 60 segments with 1s length (60 9x9x128), and the 1s length was reported as the most suitable time window length in [20]. The final data size after processing is 76800 9x9x128.

Two class classification task contains two subtasks: the low/high arousal (LA/HA) classification task and the low/high valence (LV/HV) classification task which based on the arousal and valence value with the threshold of 5 respectively. Four class classification task contains 4 classes: low arousal low valence (LALV), high arousal low valence (HALV), low arousal high valence (LAHV), high arousal high valence (HAHV). The corresponding instance numbers in DEAP dataset are shown in Table I.

TABLE I
CORRESPONDING INSTANCE NUMBERS IN THE DEAP DATASET

Two Class Classification Task				
	Arousal		Valence	
Label	LA	HA	LV	HV
Threshold	≤5	>5	≤5	>5
Instances	32580	44220	34320	42480
Total	76800		76800	
Four Class Classification Task				
Label	LALV	HALV	LAHV	HAHV
Arousal	≤5	>5	≤5	>5
Valence	≤5	≤5	>5	>5
Instances	16440	17880	16140	26340
Total	76800			

TABLE II
RECOGNITION ACCURACY (%) IN DEAP DATASET

Fold ID	2 Classes		4 Classes
mean & std.dev	Arousal	Valence	
Fold 1	95.94	95.15	94.04
Fold 2	96.67	97.25	93.18
Fold 3	96.22	97.05	93.76
Fold 4	96.58	96.70	93.37
Fold 5	96.38	96.53	94.23
Fold 6	97.63	96.66	93.63
Fold 7	96.19	95.91	93.10
Fold 8	96.58	97.05	93.32
Fold 9	96.75	96.25	93.53
Fold 10	97.17	95.73	93.16
Mean	96.61	96.43	93.53
StandardDeviation	0.47	0.63	0.36

We use 10-fold cross-validation to evaluate the performance of our model. The result of each cross-validation round is shown in Table II, and the average accuracy of the 10-fold validation processes is taken as the task's final results.

For the Two class classification task, the proposed model can achieve a better accuracy of 96.61% and 96.43% in arousal and valence respectively. The comparison of our model with previous studies which mostly using 10-fold cross-validation on the DEAP database is shown in Table III. The previous studies in comparison contain two traditional machine learning approaches and seven deep learning approaches. Compare with the traditional machine learning approaches proposed by Gupta et al. [6] and Soroush et al. [5], our model achieved a better result about 16% and 9% higher respectively. The deep learning method proposed in Yang et al. [16] and Liu et al. [11] achieved the best performance to our knowledge, and our model achieves about 5% higher than these models. Compare with the 3D CNN models proposed in Wang et al. [15] and Salama et al. [14], our model uses a simpler and more efficient structure, moreover, a pre-processing method is used to relocate the electrodes into 2D topological structure and data process. The accuracy of our model achieved about 8% and 23% of rising respectively.

For the Four class classification task, the proposed model can achieve better accuracy of 93.53% in the DEAP dataset. The comparison of our model with previous studies which mostly using 10-fold cross-validation on the DEAP database as shown in Table IV. The previous studies in comparison contain three traditional machine learning approaches and three deep

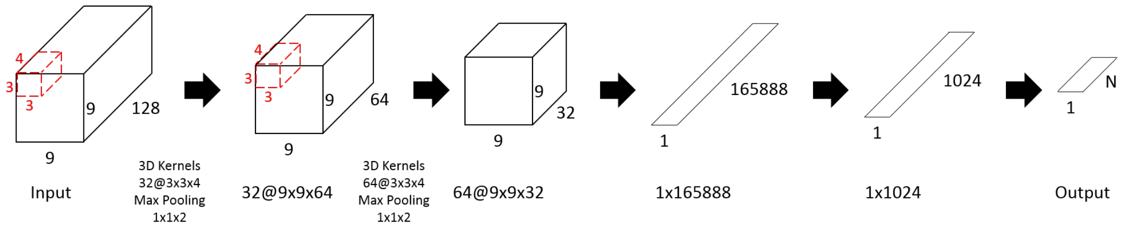


Fig. 3. 3D Convolutional Neural Network Model

TABLE III
THE COMPARISON OF OUR MODEL WITH PREVIOUS STUDIES IN THE TWO CLASS CLASSIFICATION TASK (DEAP)

Research	Method	Accuracy		Increase		Year	EEG Only
		Arousal	Valence	Arousal	Valence		
Mei et al. [12]	2D-CNN	83	83.6	13.61	12.83	2017	Yes
Lin et al. [10]	AlexNet+MF	87.3	85.5	9.31	10.93	2017	No
Wang et al. [15]	3D-CNN	73.3	72.1	23.31	24.33	2018	Yes
Salama et al. [14]	3D-CNN	88.49	87.44	8.12	8.99	2018	Yes
Kwon et al. [13]	2D-CNN+STZCR	76.56	80.46	20.05	15.97	2018	No
Liu et al. [11]	ResNet+LFCC+KNN	89.06	90.39	7.55	6.04	2018	Yes
Yang et al. [16]	CNN+LSTM	91.03	90.8	5.58	5.63	2018	Yes
Gupta et al. [6]	FAWT+RF	79.95	79.99	16.66	16.44	2019	Yes
Soroush et al. [5]	HcF+KNN+MSVM	87.42	84.59	9.19	11.84	2019	Yes
Yang et al. [21]	Multi-Column CNN	88.49	87.44	8.12	8.99	2019	Yes
Our Model	**3D-CNN**	**96.61**	**96.43**	\	\	**2020**	**Yes**

learning approaches. Compare with the deep learning method proposed in Mei et al. [12], Kwon et al. [13] and Liu et al. [11], our model achieved a better result about 20%, 20%, 7% higher respectively.

In addition, the models which proposed by Mei et al. [12], Kwon et al. [13], Liu et al. [11], Gupta et al. [6] and Soroush et al. [5] had evaluated the performance in both of the Two class and Four class classification task, and the last three models using 10-fold cross-validation method. The deep learning approach proposed by Liu et al. [11] achieved the best performance in these previous models. Compare with these models, our model achieved the best performance is shown in Fig.4.

B. Result in AMIGOS Dataset

Here we use the signals which were recorded in short videos experiment. The participant ID of 9, 12, 21, 22, 23, 24 and 33 has been removed because there are some invalid data in the preprocessed version. The data pre-processing in the AMIGOS dataset is the same as in the DEAP dataset, and the signals also are segmented with 1s length. The 14 channels EEG signals are mapped into a 9x9 matrixes (as shown in Fig.2). After processing, the final data size of EEG signals is 45474 9x9x128.

The corresponding instance numbers of the Two class classification task and Four class classification task in the AMIGOS dataset are shown in Table V.

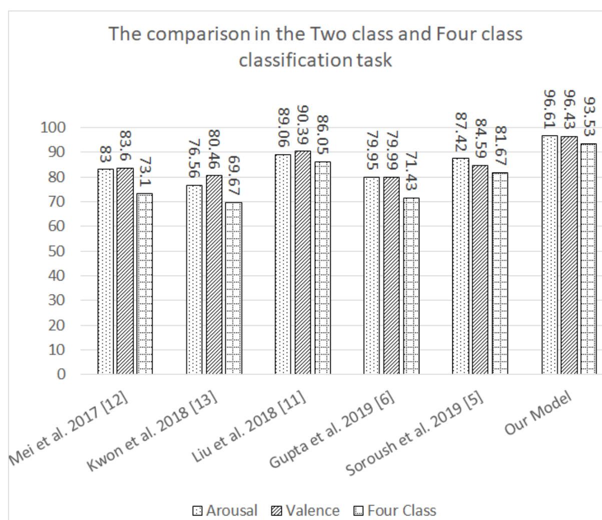


Fig. 4. The comparison of our model with previous studies in Two class and Four class classification task in DEAP dataset

The result of each cross-validation round is shown in Table VI, and the average accuracy of the 10-fold validation processes is taken as the task's final results.

For the classification in the Two class classification task, the proposed model can achieve a better accuracy of 97.52% and 96.96% in arousal and valence respectively, and 95.86% in

TABLE IV
THE COMPARISON OF OUR MODEL WITH PREVIOUS STUDIES IN THE FOUR CLASS CLASSIFICATION TASK (DEAP)

Research	Method	Accuracy	Increase	Year	EEG Only
Mei et al. [12]	2D-CNN	73.1	20.43	2017	Yes
Zheng et al. [2]	DE+GELM	69.67	23.86	2017	Yes
Kwon et al. [13]	2D-CNN+STZCR	73.43	20.1	2018	No
Liu et al. [11]	ResNet+LFCC+KNN	86.05	7.48	2018	Yes
Gupta et al. [6]	FAWT+RF	71.43	22.1	2019	Yes
Soroush et al. [5]	HcF+KNN+MSVM	81.67	11.86	2019	Yes
Our Model	**3D-CNN**	**93.53**	\	**2020**	**Yes**

TABLE V
CORRESPONDING INSTANCE NUMBERS IN THE AMIGOS DATASET

Two Class Classification Task				
	Arousal		Valence	
Label	LA	HA	LV	HV
Threshold	≤ 5	>5	≤ 5	>5
Instances	22901	22573	24622	20852
Total	45474		45474	
Four Class Classification Task				
Label	LALV	HALV	LAHV	HAHV
Arousal	≤ 5	>5	≤ 5	>5
Valence	≤ 5	≤ 5	>5	>5
Instances	12295	12327	10606	10246
Total	45474			

TABLE VI
RECOGNITION ACCURACY (%) IN AMIGOS DATASET

Fold ID	Two Class		Four Class
mean & std.dev	Arousal	Valence	
Fold 1	97.22	97.36	95.87
Fold 2	97.85	97.28	96.01
Fold 3	97.70	97.48	95.20
Fold 4	97.50	97.43	95.90
Fold 5	97.64	96.70	96.10
Fold 6	96.65	96.86	95.28
Fold 7	98.01	96.53	95.78
Fold 8	97.46	96.09	96.09
Fold 9	97.48	96.92	96.13
Fold 10	97.69	96.92	96.24
Mean	97.52	96.96	95.86
StandardDeviation	0.36	0.42	0.34

the Four class classification task. Compared with the studies of emotion recognition based on EEG signals in the DEAP dataset, the studies in the AMIGOS dataset are less [22]–[24]. The comparison of our model with previous studies is shown in Table VII. It is easy to see, our model makes a good performance in the AMIGOS dataset.

V. CONCLUSION

In this paper, we have proposed a simple and effective 3D Convolutional Neural Network Model for emotion recognition using EEG signals, and this model could be used for different tasks and datasets. The 3D CNN Model could extract spatial and temporal features simultaneously, and it made a good performance on the open dataset DEAP and AMIGOS. Compare with previous studies, the recognition accuracy of this model increases more than 5% in the DEAP dataset, and increases more than 14% in the AMIGOS dataset. In addition, the architecture of the 3D CNN model and the parameters are unchanged when evaluating the proposed model in DEAP and AMIGOS datasets. It can be proved that this model has good generality. The advantage of the traditional machine learning approaches in emotion recognition is comprehensible, in contrast, the deep learning approaches often be regarded as a black box system, and we will try to improve the explainable of our model and find the important factors in emotion recognition. Furthermore, we will use this model to challenge other tasks, such as motor imagery EEG decoding [25], [26], dementia stages classification from EEG signals [27].

TABLE VII
THE COMPARISON OF OUR MODEL WITH PREVIOUS STUDIES (AMIGOS)

Research	Method	Accuracy			Increase			Year	EEG Only
		Arousal	Valence	Four Class	Arousal	Valence	Four Class		
Miranda et al. [22]	CNN+RNN	61	59	\	36.52	37.96	\	2018	Yes
Santamaria et al. [23]	DNN-FCN	76	75	65.25	21.52	21.96	30.61	2019	No
Chao et al. [24]	LSTM-RNNs+DNN	83.3	79.4	\	14.22	17.56	\	2020	No
Our Model	**3D-CNN**	**97.52**	**96.96**	**95.95**	\	\	\	**2020**	**Yes**

REFERENCES

[1] 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.

[2] 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*, pp. 1–1, 2017.

[3] C. A. Frantzidis, B. Charalampos, C. L. Papadelis, K. Evdokimos, P. Costas, and P. D. Bamidis, "Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 589–597, 2010.

[4] A. Samara, M. L. R. Menezes, and L. Galway, "Feature extraction for emotion recognition and modelling using neurophysiological data," in *International Conference on Ubiquitous Computing and Communications and International Symposium on Cyberspace and Security*, Conference Proceedings.

[5] M. Zangeneh Soroush, K. Maghooli, S. K. Setarehdan, and A. M. Nasrabadi, "A novel eeg-based approach to classify emotions through phase space dynamics," *Signal, Image and Video Processing*, vol. 13, no. 6, pp. 1149–1156, 2019.

[6] V. Gupta, M. D. Chopda, and R. B. Pachori, "Cross-subject emotion recognition using flexible analytic wavelet transform from eeg signals," *IEEE Sensors Journal*, vol. 19, no. 6, pp. 2266–2274, 2019.

[7] M. Zangeneh Soroush, K. Maghooli, S. K. Setarehdan, and A. M. Nasrabadi, "Emotion recognition through eeg phase space dynamics and dempster-shafer theory," *Medical Hypotheses*, vol. 127, pp. 34–45, 2019.

[8] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using eeg signals: A survey," *IEEE Transactions on Affective Computing*, 2017.

[9] M. Mahmud, M. S. Kaiser, and A. Hussain, "Deep learning in mining biological data," *arXiv preprint arXiv:2003.00108*, 2020.

[10] W. Lin, C. Li, and S. Sun, *Deep Convolutional Neural Network for Emotion Recognition Using EEG and Peripheral Physiological Signal*, ser. Lecture Notes in Computer Science, 2017, book section Chapter 33, pp. 385–394.

[11] N. Liu, Y. Fang, L. Li, L. Hou, F. Yang and Y. Guo,, "Multiple feature fusion for automatic emotion recognition using eeg signals," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[12] M. Han and X. Xiangmin, "Eeg-based emotion classification using convolutional neural network," in *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*. IEEE, Conference Proceedings, pp. 130–135.

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

[14] E. S. Salama, R. A.El-Khoribi, M. E.Shoman, and M. A.Wahby, "Eeg-based emotion recognition using 3d convolutional neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 8, 2018.

[15] Y. Wang and Z. Huang, "Emotionet: A 3-d convolutional neural network for eeg-based emotion recognition," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, Conference Proceedings, pp. 1–7.

[16] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, "Emotion recognition from multi-channel eeg through parallel convolutional recurrent neural network," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, Conference Proceedings, pp. 1–7.

[17] S. Koelstra, C. Muhl, M. Soleymani, L. Jong-Seok, 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.

[18] J. A. Miranda Correa, M. K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," *IEEE Transactions on Affective Computing*, pp. 1–1, 2018.

[19] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard, "Tensorflow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, Conference Proceedings, pp. 265–283.

[20] 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, 2014.

[21] H. Yang, J. Han, and K. Min, "A multi-column cnn model for emotion recognition from eeg signals," *Sensors (Basel)*, vol. 19, no. 21, 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/31683608

[22] J. A. Miranda-Correa and I. Patras, "A multi-task cascaded network for prediction of affect, personality, mood and social context using eeg signals," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, Conference Proceedings, pp. 373–380.

[23] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-Gonzalez, E. Abdulhay, and N. Arunkumar, "Using deep convolutional neural network for emotion detection on a physiological signals dataset (amigos)," *IEEE Access*, vol. 7, pp. 57–67, 2019.

[24] C. Li, Z. Bao, L. Li, and Z. Zhao, "Exploring temporal representations by leveraging attention-based bidirectional lstm-rnns for multi-modal emotion recognition," *Information Processing & Management*, vol. 57, no. 3, p. 102185, 2020.

[25] Y. Li, M. Lei, X. Zhang, W. Cui, Y. Guo, T.-W. Huang, and H.-L. Wei, "Boosted convolutional neural networks for motor imagery eeg decoding with multiwavelet-based time-frequency conditional granger causality analysis," *arXiv preprint arXiv:1810.10353*, 2018.

[26] N. Mammone, C. Ieracitano, and F. C. Morabito, "A deep cnn approach to decode motor preparation of upper limbs from time-frequency maps of eeg signals at source level," *Neural Netw*, vol. 124, pp. 357–372, 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/32045838

[27] C. Ieracitano, N. Mammone, A. Bramanti, A. Hussain, and F. C. Morabito, "A convolutional neural network approach for classification of dementia stages based on 2d-spectral representation of eeg recordings," *Neurocomputing*, vol. 323, pp. 96–107, 2019.