Merged LSTM Model for emotion classification using EEG signals

Anumit Garg
Department of Electronics and Communication Engineering
Dr. B.R. Ambedkar National Institute of Technology
Jalandhar, India
anumitgarg@gmail.com

Ashna Kapoor
Department of Electronics and Communication Engineering
Dr. B.R. Ambedkar National Institute of Technology
Jalandhar, India
ashnakapoor07@gmail.com

Anterpreet Kaur Bedi
Department of Electronics and Communication Engineering
Dr. B.R. Ambedkar National Institute of Technology
Jalandhar, India
anterpreetbedi27@gmail.com

Ramesh K. Sunkaria
Department of Electronics and Communication Engineering
Dr. B.R. Ambedkar National Institute of Technology
Jalandhar, India
sunkariark@gmail.com

Abstract— **The applicability of contemporary deep learning techniques have seen considerable improvements in the field of biomedical signal analysis. Emotion analysis using EEG signals is one such problem that has been studied and worked upon extensively in recent times. In this paper we have proposed a novel methodology to classify emotions using signal processing techniques such as wavelet transform and statistical measures for feature extraction and dimensionality reduction followed by developing state of the art neural architecture for the classification task. A merged LSTM model has been proposed for binary classification of emotions. The model's applicability and accuracy has been validated using DEAP dataset which is the benchmark dataset for emotion recognition.**

Keywords—deep recurrent neural networks (RNN), long short term memory (LSTM), discrete wavelet transform, statistics, dimensionality reduction, feature extraction, merged LSTM, DEAP dataset.

I. INTRODUCTION

Emotions are non verbal cues that define the mental state of an individual associated with his/ her thoughts, feelings, external stimuli etc. According to Charles Darwin emotions are adaptations gained through evolution that allow both humans and animals to maximize their chances of survival and reproduction [1]. Emotions play an instrumental role in decision making capability of an individual [2]. With maturation of technology, availability of huge amount of dataset and deeper understanding of human psychology, the field of cognitive analytics [3] has seen many breakthroughs in recent years.

Study of emotions based on text, speech, facial expressions, body pose, gesture have contributed a lot in this field [4] [5] [6] [7] [8] but analysis based on physiological signals is an emerging and more promising domain of research as these signals are spontaneous and highly involuntary in nature. Physiological signals are generated by the body in response to the functioning of various physiological systems. Hence these signals hold information which can be extracted to find out the state of these physiological systems. Most commonly used physiological signals for the purpose of emotion analysis include [9] [10] skin temperature variation, Electro Dermal Activity (EDA), Electrocardiogram (ECG), Electroencephalogram (EEG), etc.

In the present work, we use Electroencephalogram (EEG) signals for the purpose of emotion recognition. Here we study electroencephalogram (EEG) signals, their utility for the application of emotion recognition and propose a Merged Recursive Neural Network (LSTM) architecture framework for emotion classification using EEG data. We have validated our procedure and model on DEAP dataset which is a benchmark dataset for emotion recognition using EEG signals [11].

II. ELECTROENCEPHALOGRAM (EEG) SIGNAL BASICS

EEG signals are the recordings of spontaneous electrical activity of brain via metal electrodes and conductive media from the scalp. The voltage fluctuations are mainly caused due to the flow of ionic currents within neurons and reflect cortical electrical activity.

Despite limited spatial resolution, EEG data proves to be a valuable asset for research and diagnostic purposes. The disorders like various fluctuations in the EEG signals are of utmost importance for the diagnosis of epilepsy, sleep disorders, coma and cases of brain death [12]. Unlike CT, MRI or PET, it offers millisecond range temporal resolution. These signals are non-invasive in nature and are acquired without much effort from the body. They greatly reflect the influence of emotions on automatic nervous system.

Though the spectrum of acquired EEG signal is continuous, ranging from 0 Hz to half of the sampling frequency, the brain state of an individual may make certain frequencies that are dominant over other frequencies. On this basis EEG signals have been classified into four major groups:

1) *Beta: frequency greater than 13 Hz.*
2) *Alpha: frequency range from 8 to 13 Hz.*
3) *Theta: frequency ranges from 4 to 8 Hz.*
4) *Delta: frequency ranges from 0.5 to 4 Hz.*

Based on experiments the highest electrical activity is observed from cerebral cortex due to its proximity to the surface and most of the meaningful information is acquired from this particular region.

III. RELATED WORKS

Emotions are psycho- physiological process tantamount of a person's mental health and state of mind. Emotions

978-1-7281-2087-4/19/$31.00 ©2019 IEEE

represent the perception of a being about the environment and world around him. As this field gains more impetus, EEG technology has become more affordable and less intrusive, thus making it suitable for effective deployment in healthcare industry.

The public release of DEAP dataset has motivated medical practitioners and researchers to dive deeper into the field of emotion analysis and has laid a firm foundation for them to build, test and deploy their novel ideas in this domain. Since the release of DEAP dataset many researchers have made a huge contribution and built effective algorithms to be validated on this dataset.

The preprocessing and dimensionality reduction of DEAP dataset is based upon the work of Jahankhani et al. [13]. This preprocessing methodology provides a starting baseline for feature extraction without much loss of information. It includes application of discrete wavelet transform to extract features from time series EEG signals [14]. This methodology provides compact representation of EEG signals in time- frequency domain. Further, to decrease the dimensionality statistical features such as maximum, minimum, mean and standard deviation were calculated.

Wavelet representation is better than Fourier representation because Fourier representation localizes a function only to its frequency domain and the signal cannot be significantly represented in spatial domain while wavelets are used to localize the function in frequency domain as well as spatial domain. Similarly the work of Kolte et al. [15] analyses the effects of various mother wavelet function on EEG signal analysis and concludes Daubachies (db4) wavelet transform for denoising and preprocessing the time series data [16].

Further the work of Samarth Tripathi et al. [17] explores the effectiveness of Neural Networks to classify user emotions based on DEAP dataset. The paper suggests two different neural models, a simple Deep Neural Network and a Convolution Neural Network for emotion classification. Their work is the testament of the fact that neural networks form a robust set of classifiers for analyzing physiological signals particularly EEG signals.

Hussein et al. [18] proposed the use of deep recurrent neural networks, particularly the long short-term memory (LSTM) model [22], for epileptic seizure diagnosis. Initially, the raw EEG signals were divided into non overlapping segments which were fed into deep recurrent neural layer (LSTM), in order to obtain a high level representation of EEG signals. Further the output of LSTM was fed into a dense layer followed by a softmax unit to create label predictions.

Compared to other works that are quite sensitive to noise, the LSTM model maintains its high detection performance even in the presence of common EEG artifacts and white noise as well. This model is thus robust in noisy as well as real life conditions.

IV. METHODOLOGY

Deep learning architectures have led us to achieve some of the most remarkable solutions to solve many contemporary problems. The power of deep learning lies in its supremacy over other statistical models due to high accuracy when trained on large dataset, which could only be achieved with huge amount of data, computation power and its layered architecture. Unlike traditional machine learning algorithms, deep learning algorithms learn about the data in hierarchical fashion which eliminates the need of domain expertise and reduces the need of extensive feature engineering.

However the EEG signal acquisition gives us a huge amount of data for processing. For instance, DEAP dataset consists of 40 trials for each volunteer and each trial includes readings from 40 EEG channels. Each channel consists of 8064 data points, thus accounting for total of 322560 readings for a single trial. With limited hardware resources, this data is very difficult to process and might contain some data that does not contribute much to the accuracy of the proposed model. Therefore we begin by cleaning our dataset and reducing its dimensionality by suitable preprocessing methodology.

Feature Extraction from EEG Signal is achieved using discrete wavelet transforms. The choice of optimal wavelet algorithm depends entirely on the application desired. In EEG Signal Feature Extraction, Db of order 2 is proved to be the optimal mother wavelet function [15] as compared to other mother wavelet functions. In this work there is overlap between iterations to pick up each detail that may be missed by other wavelet algorithms. At each step of iterations wavelet function is applied to input data. The looping variable is incremented by two at each iteration and thus if original data has N values, the computation will calculate N/2 differences.

In the present work, we propose a neural model for our research, a merged recurrent neural network followed by fully connected dense layers. The model uses contemporary deep learning techniques such as dropouts [19] to reduce over fitting of the data, LSTM cells to remove the problem of vanishing gradient and short term memory; and ReLu activation function is also used to include non-linearity in our model. The model architecture is implemented using Keras and Tensor flow libraries in python.

A. Recurrent Neural Network:

Neural networks form the basis of deep learning algorithms. In these networks, we assume all inputs and outputs to be independent of each other; i.e. the parameters of particular neuron are not affected by the input- output sequence of any other neuron. But for sequential data in which inputs and outputs are dependent on each other, the assumptions made by neural networks does not allow us to capture the temporal information contained in the data.

There are countless learning tasks that deals with sequential data such as image captioning, video analysis, handwriting analysis, study of genomes and numerical time series data obtained through sensors. In these tasks neural networks fails miserably due to their incapability to capture temporal information [20]. Unlike neural networks, recurrent neural networks (RNNs) have internal memory which enables it to retain past information as well. Thus, RNNs have a deeper understanding of sequential data and its context as compared to other algorithms.

A single layer of recurrent neural network has large number of processing units commonly called RNN cells. From equation (1) and (2) we see that each of these cells takes two inputs: activation from previous cell ($a^{<t-1>}$) and current input from dataset ($x^{<t>}$). These input variables are multiplied with their respective weight matrices and a bias term (b_a) is added to them. Further they are passed through corresponding activation functions to generate an input for the next cell. Output ($\hat{y}^{<t>}$) is calculated by multiplying ($a^{<t>}$) with the weight matrix (w_{ya}) and adding a bias term (b_y) followed by a softmax function. The structure on an RNN cell and flow of data through it is shown in Fig. 1.

$$a^{<t>} = tanh(w_{aa}\, a^{<t-1>} + w_{ax}\, x^{<t>} + b_a) \quad (1)$$

$$\hat{y}^{<t>} = softmax(w_{ya}\, a^{<t>} + b_y) \quad (2)$$

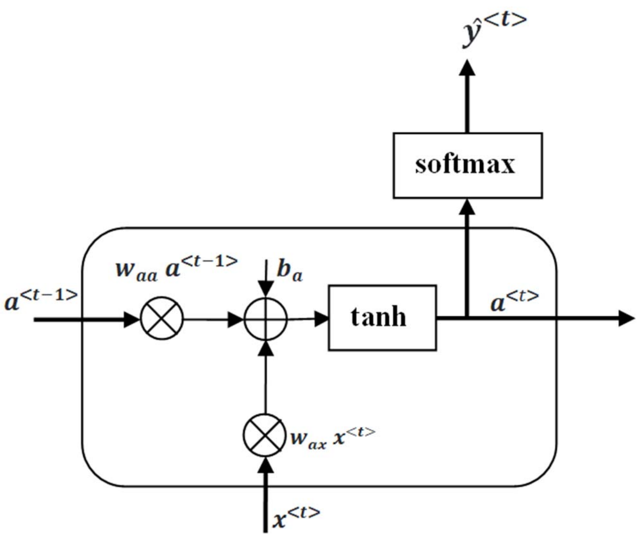

Fig. 1. Recurrent Neural Network cell

B. LSTM basics:

RNN effectively enables us to remember short sequences and model time- dependent data. However, they do suffer from one of the major problems of vanishing gradient [21]. During back propagation, the gradient reduces to such a small value that no parameters are significantly updated, thus restricting RNN to learn effectively from the data.

There are number of measures to overcome this problem. In the present work, LSTMs [22] have been used to rectify the problem of vanishing gradient and also enable our model to learn long sequences. The distinguishing feature of LSTMs that makes them more appropriate choice over RNNs is its internal architecture shown in Fig. 2. There are three gates which provide LSTM its functionality.

Here, $c^{<t>}$ represents the cell state, $a^{<t>}$ represents output from the block while $c'^{<t>}$ represents candidate for cell state at timestamp (t). Now, to obtain the memory vector for current time stamp the candidate is calculated as given in equation (3).

$$c'^{<t>} = tanh(w_c\, [a^{<t-1>}, x^{<t>}] + b_c) \quad (3)$$

1) Forget gate: it is responsible for the removal of information from the cell state that is no longer needed for the LSTM to understand and learn from the data. As shown in equation (4), the gate takes two inputs, input ($x^{<t>}$) at time stamp t and the activation from previous state ($a^{<t-1>}$). These set of input states are multiplied to the weight matrix (w_f) and a bias term (b_f) is added to it. The resultant is passed through sigmoid function. The output of sigmoid function further decides which value to keep and which value to discard. If the output is ‘0’, the forget gate no longer remembers the past value while for output equals ‘1’, the past value is retained.

$$\Gamma_f = \sigma(w_f\, [a^{<t-1>}, x^{<t>}] + b_f) \quad (4)$$

2) Update gate: it is primarily responsible for the addition of information to the cell state. As shown in equation (5), the gate takes two input, input ($x^{<t>}$) at time stamp t and the activation from previous state ($a^{<t-1>}$). These set of input states are multiplied to the weight matrix (w_u) and a bias term (b_u) is added to it, then the resultant is passed through sigmoid function.

Based on the output of sigmoid function, the gate ensures that information of significant importance is added.

$$\Gamma_u = \sigma(w_u\, [a^{<t-1>}, x^{<t>}] + b_u) \quad (5)$$

3) Output gate: Output gate filters the information that is not required in the output by the LSTM cell. As shown in equation (6), the gate takes two input, input ($x^{<t>}$) at time stamp t and the activation from previous state ($a^{<t-1>}$). These set of input states are multiplied to the weight matrix (w_o) and a bias term (b_o) is added to it. The resultant is passed through sigmoid function. The output gate selects only useful information from current cell and propagate it as an input to the next cell.

$$\Gamma_o = \sigma(w_o\, [a^{<t-1>}, x^{<t>}] + b_o) \quad (6)$$

Once the value of all the gates are calculated final cell state and activation for the next cell are predicted using these values as shown in equation (7) and (8).

$$c^{<t>} = \Gamma_u * c'^{<t>} + \Gamma_f * c^{<t-1>} \quad (7)$$

$$a^{<t>} = \Gamma_o * \tanh(c^{<t>}) \quad (8)$$

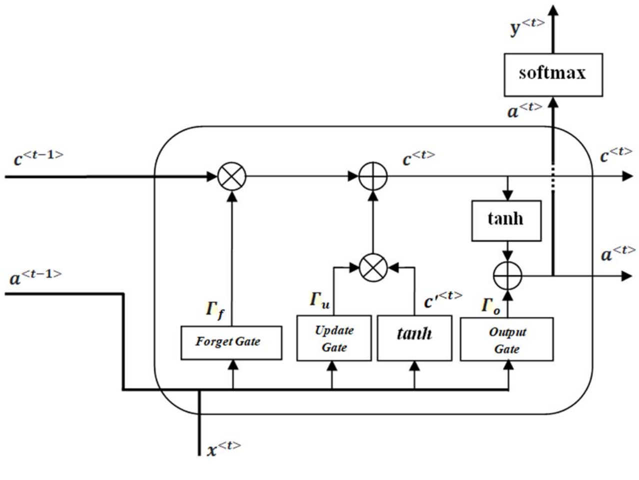

Fig. 2. LSTM cell

V. Proposed Algorithm

Biomedical signals are highly non stationary in nature and their statistical characteristics changes abruptly over time. Therefore in our proposed algorithm we begin by dimensionality reduction and feature extraction from our dataset using discrete wavelet transform followed by use of statistics to capture the nature and trend in variations of the dataset. Validation of the algorithm is done using DEAP Dataset. Mean, median, maximum, minimum, standard deviation, range, skewness and kurtosis are statistical measures used over subsets of our dataset.

In our research work we have used a merged LSTM model for the classification problem. The features extracted from the dataset are split into training and testing data and fed to the model. The model is trained and validated on the test set; a detailed description of the flow of algorithm is shown in Fig. 4. Our merged LSTM model uses LSTM layers and dense layers as the first half of the network. The preprocessed data of each of the 40 channels is fed simultaneously to this network, and this is done to provide more detailed learning to each network specific to a particular channel. Later the second half of this network consists of two fully connected layers followed by an output layer. The outputs from each of the first half of the network are concatenated and fed as input to the second half of the network as shown in Fig. 3.

It is evident from numerous researches that not all channels or regions of the brain contribute equally towards predicting the occurrence of each emotion and some regions show more activity and display a particular emotion more prominently [23]. For example disgust is found to be associated with right-sided activation in the frontal and anterior temporal regions as compared to happy condition. Thus intuitively the merged LSTM model captures the very contribution of each channel and helps us in predicting the emotional state of a person more effectively.

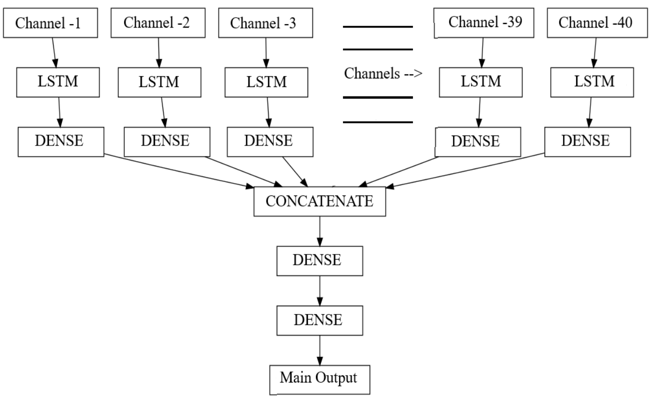

Fig. 3. Merged LSTM model

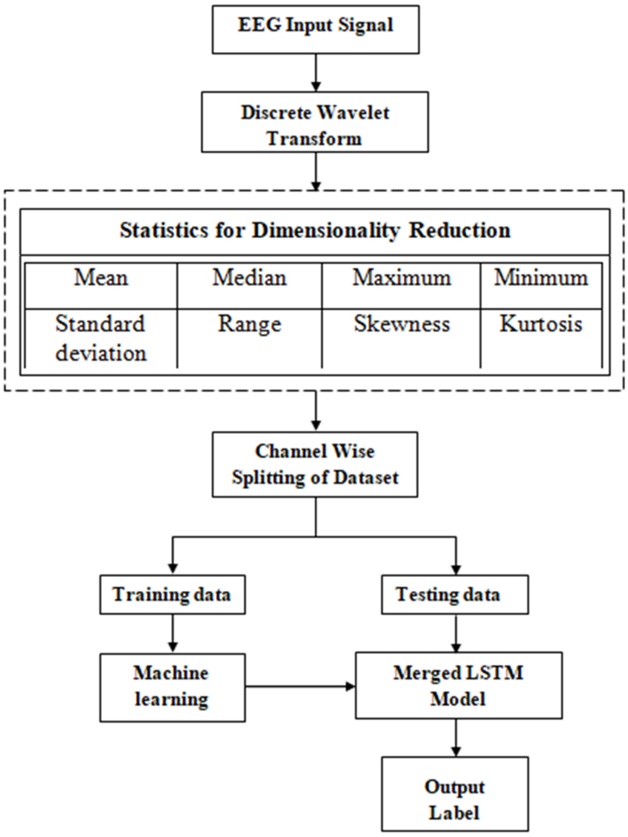

Fig. 4. Flow of algorithm

VI. Results

We have validated our algorithm on DEAP dataset which is a multimodal dataset for emotion analysis using EEG, physiological and video signals. The dataset consists of two parts. First part comprises of the ratings from an online self-assessment where 120 one-minute extracts of music videos were each rated by 14-16 volunteers based on arousal, valence and dominance. While second part comprise of the participant ratings, physiological recordings and face video of an experiment where 32 volunteers (16 men and 16 women, aged between 19 and 37) watched a subset of 40 above mentioned music videos. The proposed music videos induced various emotions in different users who then rated these videos on the valence- arousal scale [25], the same videos had an on-line evaluation that could be used for comparison.

However in present work we have used preprocessed and segmented version of data that has been made available in MATLAB and pickled Python/ Numpy formats. The data is segmented into 60 sec trial and a 3 sec pre-trial baseline removed. This version of data is well suited for testing a classifier or regression technique without the hassle of processing all data first. The entire dataset contains files of 32 participants and each file contains two arrays as illustrated in Table I.

TABLE I. DEAP DATASET DESCRIPTION

Array name	Array shape	Array contents
Data	40x40x 8064	video/trial x channel x data
Labels	40x4	video/trial x label (valence, arousal, dominance, liking)

Comparing our results with the existing state of the art methods clearly depicts supremacy of the proposed algorithm over existing methods. Samarth Tripathi et al. [17] Deep Neural Model achieves accuracy of 75.78% and 73.125% on valence and arousal two class emotion classification. Li et al. [26] propose a method of Deep Belief Networks for Emotion Identification and achieve an accuracy of 58.4% on valence, 64.2% on arousal, 65.8% on dominance and 66.9% on liking. The proposed algorithm gave us staggering accuracy of 84.89% on two class valence classification and 83.85% on two class arousal classification using a batch size of 128 when the model is run on 60 epochs. The same model was trained on all four emotions and the results are shown in TABLE II.

TABLE II. ACCURACY ON DIFFERENT EMOTIONS

Emotion	Accuracy
Valence	84.89%
Arousal	83.85%
Dominance	84.37%
Liking	80.72%

VII. CONCLUSION

In this paper, we built upon the prior research work in the field of EEG data analysis and Emotion recognition. Further we explored the possibility of using Recurrent Neural networks to classify emotions using EEG signals. Our study once again proved the efficiency of neural networks in solving some of the biggest biomedical problems in a profound way. The proposed merged LSTM model provides state of the art classification accuracy; we achieved an accuracy of 84.89%, 83.85%, 84.37% and 80.72% on binary classification of valence, arousal, dominance and liking respectively. This also proves that neural networks form a robust set of classifiers for EEG signals surpassing traditional learning techniques.

The present model also lays a firm foundation for future work on emotion analysis. Our future work would focus more on development and improvement of existing model for multiclass classification of a each emotion.

VIII. REFERENCES

[1] Darwin, Charles, and Phillip Prodger. The expression of the emotions in man and animals. Oxford University Press, USA, 1998.

[2] Loewenstein, George, and Jennifer S. Lerner. “The role of affect in decision making.” Handbook of affective science619.642 (2003): 3.

[3] Gudivada, Venkat N., et al. “Cognitive analytics: Going beyond big data analytics and machine learning.” Handbook of statistics. Vol. 35. Elsevier, 2016. 169-205.

[4] Lang, Peter J. "A bio‐informational theory of emotional imagery." Psychophysiology 16.6 (1979): 495-512.

[5] Fasel, Beat, and Juergen Luettin. "Automatic facial expression analysis: a survey." Pattern recognition 36.1 (2003): 259-275.

[6] Strapparava, Carlo, and Rada Mihalcea. "Learning to identify emotions in text." Proceedings of the 2008 ACM symposium on Applied computing. ACM, 2008.

[7] Gunes, Hatice, and Massimo Piccardi. "Bi-modal emotion recognition from expressive face and body gestures." Journal of Network and Computer Applications 30.4 (2007): 1334-1345.

[8] Busso, Carlos, et al. "Analysis of emotion recognition using facial expressions, speech and multimodal information." Proceedings of the 6th international conference on Multimodal interfaces. ACM, 2004.

[9] Kim, Kyung Hwan, Seok Won Bang, and Sang Ryong Kim. "Emotion recognition system using short-term monitoring of physiological signals." Medical and biological engineering and computing 42.3 (2004): 419-427.

[10] Bong, Siao Zheng, M. Murugappan, and Sazali Yaacob. "Analysis of electrocardiogram (ecg) signals for human emotional stress classification." International Conference on Intelligent Robotics, Automation, and Manufacturing. Springer, Berlin, Heidelberg, 2012.

[11] Koelstra, Sander, et al. "Deap: A database for emotion analysis; using physiological signals." IEEE transactions on affective computing 3.1 (2012): 18-31.

[12] Teplan, Michal. "Fundamentals of EEG measurement." Measurement science review 2.2 (2002): 1-11.

[13] Jahankhani, Pari, Kenneth Revett, and Vassilis Kodogiannis. "Data mining an EEG dataset with an emphasis on dimensionality reduction." 2007 IEEE Symposium on Computational Intelligence and Data Mining. IEEE, 2007.

[14] Adeli, Hojjat, Ziqin Zhou, and Nahid Dadmehr. "Analysis of EEG records in an epileptic patient using wavelet transform." *Journal of neuroscience methods* 123.1 (2003): 69-87.

[15] Chavan, A., and M. Kolte. "Optimal Mother Wavelet for EEG Signal Processing." International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 2.12 (2013): 5959-5963.

[16] Raghuveer R. Bapordikar A , „ Wavelet Transforms – Introduction to theory and applications.‟ Addis on – Wesley , 2000

[17] Tripathi, Samarth, et al. "Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset." Twenty-Ninth IAAI Conference. 2017.

[18] Hussein, Ramy, et al. "Epileptic seizure detection: a deep learning approach." arXiv preprint arXiv:1803.09848 (2018).

[19] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." The Journal of Machine Learning Research 15.1 (2014): 1929-1958.

[20] Sundermeyer, Martin, et al. "Comparison of feedforward and recurrent neural network language models." 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013.

[21] Hochreiter, Sepp. "The vanishing gradient problem during learning recurrent neural nets and problem solutions." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 6.02 (1998): 107-116.

[22] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

[23] Davidson, Richard J., et al. "Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology: I." Journal of personality and social psychology58.2 (1990): 330.

[24] Tieleman, Tijmen, and Geoffrey Hinton. "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude." COURSERA: Neural networks for machine learning 4.2 (2012): 26-31.

[25] Russell, James A. "A circumplex model of affect." Journal of personality and social psychology 39.6 (1980): 1161.

[26] Li, Xiang, et al. "EEG based emotion identification using unsupervised deep feature learning." (2015).