

# Emotion Recognition based on EEG using LSTM Recurrent Neural Network

Salma Alhagry

Faculty of Computer and Information  
Cairo University  
Cairo, Egypt

Aly Aly Fahmy

Faculty of Computer and Information  
Cairo University  
Cairo, Egypt

Reda A. El-Khoribi

Faculty of Computer and Information  
Cairo University  
Cairo, Egypt

**Abstract**—Emotion is the most important component in daily interaction between people. Nowadays, it is important to make the computers understand user's emotion who interacts with it in human-computer interaction (HCI) systems. Electroencephalogram (EEG) signals are the main source of emotion in our body. Recently, emotion recognition based on EEG signals have attracted many researchers and many methods were reported. Different types of features were extracted from EEG signals then different types of classifiers were applied to these features. In this paper, a deep learning method is proposed to recognize emotion from raw EEG signals. Long-Short Term Memory (LSTM) is used to learn features from EEG signals then the dense layer classifies these features into low/high arousal, valence, and liking. DEAP dataset is used to verify this method which gives an average accuracy of 85.65%, 85.45%, and 87.99% with arousal, valence, and liking classes, respectively. The proposed method introduced high average accuracy in comparison with the traditional techniques.

**Keywords**—*Electroencephalogram; emotion; emotion recognition; deep learning; long-short term memory*

## I. INTRODUCTION

Emotion is the most important component of being human, and very essential for everyday activities, such as the interaction between people, decision making, and learning. It eases the communication between people and makes it representative. It is important to detect and recognize the emotion in computer systems which people interact with, to enhance the communication between users and machines. Moreover, we need to know the current state of the user to enhance the accuracy and throughput of the system.

In order to make the computer understand and recognize emotion, we need to understand the sources of them in our body. Emotion could be expressed verbally like some known words or non-verbally like the tone of voice, facial expression and physiological changes in our nervous system. Voice and facial expression are not reliable indicators of emotion because they can either be fake by the user or may not be produced as a result of a specific emotion.

The physiological signals are more accurate because the user cannot control it. Physiological changes are the main sources of emotion in our body. There are two types of physiological changes: one is related to Central Nervous System (CNS) and the other is related to Peripheral Nervous System (PNS). CNS consists of the brain and spinal cord. The brain is the center of everything in our body and every change in the electrical activity is translated into different actions and

emotion. Electroencephalogram (EEG) is a measure of these electrical changes. EEG is defined as the electrical activity of an alternating type recorded from the scalp surface after being picked up by metal electrodes and conductive media [1].

Emotion recognition based EEG signals will provide an accurate emotion to use it in many fields. It can be used in automatic healthcare applications, helps autism to express their emotion and detect the state of the learner in E-learning system to develop an adaptive E-learning system.

In recent years, many emotion recognition approaches based on EEG signals have been proposed by many researchers. Koelstra [2] introduced a database for emotion analysis using physiological signals (DEAP). The main purpose of this database is to create music video recommendation system based on emotional state of the user. The music video clips are used to elicit different emotion. The physiological signals of 32 participants were recorded such as galvanic skin response, plethysmograph, skin temperature, breathing rate, Electromyogram, and EEG signals. 216 EEG features were extracted which are theta (4-8 Hz), slow alpha (8-10 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (30+ Hz) spectral power for 32 electrodes, and the difference between the spectral power of all the symmetrical pairs of electrodes on the right and left hemisphere. Fisher's linear discriminant was used for feature selection, then Gaussian naive Bayes used for classification with 3 different binary classification problems which are low/high arousal, low/high valence, and low/high liking.

Atkinson and Campos [3] proposed an EEG feature-based emotion recognition approach. DEAP [2] dataset was used to verify this approach. Statistical features were extracted which were the median, standard deviation, and kurtosis coefficient. Furthermore, band power of frequencies (theta (4-8 Hz), slow alpha (8-10 Hz), alpha (8-12 Hz), beta (12-30 Hz), gamma (30+ Hz)), Hjorth parameters (HP) and Fractal dimension (FD) were extracted for each channel. The minimum-Redundancy Maximum-Relevance (mRMR) was used to select a relevant set of extracted features. Then support vector machine (SVM) was used to classify features into low/high arousal and low/high valence classes.

Jadhav et al. [4] used EEG spectrogram image for emotion recognition. Gray-Level Co-occurrence Matrix (GLCM) features were extracted from EEG spectrogram image. DEAP dataset was used in this work to classify four emotions which are Happy, angry, relax, and sad. K-nearest neighbor was used

for classification.

Chanel et al. [5] proposed a new approach for adapting game difficulty according to the current emotion of the player. EEG signals were recorded from 14 players playing a Tetris game at three different levels easy, medium, and hard which are related to boredom, engagement, and anxiety emotions, respectively. For each electrode, the energy of different frequency bands which are theta (4-8 Hz), alpha (8-12 Hz), and beta (12-30 Hz) was computed using Fourier Transform. Furthermore,  $EEG_W$  features were computed for all electrodes as shown in (1).

$$EEG_W = \log \frac{\sum_{i=1}^{N_e} \beta_i}{\sum_{i=1}^{N_e} \theta_i + \alpha_i} \quad (1)$$

Different feature selection methods were experienced with different classifiers. The best accuracy was 56% which was obtained by using analysis of variance (ANOVA) as feature selection method and Linear discriminant analysis (LDA) as a classifier.

Yoon and Chung [6] proposed a new methodology for emotion recognition from EEG signals. DEAP dataset was used to verify this method. Fast Fourier Transform analysis was used in feature extraction. Then, feature selection based on Pearson correlation coefficient was applied on extracted features. They proposed a probabilistic classifier based on Bayes theorem and a supervised learning using a perceptron convergence algorithm.

Naser and Saha [7] proposed a new method for emotion recognition from EEG signals. DEAP dataset was used to verify this method. Dual-tree complex wavelet packet transform (DT-CWPT) was used for feature extraction. Then redundant features were eliminated using Singular value decomposition (SVD), QR factorization with column pivoting (QRcp), and F-ratio. Support vector machine was used for classification.

Liu et al. [8] proposed a new method for emotion recognition from EEG signals using DEAP dataset. Twelve different features were extracted from the time domain, frequency domain, time-frequency domain, and multi-electrode features. Minimum Redundancy Maximum Relevance (mRMR) was used for feature selection. K-Nearest Neighbour (KNN) and Random Forest (RF) were used for classification.

Bhagwat and Paithane [9] proposed a new method to classify four emotions which are happy, angry, cry, and sad. Wavelet Transform (WT) was used to extract features from raw EEG signals. Hidden Markov Model (HMM) was used for classification.

Hatamikia and Nasrabadi [10] proposed a new method for emotion recognition from EEG signals. Four feature extraction methods were used which are Approximate Entropy, Spectral Entropy, Katz's fractal dimension, and Petrosian's fractal dimension. In order to select the most informative features; two-stage feature selection method based on Dunn index and Sequential forward feature selection algorithm were used. Self-Organization Map (SOM) was used to classify emotions.

In short, in the previously presented work, researchers proposed various methods to extract different features from raw EEG signals. Different types of classifiers were applied to extracted features to recognize emotion. In order to improve

the accuracy of emotion recognition system, In this paper, a new method is proposed to recognize emotion from raw EEG signals directly by using an end-to-end deep learning approach. The proposed method improves the accuracy as discussed in Section V.

The rest of the paper is organized as follows: in Section II DEAP dataset is described. Long Short-Term Memory Recurrent Neural Network (LSTM-RRN) is described in Section III. The proposed method is presented in Section IV. Results are shown in Section V. The whole paper is concluded in Section VI.

## II. DATASET

DEAP dataset was recorded in order to create an adaptive music video recommendation system based on the current state of the user. Physiological signals were recorded from 32 healthy participants aged between 19 and 37 (with mean age 26.9 years). Each participant watched a one-minute long music video. After each trial/video, each participant performs self-assessment of their level of arousal, valence, like/dislike, and dominance. For 22 participants of the 32 participants, the frontal face video was also recorded. EEG and peripheral signals were recorded at a sampling rate of 512 Hz. Their EEG data were downsampled to 128 Hz, then, averaged to the common reference, after that, eye artifacts were removed, and a high-pass filter was applied. Also, the Peripheral signals were downsampled to 128 Hz.

Each participant's file contains two arrays as described in Table I.

TABLE I. DEAP ARRAYS OF EACH PARTICIPANT

Array Name	Array shape	Array contents
data	$40 \times 40 \times 8064$	video/trail $\times$ channel $\times$ data
labels	$40 \times 4$	video/trail $\times$ label (valence,arousal,dominance,liking)

As mentioned in Table I each participant have an array of 40 watched videos  $\times$  40 (EEG+peripheral) channels  $\times$  8064 reading. In this paper, only EEG signals are used. The 8064 reading per EEG channels divided into 12 segments which each is 5 seconds long of approximately 21504 reading.

## III. LONG SHORT TERM MEMORY (LSTMS)

Long Short-Term Memory Networks (LSTMs) are special kind of Recurrent neural network (RNN). It was introduced by Hochreiter and Schmidhuber in 1997 [11] in order to overcome the problem of long-term dependency in RNN. Long sequences can be difficult to learn from standard RNN because it's trained by back-propagation through time (BPTT) and that causes the problem of vanishing/exploding gradient. To solve this, the RNN cell is replaced by a gated cell, like LSTMs cell. Figure 1 shows the basic architecture of LSTMs cell.

These gates control which information must be remembered in memory and which are not. The memory added to LSTMs cell makes it able to remembers previous steps. The key to LSTMs is the cell state (the horizontal line on the top of figure 1 ( $C_t$ )). LSTM has the ability to remove or add

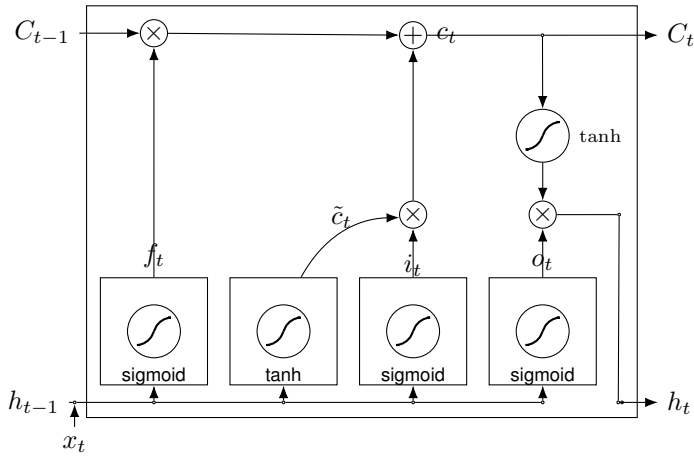


Fig. 1. LSTM cell architecture.

information to the cell state by using three gates. The first gate is a forget gate to decide what information to throw away from the cell state, this decision made by a sigmoid layer

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

The second gate is an input gate which consists of sigmoid layer to decide which values will be updated, and tanh layer which creates a vector of new updated values as described in (3) and (4)

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

Then the cell state updated from equations 2, 3, and 4 by

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (5)$$

Finally, the output of the current state will be calculated based on the updated cell state and a sigmoid layer which decides what parts of the cell state will be the final output as described in equations 6 and 7

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$

$$h_t = o_t * \tanh(C_t) \quad (7)$$

where  $\sigma$  is sigmoid activation function which squashes numbers into the range (0,1),  $\tanh$  is hyperbolic tangent activation function which squashes numbers into the range(-1,1),  $W_f$ ,  $W_i$ ,  $W_c$ ,  $W_o$  are the weight matrices,  $x_t$  is the input vector,  $h_{t-1}$  denote the past hidden state and  $b_f$ ,  $b_i$ ,  $b_c$ ,  $b_o$  are bias vectors.

#### IV. PROPOSED METHOD

In this paper, end-to-end deep learning neural network is applied to raw EEG signals of 32 participants who watched the 40 videos, in order to recognize the emotion elicited from these videos. Each video segmented into 12 segments with a length of 5 seconds. DEAP [2] dataset was used to verify the algorithm in this work. As mentioned in Section II each participant has an array of data and array of labels. label array represents the ratings of each video performed by each participant. The ratings represent their levels of arousal, valence, and liking in continuous scale ranged from 1 to 9.

Three different classification problems were posed: low/high arousal, low/high valence, and low/high liking. Since there are two levels only per each classification problem, then the continuous rating range per class is thresholded in the middle such that if the rating is greater than or equal to five then the video/trail belongs to high class otherwise, it belongs to the low class.

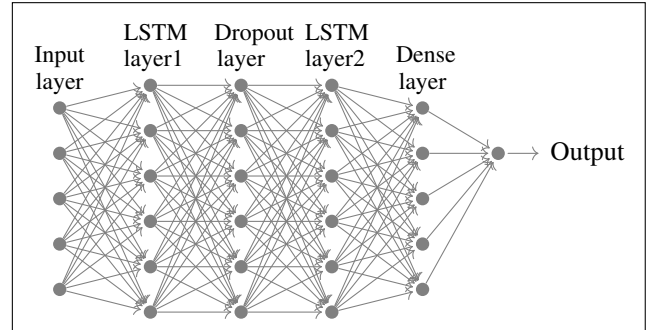


Fig. 2. Neural network model.

Fig. 2 shows the proposed deep learning neural network model. It consists of fully connected two LSTM layer, dropout layer, and dense layer. The dropout layer used to reduce the Overfitting by preventing units from co-adapting too much. The LSTM and dropout layers are used to learn features from raw EEG signals and the dense layer is used for classification.

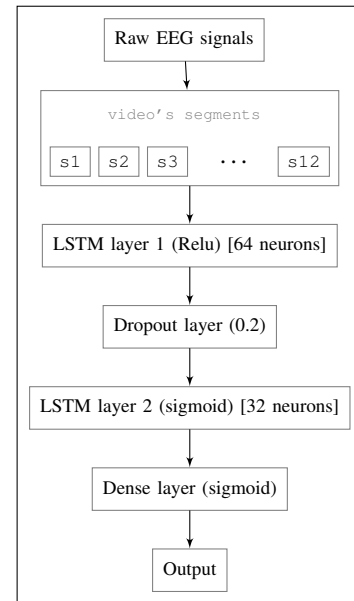


Fig. 3. Detailed proposed model.

Each participant data consists of 8064 readings for 32 EEG channels for each video. each video segmented into 12 segments with a length of 5 seconds. each segment consists of 672 readings for 32 EEG channels. The Deep learning model used in this paper consists of the input layer, the first sequence-to-sequence LSTM layer, a dropout layer with a probability of 0.2, a many-to-one LSTM layer, and a dense layer for

binary classification as shown in figure 3. The first hidden layer contains 64 neurons and used Relu as an activation function.

$$ReLU(x) = \max(0, x) \quad (8)$$

The second LSTM layer contains 32 neurons and used sigmoid activation function. The dense layer used sigmoid activation function. The total number of parameters in this model is 5534113.

The model is trained on 75% of the videos using 4 fold cross-validation and tested on 25% of them. Thirty epochs used for each cross-validation iteration. RMSprop optimizer used in training process with 0.001 learning rate. Keras [12] library with TensorFlow backend is used to implement the proposed deep learning method.

## V. RESULTS

DEAP [2] dataset is used to verify the algorithm proposed in this paper. As described in Section II, EEG signals are recorded from 32 participants. The algorithm is applied to each participant separately and the accuracy is calculated. therefore, the average accuracy over all participants is calculated for three classification problems which are low/high arousal, valence, and liking.

The results provided by the proposed method has been compared with four different methods which all used DEAP dataset as shown in Table II. The proposed method achieves the higher accuracy for all classes as shown in Fig. 4.

TABLE II. THE AVERAGE ACCURACY

	Arousal	Valence	liking
Koelstra et al. [2]	0.620	0.567	0.554
Atkinson and Campos [3]	0.7306	0.7341	-
Yoon and Chung [6]	0.701	0.709	-
Naser and Saha [7]	0.662	0.643	0.702
proposed method	<b>0.8565</b>	<b>0.8545</b>	<b>0.8799</b>

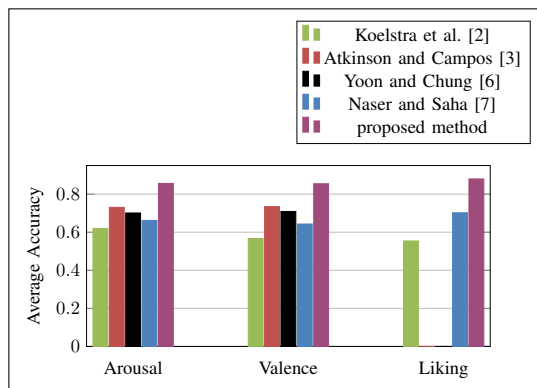


Fig. 4. Average accuracy.

## VI. CONCLUSION

This paper presents an end-to-end deep learning neural networks method to recognize emotion from raw EEG signals. LSTM-RNN is used to learn features from EEG signals then the dense layer is used for classification. Results show that the proposed method is a very promising choice for emotion recognition, because of its powerful ability to learn features from raw data directly. It achieves high average accuracy over participants compared to the traditional feature extraction techniques.

## REFERENCES

- [1] Michal Teplan. Fundamentals of eeg measurement. *Measurement science review*, 2(2):1–11, 2002.
- [2] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2012.
- [3] John Atkinson and Daniel Campos. Improving bci-based emotion recognition by combining eeg feature selection and kernel classifiers. *Expert Systems with Applications*, 47:35–41, 2016.
- [4] Narendra Jadhav, Ramchandra Manthalkar, and Yashwant Joshi. Electroencephalography-based emotion recognition using gray-level co-occurrence matrix features. In *Proceedings of International Conference on Computer Vision and Image Processing*, pages 335–343. Springer, 2017.
- [5] Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun. Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(6):1052–1063, 2011.
- [6] Hyun Joong Yoon and Seong Youb Chung. Eeg-based emotion estimation using bayesian weighted-log-posterior function and perceptron convergence algorithm. *Computers in biology and medicine*, 43(12):2230–2237, 2013.
- [7] Daimi Syed Naser and Goutam Saha. Recognition of emotions induced by music videos using dt-cwpt. In *Medical Informatics and Telemedicine (ICMIT), 2013 Indian Conference on*, pages 53–57. IEEE, 2013.
- [8] Jingxin Liu, Hongying Meng, Asoke Nandi, and Maozhen Li. Emotion detection from eeg recordings. In *Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2016 12th International Conference on*, pages 1722–1727. IEEE, 2016.
- [9] Anuja R Bhagwat and AN Paithane. Human disposition detection using eeg signals. In *Computing, Analytics and Security Trends (CAST), International Conference on*, pages 366–370. IEEE, 2016.
- [10] Sepideh Hatamikia and Ali Motie Nasrabadi. Recognition of emotional states induced by music videos based on nonlinear feature extraction and som classification. In *Biomedical Engineering (ICBME), 2014 21th Iranian Conference on*, pages 333–337. IEEE, 2014.
- [11] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [12] François Chollet et al. Keras. <https://github.com/fchollet/keras>, 2015.