

# EEG-based Deep Emotional Diagnosis: A Comparative Study

---

Geyi Liu<sup>1</sup>, Zhaonian Zhang<sup>1</sup>, Richard Jiang<sup>1</sup>, Danny Crookes<sup>2</sup> and Paul Chazot<sup>3</sup>

<sup>1</sup> School of Computing and Communication, Lancaster University, Lancaster, UK

<sup>2</sup> ECIT Institute, Queen's University Belfast, Belfast, UK

<sup>3</sup> School of Bioscience, Durham University, Durham, UK

## 1. Introduction

Emotion plays an important role in human decision, interaction and cognition [1]. Due to the rapid development of society and the accelerated pace of life, people often feel pressure and anxiety. The persistence of this situation may lead to a variety of health problems or depression, thereby affecting people's daily life and self-development. Therefore, emotion recognition is gradually becoming a practical topic of researchers. Nowadays, emotion recognition is used in many fields, such as text, speech, expression, and posture. But these methods are subjective and cannot guarantee the authenticity of emotion. Physiological and psychological studies show that the changes of physiological signals are often closer to people's real emotions than facial expression, posture or voice [2]. Electroencephalogram (EEG) reflects all kinds of electrical activities and functional states of the brain, and contains the effective information of human emotional state. Using EEG signals for emotion detection has more advantages than other methods [3]. EEG based on emotion recognition will provide an accurate emotion in many fields.

In this paper, the performance of three deep learning technologies: CNN, LSTM, and DNN in binary classification is compared. The relevant parameters of the model are given and further comparative analysis is carried out. In addition, ensemble learning method, based on major voting, is proposed to further improve the accuracy of binary classification task. The median value 5 was used as the threshold to divide the rating values into two categories. Russell proposed a 2D model of emotion classification. In his model, valence is represented from negative to positive, whereas arousal is represented from not excited to aroused or dull to intense [15]. His model is commonly used previous studies. Arousal and valence are the target dimensions in this paper.

DEAP data set was used to conduct an emotion binary classification work to validate the efficacy of these deep learning techniques. The main purpose of the dataset is to establish a music video recommendation system based on the user's emotional state. The skin electrical response, plethysmograph, skin temperature, respiratory rate, EMG, EEG, and other physiological signals of 32 subjects were recorded [4, 5]. In this study, we mainly focused on the EEG signal in the dataset. The frequency domain features of the EEG signal were extracted and inputted into the model for binary classification work.

The rest of this paper is divided into six parts. The second section briefly reviews related works. The third section introduces the DEAP database and features of EEG signal used in this study. The fourth section introduces the basic principles of CNN, DNN, and LSTM. The fifth section presents the results of this study. Section sixth constitutes the conclusion.

## 2. Related Works

In recent years, many emotion recognition methods based on EEG signals have been proposed. Liu *et al.* [6] proposed a support vector machine (SVM) method for data classification using time-domain and frequency-domain features, where classification results are 70.3% and 72.6%. A supervised learning algorithm based on perceptron convergence algorithm and Bayesian weighted log a posteriori function was introduced by Yoon and Chung [7], and the accuracy of valence and arousal was 70.9% and 70.1%. Alhagry *et al.* [8] proposed an LSTM recurrent neural network model. The average accuracy of arousal and valence were 85.65% and 85.45% respectively. Zhan *et al.* [9] designed a shallow depth-wise parallel convolutional neural network (CNN). In DEAP dataset, the prediction accuracy of arousal and valence are 84.07% and 82.95%. Xing *et al.* [10] established a stacked automatic encoder (SAE) to decompose EEG signals and classify them through the LSTM model. The accuracy of valence and arousal observation were 81.1% and 74.38%. Bagzir *et al.* [11] established a valence/arousal emotion recognition system. Discrete wavelet transform was used to transform EEG signals, which are decomposed into alpha, beta, theta, and gamma bands, and the spectral features of each band are extracted. They proposed SVM, k nearest neighbor (KNN) and artificial neural network (ANN) for emotion recognition. The best accuracy on valence and arousal were 91.1% and 91.3%. A 3D convolutional neural network was invoked by Salama *et al.* for emotion recognition of multi-channel EEG signal [12]. The obtained accuracy of valence and arousal dimension were 87.44% and 88.49%. Chen *et al.* [13] proposed a new emotion-recognition method based on the ensemble learning method, AdaBoost. The best average accuracy could reach up to 88.70%.

In general, traditional machine learning, deep learning and ensemble learning methods have been used for emotion recognition of EEG signals. However, there is still a problem that EEG signal information cannot be well reflected in feature extraction. At the same time, the methods of feature extraction are also different, causing some difficulties in the comparison of model performance.

### 3. DEAP Dataset

In this paper, the DEAP database is applied for emotion recognition task. This database recorded the physiological signals of 32 subjects watching 40 minutes of music video (1 minute for each music video) and the psychological scales of valence, arousal, dominance, and liking of the subjects [14]. Each subject had 40 channels records. These channels consisted of 32 EEG channels and remaining eight channels were peripheral physiological signals, and the sampling frequency of these pre-processed signals was 128 Hz. The structure of the DEAP dataset is shown in Table 1.

Table 1: DEAP dataset description

Array Name	Array Shape	Array Contents
Data	$40 \times 40 \times 8064$	(Video/trial) $\times$ channels $\times$ data
Labels	$40 \times 4$	(Video/trial) $\times$ label (valence, arousal, dominance, liking)

In this paper, all EEG channels are used. These 32 channels are denoted as Fp1, AF3, F3, F7, FC5, FC1, C3, T7, CP5, CP1, P3, P7, PO3, O1, Oz, Pz, Fp2, AF4, Fz, F4, F8, FC6, FC2, Cz, C4, T8, CP6, CP2, P4, P8, PO4 and O<sub>2</sub> respectively [14]. Figure 1 shows the placement position of the electrodes.

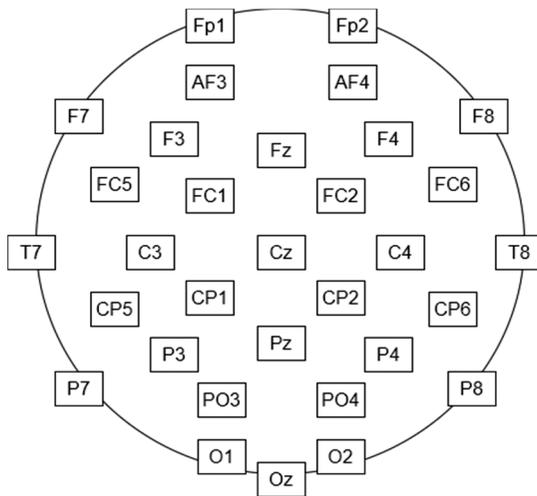


Fig. 1: Electrodes placement for EEG recording (32 electrodes)

People can observe brain waves in different frequency ranges through EEG and brain waves differ, depending on the information or instructions they convey [16]. For instance, high frequency brain waves can be observed when people are happy or excited. On the other hand, low frequency brain waves can be observed when people feel lazy or bored. According to the frequency range, EEG can be divided into five types as shown in Table 2 [17]. Gamma wave is closely related to learning,

memory and information processing. Beta waves are associated with arousal. Too much stress and anxiety will be evident in the beta band. Alpha is the frequency range between beta and theta. This band is related to the state when we relax. Theta is involved in sleep. Too much  $\theta$  activity may predispose people to depression. Delta is not our focus in this study because it is associated with the deepest levels of sleep and relaxation.

**Table 2:** EEG frequency bands

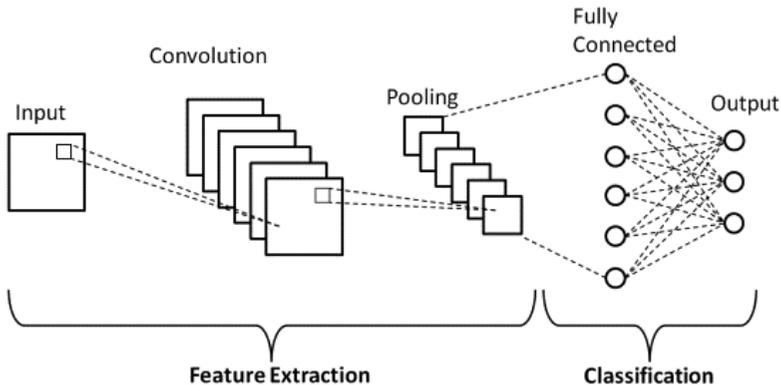
Name	Frequency Range
Gamma	40 Hz to 100 Hz
Beta	12 Hz to 40 Hz
Alpha	8 Hz to 12 Hz
Theta	4 Hz to 8 Hz
Delta	0 Hz to 4 Hz

In this paper, Fast Fourier Transform (FFT) was used to extract the specified frequency band features. We mainly extracted the information of five frequency bands, which were [4-8 Hz]: theta band, [8-12 Hz]: alpha band, [12-16 Hz]: low beta band, [16-25 Hz]: high beta band and [25-40 Hz].

## 4. Models

### 4.1 Convolutional Neural Network (CNN)

In recent years, convolutional neural network (CNN) has been widely used in satellite image analysis [18], medical image analysis [19], behavior analysis [20], and facial modelling [21, 22]. A CNN consists of an input and an output layer, as well as multiple hidden layers. Typically, hidden layers consist of three main feature layers: convolution layer, pool layer and fully connected layer [23]. The typical structure of CNN is shown in Fig. 2.



**Fig. 2:** Typical structure of convolutional neural network.

Convolution layer is the core of CNN operation. Convolution layer is composed of several feature maps and applies a convolution operation to the input, transferring the result to the next layer. And the Rectified Linear Unit (ReLU) aims to apply an activation function, such as sigmoid to the output produced by the previous layer [24]. The pooling layer aims to down sample in order to reduce the complexity for further layers and increase the robustness of feature extraction. After all of the features are generated by the neural network, they are passed to the fully connected Softmax layer. The function of fully connected layer in CNN is the same as that in standard ANN. It will produce probability distribution of all labels used for classification. In order to improve network performance, ReLU is usually used between these layers. Compared with traditional classifiers, CNNs have obvious advantages in analyzing high-dimensional data [23]. The convolution layers in CNN can control and reduce the number of parameters through parameter sharing scheme. And pooling layer could progressively reduce the spatial size of the representation and the number of parameters in the network, before solving the problem of overfitting to a certain extent.

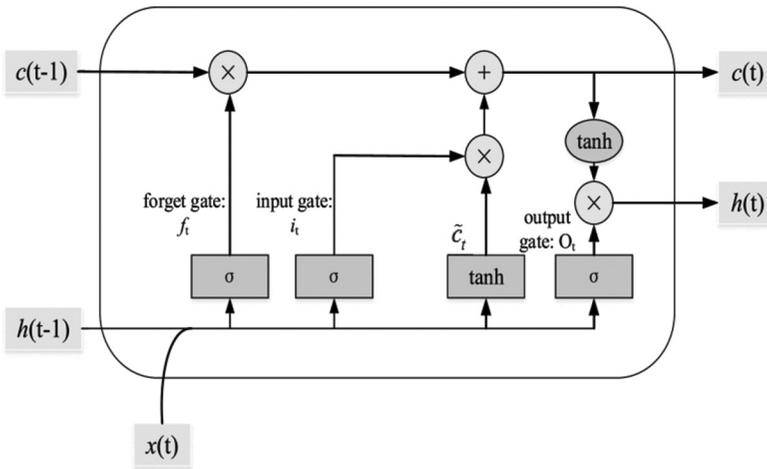


Fig. 3: LSTM block structure [25]

### 4.2 Long Short-Term Memory (LSTM)

The typical structure of an LSTM block is shown in Fig. 3 [25]. Compared with RNN, LSTM network performs better in dealing with explosion of long-term dependences and gradient vanishing. LSTM block includes three gate controllers, which are input gate, forget gate as well as output gate. And the input of LSTM block has its previous cell state  $c_{(t-1)}$ , previous hidden state  $H_{(t-1)}$ , and current input vector  $x_{(t)}$ .

The forget gate decides what information should be discarded or retained. Information from the previously hidden state and the current input are passed through the sigmoid function. The output value of sigmoid function is between 0 and 1. The closer the output value is to 0, the easier it will be forgotten, and the closer it is to 1, the more it will be retained.

The input gate is used to update the cell status. First, the previous hidden state and the current input will be passed to the sigmoid function. At the same time, the hidden state and the current input will be passed to the tanh function, compressing their values between  $-1$  and  $1$  to help adjust the network. Finally, the tanh function output will be multiplied by the sigmoid function output, and sigmoid function output will determine which information is important.

The output gate determines what the next hidden state is. First, the previous hidden state and the current input are passed to the sigmoid function. The new cell state is then passed to the tanh function. And tanh output will be multiplied by the sigmoid output to determine the information to be carried in the hidden state. Finally, the new cell state and the new hidden state will be passed to the next time step.

The relevant calculation formula for three gates are as follows:  $b$  is the corresponding bias terms;  $\sigma$  is the nonlinear activation function like sigmoid function;  $W$  is the corresponding weights.

$$\begin{aligned} f_{(t)} &= \sigma(W_{fx}x_{(t)} + W_{fh}h_{(t-1)} + b_f) \\ i_{(t)} &= \sigma(W_{ix}x_{(t)} + W_{ih}h_{(t-1)} + b_i) \\ o_{(t)} &= \sigma(W_{ox}x_{(t)} + W_{oh}h_{(t-1)} + b_o) \end{aligned}$$

The memory cell  $C_{(t)}$ , intermediate state  $\tilde{C}_{(t)}$  and hidden state of LSTM  $h_{(t)}$  are updated as

$$\begin{aligned} \tilde{C}_{(t)} &= \tanh(W_{\tilde{c}}x_{(t)} + W_{\tilde{c}h}h_{(t-1)} + b_{\tilde{c}}) \\ C_{(t)} &= f_{(t)} \cdot C_{(t-1)} + i_{(t)} \cdot \tilde{C}_{(t)} \\ h_{(t)} &= O_{(t)} \cdot \tanh(C_{(t)}) \end{aligned}$$

“ $\cdot$ ” represents the point-wise multiplication of two vectors.

### 4.3 Deep Neural Network (DNN)

Deep neural network is a type of ANN which includes input layer, output layer, and several hidden layers [26]. The nodes of each layer are fully connected with the nodes of the next layer and all connections have their weighted values. At the same time, multiple hidden layers enhance the expression ability and complexity of the model. The structure of DNN model is shown in Fig. 4.

## 5. Methods

### 5.1 Pre-processing

As mentioned before, Fast Fourier Transform (FFT) was used to extract the specified frequency band features. There are five bands which were extracted: [4-8 Hz], [8-12 Hz], [12-16 Hz], [16-25 Hz] and [25-40 Hz]. The extracted features will be used as the input of the neural network model. In terms of data partition, the ratio of training set to test set is 8:2. In the division of classification labels, we set the label value less than 5 as negative (0), and the label value greater than 5 as positive (1).

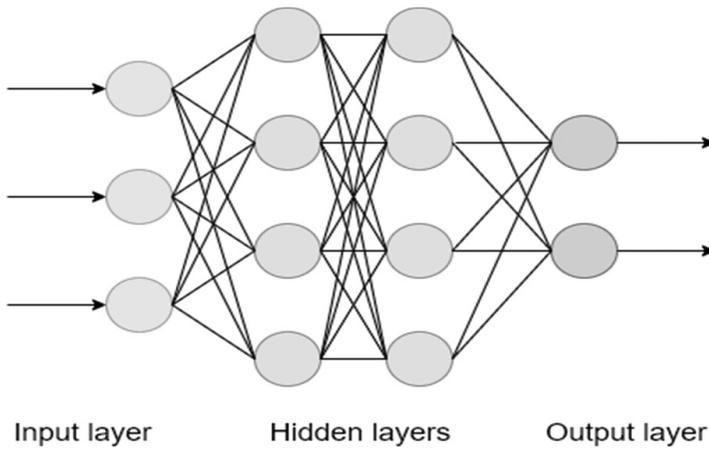


Fig. 4: Description of deep neural network (DNN) model

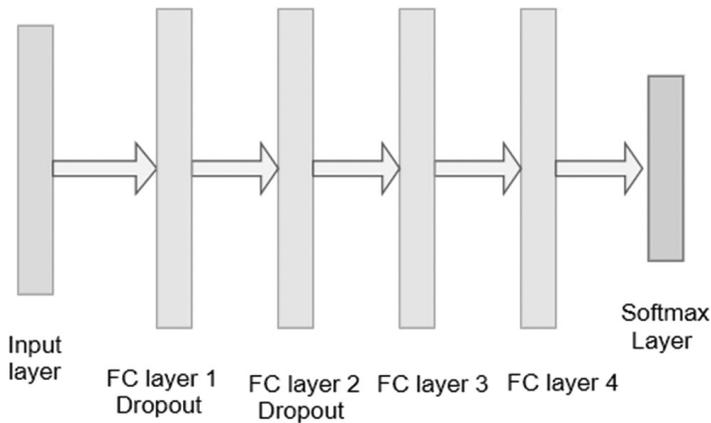


Fig. 5: Architecture of deep neural network (DNN) model

### 5.2 Model Architecture

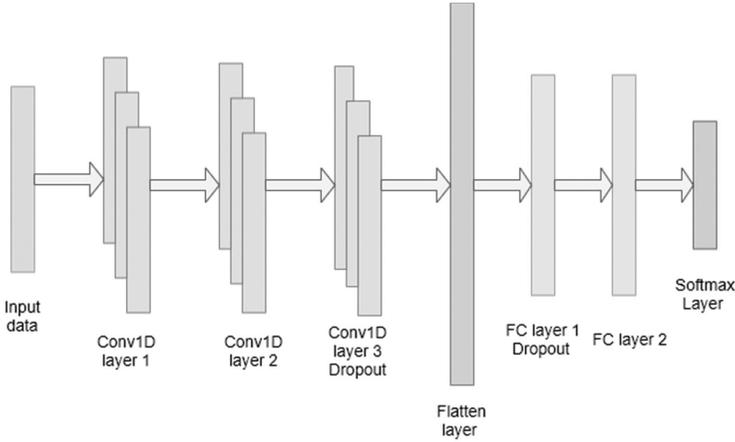
First, the architecture of DNN will be introduced. The DNN model uses four fully connected, dense neural layers. And the output of one layer serves as the input for the next layer. The architecture of Deep Neural Network (DNN) model is shown in Fig. 5. Dropout is applied in the first two full connection layers with a probability of 0.2. Because dropout technology is added, it is very important to use high epoch. In the three models this time, more than 500 epochs are used.

In the selection of activation function, rectified linear units (ReLU) are applied to all hidden layers, and Softmax is used for the output layer. The neurons number in each layer are shown in Table 3.

The proposed CNN model is composed of three convolution layers, one Flatten layer, two fully connected (FC) layers, and an output layer. The specific model structure is shown in Fig. 6.

**Table 3:** Neurons number in each layer of DNN model

Layer	Neurons Number
Input layer	70
FC layer 1	512
FC layer 2	1024
FC layer 3	1024
FC layer 4	256



**Fig. 6:** Architecture of Convolutional Neural Network (CNN) model

In order to avoid over-fitting and improve network performance, batch normalization is added after the first two convolution layers, and dropout layer is added after the third convolution layer and the first fully connected (FC) layer. The drop probability of convolution layer is 0.5. And the fully connected (FC) layer has a dropout probability of 0.2. Furthermore, we deploy rectified linear units (ReLU) and Softmax as the non-linear activation function. In the selection of optimizer, we use the most common method, Adam. The kernels of the three convolution layers are 8, 3 and 3, respectively. The neurons of each layer are shown in Table 4.

**Table 4:** Neurons number in each layer of CNN model

Layer	Neurons Number
Convolution 1D layer 1	256
Convolution 1D layer 2	128
Convolution 1D layer 3	64
FC layer 1	64
FC layer 2	16

The third model is CNN-LSTM model. Many previous studies have achieved good results in CNN model and LSTM model. So this time we try to combine the two techniques to explore the feasibility of CNN-LSTM model in the classification task. The specific layer structure is shown in Fig. 7. Convolution layer is used to extract feature information, and batch normalization technology is added in the third convolution layer to avoid over-fitting. Compared with the previous CNN model, LSTM replaces the full connection layer, and the output of the convolution layer is then inputted into the LSTM layer.

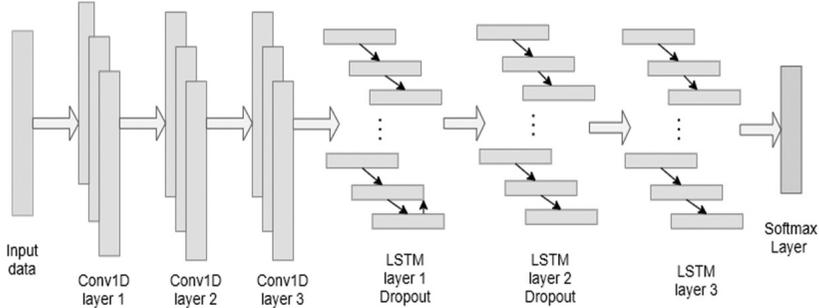


Fig. 7: Architecture of CNN-LSTM model

Table 5 records the neuron numbers in each layer of CNN-LSTM model. As in the previous models, the activation functions are ReLu and Softmax. The dropout probability of LSTM layer is set to 0.3. The kernel of convolution layer 1 is 15, the kernel of convolution layer is 2 as well as convolution layer 3 is 3. And the strides of convolution layer 1 and convolution layer 3 are all 2.

Table 5: Neurons number in each layer of CNN-LSTM model

Layer	Neurons Number
Convolution 1D layer 1	256
Convolution 1D layer 2	128
Convolution 1D layer 3	64
LSTM layer 1	512
LSTM layer 2	256
LSTM layer 3	128

### 6. Result

In this section, we compare the performance of the three models mentioned above, and compare and analyze the different settings of some hyperparameters of the models. Figure 8 shows the accuracy of the three models on the dimension of valence and arousal. From the figure, we can see that the average accuracy of the three models has reached more than 90%. Among them, the average accuracy of CNN model can reach about 95% on arousal dimension and about 94% on valence dimension.

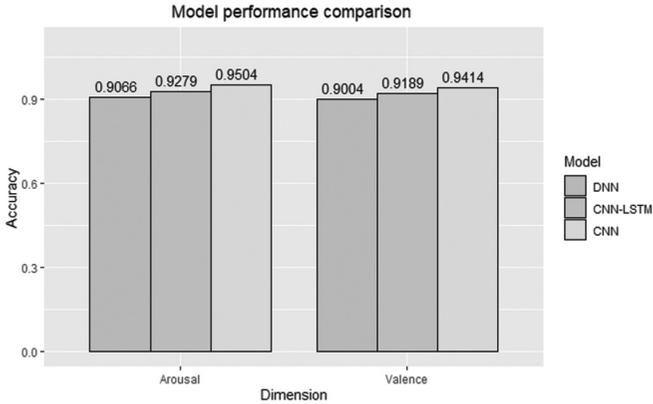


Fig. 8: Model comparison

Compared with some methods mentioned in previous articles, our method has made some progress, especially on CNN model. The comparison between the proposed method and some existing methods using DEAP data base is shown in Table 6.

Table 6: Comparison of the proposed method with some existing methods

Reference	Feature Domain	Class	Classifier	Valence (%)	Arousal (%)
Hwang <i>et al.</i> [27]	Frequency	Binary	SVM	64.90	64.90
Zhan <i>et al.</i> [9]	Frequency	Binary	CNN	82.95	84.07
Alhagry, S. [8]	Time	Binary	LSTM	85.65	85.45
Our model	Frequency	Binary	CNN-LSTM	92.79	91.89
Our model	Frequency	Binary	CNN	95.04	94.14

## 7. Conclusion

In conclusion, we examined three models and compared their performances on EEG-based mental diagnosis. In our experiments, it shows that our CNN model has a better performance than the other two models. From the results, the deep learning techniques exhibit high practical value and research potential. In our work, we used some specific frequency bands in feature extraction, where the performance of frequency bands is critical in the model. Further optimization of hyperparameters may help improve the performance, which can also be part of our future work.

## References

1. Sreeshakthy, M. and Preethi, J. (2016). Classification of human emotion from Deap EEG signal using hybrid improved neural networks with cuckoo search, *BRAIN: Broad Research in Artificial Intelligence and Neuroscience*, 6(3-4): 60-73.

2. Chen, Y., Cui, Y. and Wang, S. (2017). Review of emotion recognition based on physiological signals, *System Simulation Technology*, 13(1): 1-5.
3. Li M., Chai, Q., Kaixiang, T., Wahab, A. and Abut, H. (2009) EEG Emotion recognition system. In: Takeda, K., Erdogan, H., Hansen, J.H.L. and Abut, H. (Eds.). *In-Vehicle Corpus and Signal Processing for Driver Behavior*. Springer, Boston, MA.
4. Kumar, D.K. and Nataraj, J.L. (2019). Analysis of EEG-based emotion detection of DEAP and SEED-IV databases using SVM, *SSRN Electron*, 15: 207-211.
5. Koelstra, S., Muhl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., ... and Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals, *IEEE Transactions on Affective Computing*, 3(1): 18-31.
6. Liu, J., Meng, H., Li, M., Zhang, F., Qin, R. and Nandi, A.K. (2018). Emotion detection from EEG recordings based on supervised and unsupervised dimension reduction, *Concurrency and Computation: Practice and Experience*, 30(23): e4446.
7. Yoon, H.J. and Chung, S.Y. (2013). EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm, *Computers in Biology and Medicine*, 43(12): 2230-2237.
8. Alhagry, S., Fahmy, A.A. and El-Khoribi, R.A. (2017). Emotion recognition based on EEG using LSTM recurrent neural network, *Emotion*, 8(10): 355-358.
9. Zhan, Y., Vai, M.I., Barma, S., Pun, S.H., Li, J.W. and Mak, P.U. (2019, June). A computation resource-friendly convolutional neural-network engine for EEG-based emotion recognition. In: *2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pp. 1-6, IEEE.
10. Xing, X., Li, Z., Xu, T., Shu, L., Hu, B. and Xu, X. (2019). SAE+ LSTM: A new framework for emotion recognition from multi-channel EEG, *Frontiers in Neuroinformatics*, 13: 37.
11. Bazgir, O., Mohammadi, Z. and Habibi, S.A.H. (2018, November). Emotion recognition with machine learning using EEG signals. In: *2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME)*, pp. 1-5, IEEE.
12. Salama, E.S., El-Khoribi, R.A., Shoman, M.E. and Shalaby, M.A.W. (2018). EEG-based emotion recognition using 3D convolutional neural networks, *Int. J. Adv. Comput. Sci. Appl.*, 9(8): 329-337.
13. Chen, Y., Chang, R. and Guo, J. (2021). Emotion recognition of EEG signals based on the ensemble learning method: AdaBoost, *Mathematical Problems in Engineering*, 2021: Article ID 8896062.
14. Dabas, H., Sethi, C., Dua, C., Dalawat, M. and Sethia, D. (2018, December). Emotion classification using EEG signals. In: *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*, pp. 380-384.
15. Russell, J. A. and Snodgrass, J. (1987). Emotion and the environment. *Handb. Environ. Psychol.*, 1: 245-281.
16. Mühl, C., Allison, B., Nijholt, A. and Chanel, G. (2014). A survey of affective brain computer interfaces: Principles, state-of-the-art, and challenges, *Brain-Computer Interfaces*, 1(2): 66-84.
17. Five types of brain waves frequencies; <https://mentalhealthdaily.com/2014/04/15/5-types-of-brain-waves-frequencies-gamma-beta-alpha-theta-delta/> [accessed on 11 July, 2018]
18. Chiang, C., Barnes, C., Angelov, P. and Jiang, R. (2020). Deep Learning based Automated Forest Health Diagnosis from Aerial Images. *IEEE Access*.
19. Storey, G., Jiang, R., Bouridane, A. and Li, C.T. (2019). 3DPalsyNet: A Facial Palsy Grading and Motion Recognition Framework using Fully 3D Convolutional Neural Networks. *IEEE Access*.

20. Jiang, Z., Chazot, P.L., Celebi, M.E., Crookes, D. and Jiang, R. (2019). Social Behavioral Phenotyping of *Drosophila* with a 2D-3D Hybrid CNN Framework. IEEE Access.
21. Storey, G., Bouridane, A. and Jiang, R. (2018). Integrated Deep Model for Face Detection and Landmark Localisation from 'in the wild' Images. IEEE Access.
22. Storey, G., Jiang, R. and Bouridane, A. (2017). Role for 2D image generated 3D face models in the rehabilitation of facial palsy. *IET Healthcare Technology Letters*.
23. O'Shea, K. and Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458
24. Albawi, S., Mohammed, T.A. and Al-Zawi, S. (2017, August). Understanding of a Convolutional Neural Network. In: *2017 International Conference on Engineering and Technology (ICET)*, pp. 1-6, IEEE.
25. Yuan, X., Li, L. and Wang, Y. (2019). Nonlinear dynamic soft sensor modeling with supervised long short-term memory network, *IEEE Transactions on Industrial Informatics*, 16(5): 3168-3176.
26. Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H. and Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals, *Computers in Biology and Medicine*, 100: 270-278.
27. Hwang, S., Ki, M., Hong, K. and Byun, H. (2020, February). Subject-independent EEG-based emotion recognition using adversarial learning. In: *2020 8th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1-4, IEEE.