2020 16th IEEE International Colloquium on Signal Processing & its Applications (CSPA 2020), 28-29 Feb. 2020, Langkawi, Malaysia

An Efficient Approach to EEG-Based Emotion Recognition using LSTM Network

Anubhav
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
anubhav2901@gmail.com

Debarshi Nath
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
debarshinath94@gmail.com

Mrigank Singh
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
mriganksingh1@gmail.com

Divyashikha Sethia
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
divyashikha@dtu.ac.in

Diksha Kalra
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
kalradiksha11@gmail.com

S. Indu
Department of Electronics and Communication Engineering
Delhi Technological University
Delhi, India
s.indu@dce.ac.in

***Abstract*—This work aims to investigate the performance of the Long Short-Term Memory (LSTM) Model for EEG-Based Emotion Recognition. For the experimentation, we use the publicly available DEAP dataset, which consists of preprocessed EEG and physiological signals. Our work limits itself to the study of only the EEG signals to have a scope for developing an efficient headgear model for real-time monitoring of emotions. In this study, we extract the band power, a frequency-domain feature, from the EEG signals and compare the classification accuracies for Valence and Arousal domain for different classifiers. The proposed Long Short-Term Memory (LSTM) model achieves the best classification accuracy of 94.69% and 93.13% for Valence and Arousal scales, respectively, illustrating a significant average increment of 16% in valence and 18% in arousal in comparison to other classifiers.**

***Keywords*—EEG Data, Emotion, Emotion Recognition, DEAP dataset, Band power, LSTM Network**

I. Introduction

Emotion represents the state of mind of a person whether a person is happy or sad, angry or calm, stressed, or relieved. Emotions are the response to a particular stimulus. Studies suggest that emotion is a subjective experience: it varies from person to person, and because of this, it is one of the most challenging and exciting research fields in psychology [1]. Recognition of emotion plays a vital role in daily life as it can help in enhancing one's psychological health which is equally important as maintaining physical fitness. Nowadays, a lot more people suffer from anxiety, stress, hypertension, and other mental health-related issues. So, Emotion Recognition here plays a crucial role in improving the lives of people. For instance, when a game becomes too dull or too exciting, the level of the game can be modified depending on the excited emotional level of the person. Also, a computer can change the music or window background according to one's mood. There are various other applications in the field of mental health where the knowledge of human emotion helps the psychologist to treat stress, tension, and anxiety issues.

Emotion is a phenomenon that is difficult to grasp, and for its better understanding, there are various models proposed by researchers like Valence and Arousal Model by Russell [2]. This model represents emotions on a 2-D circular space where arousal represents the vertical axis, and valence represents the horizontal axis. The Circular space represents the neutral valence and medium value of arousal. Bradley et al. [3] proposed another model named Approach and Withdrawal Model or the vector model. It is also a 2-D model where the value of valence determines the direction of emotion where a positive value of valence shifts the emotion in the top vector. Likewise, the negative value of valence would shift the emotion in the down vector. Watson and Tellegen [4] developed a Positive and Negative Model. In this model, the vertical axis represents low to high positive affect, and the horizontal axis represents low to high negative affect.

Earlier researches on emotions were done using facial expressions, speech processing, and various other methods. However, since it is possible to fake this behaviour and techniques, the focus has now shifted on emotion recognition using other physiological signals such as Electrocardiography (ECG), Electromyography (EMG), Galvanic Skin Response (GSR), Respiration Rate (RR) and Electroencephalogram (EEG) signals [5], [6]. Emotion Recognition through EEG has vast applications in the field of Human-Computer Interaction (HCI), where the computer can adjust its behaviour according to user emotion. For the measurement of brain signals, Electroencephalogram (EEG) device is used, which measures the electrical activity of the brain. EEG device contains a large number of electrodes that can be placed on the Human

Skull, according to the 10-20 International System. Since the understanding of the entire EEG signal at once is a complex task, the EEG signal is divided into various frequency bands. The different frequency bands are the Alpha (0-4 Hz), Beta (4-8 Hz), Delta (8-13 Hz), Theta (13-30 Hz), and Gamma (>30 Hz). Each band is associated with different activities taking place in the body. For instance, the Delta wave is related to deep sleep as well as the deepest level of relaxation. Similarly, the Theta wave is associated with REM sleep, deep and raw emotions, and cognitive processing. Likewise, in a drowsy state, the Alpha wave comes into the picture. It is associated with relaxation and calmness. In a conscious state, the Beta wave is present during the thought process. Gamma waves are current when a person tries to perceive two different senses at the same time as sound and sight. EEG signals have broad applications ranging from Emotion Recognition to diseases and disorders like Sleep Apnea, Epilepsy, and Alzheimer's disease.

Although various classification techniques have been reported in the literature Alhagry et al. [7] reports a classification accuracy of 85.65%, 85.45%, and 87.99% for valance, arousal, and liking, by using LSTM model. Similarly, Li et al. [8] employed LSTM model achieving a mean accuracy of 76.6%. We compute the band power feature for each frequency band of the EEG signals and employ the machine learning methods, namely Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM), Decision Tree and Random Forest. Section II for Related Work presents a detailed description of the previous research work. In this work, we achieve a maximum classification accuracy of 94.69% for valence and 93.13% for arousal using the LSTM classifier, outperforming the other classifiers.

In the rest of the paper, Section II describes the related work. Section III contains the proposed methodology, and Section IV includes the experimental results. Finally, we conclude the paper with section V with the conclusion and future work.

II. Related Work

There are various emotions like happy, excited, angry, afraid, sad, depressed, calm, and contentment and the proper classification of these emotions can be beneficial for the study. Bastos-Filho et al. [9] have classified the emotional state as calm when the levels of arousal are below 4, and the level of valence is between 4 and 6. Similarly, for stress, the levels of arousal should be greater than 5, and that of valence should be less than 3.

In researches concerning the problem of emotion classification, the ability to classify emotions depends on two main factors:

1) Features extracted from the dataset.
2) Classifiers used for emotion classification.

The classification accuracy compared to the original dataset can be improved by extracting a wide range of features from the dataset. There are mainly three types of features :

1) Time domain features.
2) Frequency domain features.
3) Time-Frequency domain features.

Jenke et al. [10] have described several features and their relevance to EEG signals. Some of the features are statistical features like mean, standard deviation, power. Hjorth Features like activity, mobility, complexity. Frequency domain features include band power, higher-order spectra. The time-frequency domain features are Hilbert-Huang Spectrum (HHS) and Discrete Wavelet Transform (DWT). Recent researches have shown that frequency domain features are more useful in the analysis of EEG signals. A good number of papers have used PSD, or PSD-based features generated from EEG signal datasets and achieved good accuracy to solve problems in the domain of emotion recognition and classification. Raphael Vallat also mentions the use of PSD for a myriad of analyses [11]. This motivates us to use frequency-domain features for extracting information from the EEG signals and explore various classification techniques.

This paper examines different classification techniques for Emotion Recognition on the publicly available DEAP (Dataset for Emotional Analysis using Physiological Signals) dataset [12]. We have also found Recurrent Neural Networks (RNN) being used in recent years to address the problem of emotion recognition and classification effectively. We mention some of the prominent researches employing LSTM model and other classifiers using DEAP dataset. However, all of them achieve accuracy less than the proposed model in this paper.

Dabas et al. [13] proposed a 3-D emotional model (Valence, Arousal, and Dominance) for classifying emotions using the DEAP dataset. They used machine learning algorithms like SVM, Naive Bayes, and achieved an accuracy of 58.90% and 78.06%.

Liu et al. [14] has employed the DEAP dataset for classifying emotions and features like time-domain features (mean power, standard deviation, higher-order crossings, fractal dimension, Hjorth feature), frequency domain features (power spectral density), time-frequency domain feature (discrete wavelet transform). Multi-electrode features (differential asymmetry and rational asymmetry, magnitude squared coherence estimate) are computed and uses maximum relevance minimum redundancy (mRMR) for feature selection. KNN and RF are employed as classification techniques with the highest accuracy of 66.17% for arousal using a magnitude squared coherence estimate as a feature.

Wichakam et al. [15] have experimented on the DEAP dataset using band power as the feature and SVM as classifier. The maximum accuracy achieved is 64.9% for valence and 66.8% for liking while using the 3-dimensional emotion model. They have only used ten channels and have shown that performance accuracy is not improved even if 32 channels are employed.

Salama et al. [16] designed a 3-dimensional convolutional neural network for emotion recognition from multi-channel EEG data. They have used the DEAP dataset for analysis and have achieved 87.44% and 88.49% accuracy for valence and arousal classes.

Alhagry et al. [7] have proposed an end-to-end model employing the LSTM classifier for emotion classification on the DEAP dataset. The average subject-independent accuracy achieved for arousal, valence, and liking is 85.65%, 85.45%, and 87.99%.

Li et al. [8] have used RASM as a feature which represents frequency-space domain characteristics of the EEG signal. They have employed the LSTM model on DEAP dataset and achieved a mean accuracy of 76.6%. Xing et al. [17] built Stack Autoencoder (SAE) for EEG signal decomposition and used LSTM model for classification but still observed accuracy of 81.1% in valence and 74.38% in arousal.

Yang et al. [18] propose a parallel combination of Convolutional Neural Network and LSTM Network to extract features from the DEAP dataset and then use the softmax classifier for classification. This model obtained the mean accuracy of 90.80% and 91.03% for valence and arousal.

Previous works in the Emotion Recognition establish that the LSTM models perform better than other classification techniques. Still, the reported work examines either the raw EEG signals or the time-domain features for training the LSTM model. Here, we investigate the use of the band power, a frequency-domain feature for training the LSTM model. To compare the performance of the proposed LSTM model with other classifiers, we also train classifiers, namely SVM, KNN, Decision Tree, and Random Forest. On contrasting the results, we observe a maximum classification accuracy of 94.69% for valence and 93.13% for arousal using the LSTM classifier, which is significantly better than other classifiers.

III. Methodology

A. Dataset Description

The DEAP dataset [12] is a multimodal dataset for the determination of human emotional states available for public access. For the creation of the database, experiments were performed where 32 participants were made to watch 40 one-minute-long excerpts of selected videos, during which the signals from 32 channels of standard EEG headset and physiological signals from 8 channels were captured. Equipment had a sampling frequency of 128 Hz. The data was pre-processed, and artefacts were removed. The participants rated each video on a scale of 0-9 in terms of valence, arousal, dominance, and liking. These ratings become the benchmark for the classification of emotional states.

B. Pre-processing

Fig. 1 illustrates the EEG signals for two subjects (1, and 9) subjected to the same trial. Both signals indicate differences in the magnitude and the pattern of activations of the brain for different individuals. Thus, it displays the uniqueness in the processing of information in the brain for every individual.

Fig. 2 illustrates the EEG signals for a single person subjected to two different trials. Both signals display similarity in the magnitude of the activation of the brain for the subject.

As inferred from Fig. 1 and Fig. 2, we conclude that emotion recognition is to be done for each individual separately as the analysis displays no similarity in the activation of the signals recorded for different subjects in the population. Since every individual possesses unique consciousness and emotional limits, the prediction of emotion for an individual using the learning from any other individual will drastically reduce not only the accuracy in prediction but the model will also lose its credibility for the prediction of an unknown subject. But the analysis also shows that there does exist similarity amongst the signals, valence and arousal values for the different trials of an individual. Hence, there is a possibility of finding a pattern for a certain emotion by understanding the signals obtained for that individual only. We exploit the results of this analysis to design a customized model for emotion recognition. Although by increasing the size of data, it is possible to account for this diversity in the strengths of the EEG signals. This is demonstrated by the modern extensive Image Classification datasets such as Tencent [19] which consists of more than 17 million images. Such a large dataset enables the Deep Learning models to explore the hidden features of the dataset and account for the diversity in the sample population. But the DEAP dataset consists of only 32 subjects, so it is currently not possible to account for such variance in the EEG signals. Therefore this study limits its training and testing to independent subjects.

The original signals were recorded for 63s (3s prior and 60s for the video). The preceding signal recorded is not removed as

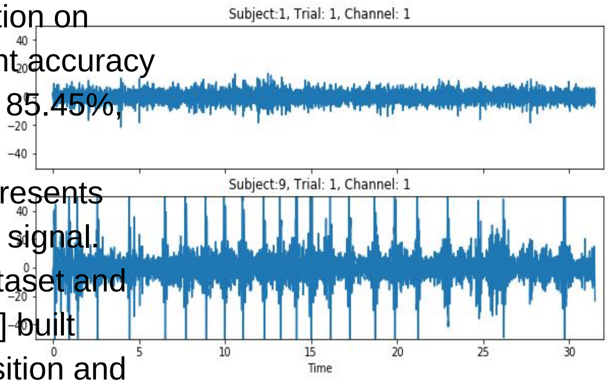

Fig. 1. EEG signal for Subjects:(1, 9), Trial: 1, and Channel: 1

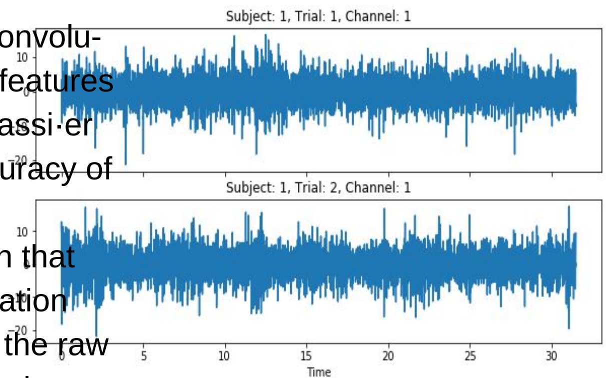

Fig. 2. EEG signal for Subject:1, Trials: (1, 2), and Channel: 1

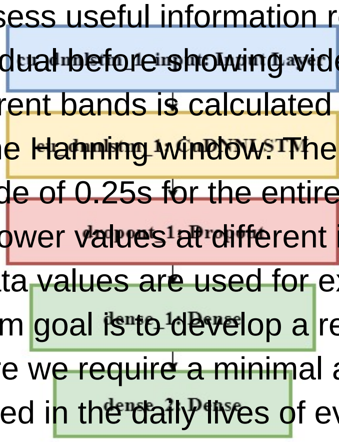

Fig. 3. Proposed LSTM Model

it does possess useful information regarding the state of mind of the individual before showing video trials. The band power for the different bands is calculated using the Welch method based on the Hanning window. The window length here is 1s and the stride of 0.25s for the entire 63s, therefore, obtaining 249 band power values at different instances of time. Only the EEG data values are used for experimental purposes, as our long term goal is to develop a real-time emotion prediction model where we require a minimal amount of hardware so that it can be used in the daily lives of every individual, especially patients.

C. LSTM Model

In this paper, we use the LSTM network as a useful tool for the prediction of the emotion of individuals. The LSTM networks are frequently used for handling sequential data such as paragraphs in NLP, previous electricity load in the case of the electricity demand prediction. LSTM cell possesses the ability to remember the distant as well as recent events to accurately predict the target variable. This property of retention can turn to be useful for emotion recognition as knowing about the past activations of the EEG signals can drastically affect the prediction of target variables and provide useful insights to the events leading to an appropriate response for the subject.

Fig. 3 shows the configuration of the Proposed LSTM model. We implement the model in Python 3 on the Google Colab platform with GPU support for the LSTM network. The LSTM layer has 40 nodes. Dense_1 layer has 10 nodes with 'tanh' activation function, and Dense_2 has a single node with 'sigmoid' activation function. We use a Dropout 25% between the LSTM Layer and Dense_1 Layer. We use Stochastic Gradient Descent (SGD) optimizer (learning rate=0.01, learning rate decay constant=1×10^{-5}, and momentum constant=0.9) to minimize the binary-cross-entropy loss function.

Other than the customized LSTM model, we also test the classification on the dataset using KNN, SVM, Decision Tree, and Random Forest. We evaluate the performance of each classifier on the pre-processed dataset. As expected from the previous studies, the proposed LSTM model outperforms the other classifiers by a huge margin.

IV. Results

To verify the effectiveness of the proposed LSTM model, we contrast the performance of KNN, SVM, Decision Tree, Random Forest, and LSTM models for classification of the preprocessed dataset. We test models several times to ensure the significance of the results observed. Table I highlights the average prediction accuracies.

TABLE I
TESTING ACCURACIES

Model	Valence	Arousal
KNN	79.69	75.78
SVM	76.56	72.66
Decision Tree	77.34	74.21
Random Forest	80.46	77.34
LSTM	**94.69**	**93.13**

On analyzing the results, we observe that the LSTM model outperforms the other classifiers by a large margin. We observe a remarkable increment of about 16% and 18% for valence and arousal when comparing the LSTM model with other classifiers. The highest increment in average testing accuracy of 18% for valence and 20% for arousal is observed when comparing the SVM classifier with the LSTM model.

We compare our results with the results of Yang et al. [18] following similar experimental procedures with the parallel Convolutional Recurrent Neural Network model. Here, we observe an increment of 4% for valence and 2% for arousal in average testing accuracies for all the subjects. Our proposed model notes a significant increment of 9% in valence and 7.5% in arousal for Alhagry et al. [7] and 14% in valence and 19% in arousal for Xing et al. [17]. Fig. 4 illustrates the average testing accuracy for valence and arousal of the 32 subjects.

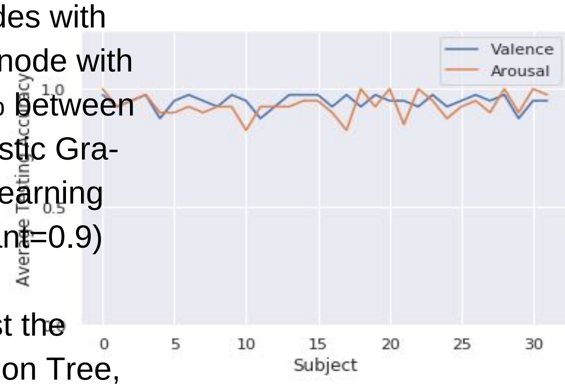

Fig. 4. Average testing accuracy of 32 subjects for LSTM model

V. Conclusion and Future Work

In this work, we evaluate power spectral density over the 32 channels of the DEAP dataset. We segregate them into five bands of frequencies, namely Alpha, Beta, Gamma, Delta, and Theta, to derive the band power of each band. We use band power as a feature for classifying valence and arousal of the subject. We evaluate and compare using KNN, SVM, Decision Tree, Random Forest, and LSTM as our classifiers. On analysis, we observe a minimum average increment of 16% in the average testing accuracies and maximum classification accuracy of 94.69% for valence and 93.13% for arousal using the LSTM classifier, which performs better than the current state-of-the-art classifiers.

In future, we can use the proposed experimental setup to obtain useful information regarding the emotions of the subjects and extend it for real-time applications. For further improvements, we can add more frequency domain features and test for their performance. As demonstrated by Wichakam and Vateekul [15], a subset of the channels for feature generation may perform better in terms of accuracy. The work may further be extended to include subject-independent models as well. The study can be also be extended to developing 3-D emotion models like the work of Dabas et al. [13].

References

[1] R. Plutchik, *The Emotions*. University Press of America, 1991.

[2] J. A. Russell, “A circumplex model of affect,” *Journal of personality and social psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[3] M. M. Bradley, M. K. Greenwald, M. C. Petry, and P. J. Lang, “Remembering pictures: Pleasure and arousal in memory,” *Journal of experimental psychology: Learning, Memory, and Cognition*, vol. 18, no. 2, pp. 379–390, 1992.

[4] W. David and T. Auke, “Toward a consensual structure of mood,” *Psychological bulletin*, vol. 98, no. 2, pp. 219–235, 1985.

[5] K. Takahashi, “Remarks on SVM-based emotion recognition from multi-modal bio-potential signals,” in *Proc. IEEE Int. Work. on Robot and Human Interactive Communication*, 2004, pp. 95–100.

[6] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, “Emotion Assessment: Arousal Evaluation Using EEG’s and Peripheral Physiological Signals,” *Multimedia Content Representation, Classification and Security. MRCS 2006. Lecture Notes in Computer Science*, vol. 4105, 2006.

[7] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, “Emotion Recognition based on EEG using LSTM Recurrent Neural Network,” *Emotion*, vol. 8, no. 10, pp. 355–358, 2017.

[8] Z. Li, X. Tian, L. Shu, X. Xu, and B. Hu, “Emotion Recognition from EEG using RASM and LSTM,” in *International Conference on Internet Multimedia Computing and Service*. Springer, 2017, pp. 310–318.

[9] T. F. Bastos-Filho *et al.*, “Evaluation of feature extraction techniques in emotional state recognition,” in *Proc. IEEE Int. Conf. Intelligent human computer interaction*. IEEE, 2012, pp. 1–6.

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

[11] R. Vallat. Bandpower of an EEG signal. [Online]. Available: https://raphaelvallat.com/bandpower.html (accessed Nov. 13, 2019).

[12] K. Sander *et al.*, “Deap: A Database for Emotion Analysis; Using Physiological Signals,” *IEEE Trans. Affective Computing*, vol. 3, no. 1, pp. 18–31, 2011.

[13] H. Dabas, C. Sethi, C. Dua, M. Dalawat, and D. Sethia, “Emotion Classification Using EEG Signals,” in *Proc. ACM Int. Conf. Computer Science and Artificial Intelligence*. ACM, 2018, pp. 380–384.

[14] J. Liu, H. Meng, A. Nandi, and M. Li, “Emotion detection from EEG recordings,” in *Proc. IEEE Int. Conf. Natural Computation, Fuzzy Systems and Knowledge Discovery*. IEEE, 2016, pp. 1722–1727.

[15] I. Wichakam and P. Vateekul, “An evaluation of feature extraction in EEG-based emotion prediction with support vector machines,” in *Proc. IEEE Int. Conf. Joint conference on computer science and software engineering*. IEEE, 2014, pp. 106–110.

[16] E. S. Salama, R. A.El-Khoribi, M. E.Shoman, and M. A. Shalaby, “Eeg-Based Emotion Recognition using 3D Convolutional Neural Networks,” *Int. Journal of Advanced Computer Science and Applications*, vol. 9, no. 8, 2018.

[17] X. Xing, Z. Li, T. Xu, L. Shu, B. Hu, and X. Xu, “SAE+LSTM: A New Framework for Emotion Recognition From Multi-Channel EEG,” *Frontiers in Neurorobotics*, vol. 13, p. 37, 2019.

[18] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, “Emotion Recognition from Multi-Channel EEG through Parallel Convolutional Recurrent Neural Network,” *Proc. IEEE Int. Conf. Joint Conference on Neural Networks*, 2018.

[19] B. Wu, W. Chen, Y. Fan, Y. Zhang, J. Hou, J. Huang, and T. Zhang, “Tencent ML-Images: A Large-Scale Multi-Label Image Database for Visual Representation Learning,” *arXiv preprint arXiv:1901.01703*, 2019.