

EEG Signal-Based Human Emotion Recognition Using Power Spectrum Density and Discrete Wavelet Transform



S. G. Shaila, B. M. Anirudh, Archana S. Nair, L. Monish, Prathyusha Murala, and A. G. Sanjana

Abstract Emotion recognition has been a problem in the field of brain–computer interface. Numerous ways are available for recognizing human emotions and one such technique is through Electroencephalogram (EEG) signals. EEG signals are recordings of the subject’s electrical activity in the brain. Feature extraction approaches such as Power Spectrum Density (PSD) and Discrete Wavelet Transform (DWT) are fed as features to various machine learning (ML) and deep learning (DL) models. This work aims to develop models that predict emotions from EEG data. In addition, the results of the above-mentioned feature extraction approaches are compared in this work. The proposed feature extraction methods and models are applied on the DEAP dataset.

Keywords Emotion recognition · EEG signals · Machine learning · Deep learning · LSTM · Power Spectrum Density · Discrete Wavelet Transform · Brain–computer interface

1 Introduction

For any social being emotions are an integral part. Emotional intelligence is what distinguishes us, humans, from animals. Emotion can be defined as “A strong feeling deriving from one’s circumstances, mood, or relationships with others”. The advent of ML and artificial intelligence (AI) has made emotion recognition a very popular topic in the field of brain–computer interface (BCI). Although, there are many biophysical signals given out by a person which can be used for emotion recognition (such as facial expressions, heart rate, perspiration, voice tempo). EEG is the most prominent and promising approach, since signals like facial expressions and change

S. G. Shaila (✉) · B. M. Anirudh · A. S. Nair · L. Monish · P. Murala · A. G. Sanjana
Department of Computer Science Engineering, Dayananda Sagar University, Bangalore, India
e-mail: shaila-cse@dsu.edu.in

L. Monish
e-mail: monishl-cse@dsu.edu.in

Fig. 1 Valence arousal model



in voice tempo can be easily faked leading to wrong interpretation. Electrical activity in the brain cannot be forged or tampered with. Although EEG signals show beyond doubt to be the best for emotion recognition, they are time-dependent and have a high signal-to-noise ratio. They require complex machinery to be recorded. This makes feature extraction for ML and DL applications a tiring task. To avoid this, two feature extraction techniques: Power Spectrum Density (PSD) and Discrete Wavelet Transform (DWT) are to be used. In our paper, we perform a comparative analysis of the above-mentioned feature extraction techniques and their implications in machine learning and deep learning models, namely, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM).

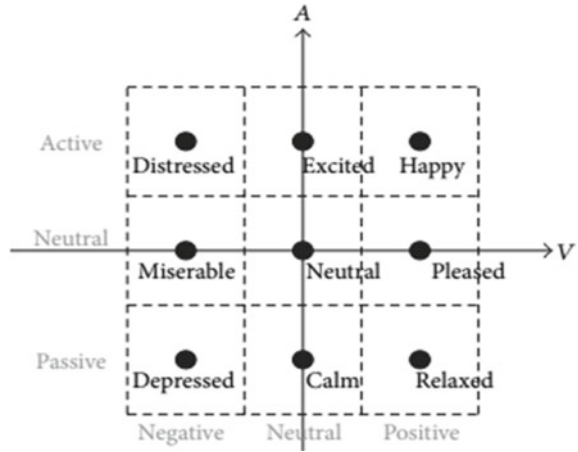
The input variables are EEG signals and the output labels are the emotions generated through the valence–arousal model. The valence–arousal model, shown in Fig. 1, is a two-dimensional model, horizontal axis represents the valence and vertical axis represents arousal, proposed by Russell in 1980 [1]. The measure of degree of aversion or attraction to a stimulus is valence. It ranges from negative to positive. Arousal is a degree of alertness to stimuli or a measure of how awake a person is to stimuli, ranging from passive to active as depicted in Fig. 1.

The valence arousal model allows us to quantify emotions and is an innovative approach to classify emotions. Further research has been done on the model giving us the nine-state model of valence arousal as shown in Fig. 2 [2].

2 Background

The DEAP dataset is a collection of EEG, physiological, and visual signals that can be used to analyze emotions. The description [3] provides a procedure for selecting stimuli, the experimental setup (explanation of the 10–20 system), self-assessment mannequins based upon valence, arousal, familiarity, and dominance ratings, and the

Fig. 2. Nine-state model of valence arousal



correlation between EEG frequencies and participant ratings. The results of [1] are contained in a database to examine spontaneous emotions, which contains psychological signals from 32 participants who watched and assessed their emotional responses to a 40 one-minute films. A brain-computer interface (BCI), commonly known as a brain-machine interface (BMI) or mind-machine interface (MMI), dispenses a non-muscular channel of communication between the human brain and a computer system [4]. BCI deals with how humans and machines interact and interface [5]. The research in this field is responsible for bridging the gap between humans and machines. Brain-computer interface is used to read and interpret signals from the brain, and great success is achieved using this approach in the clinical front.

The brain waves are used to control the actions of robotic arm to help the people affected from stroke and locked-in syndrome allowing them to move their wheel chair or drink coffee from the cup on their own [6]. Dreamer (A Database for Emotion Recognition through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices) dataset is used in [7] by capturing all the signals using portable, low-cost, wireless, wearable, and store brought equipment, that made it easier and shows the potential to use of effective computing methods. RNN, CNN, and GRU algorithms were used in [8] to predict EEG abnormalities. Deep learning algorithms performed well, with an 86.7 percent testing accuracy in predicting EEG abnormalities. Algorithms like Principal Component Analysis for extracting the features and SVM, KNN, and ANN were used for classification in [9]. SVM has 91.3% accuracy on ten channels. In [2], Principal Component Analysis was used to extract the most important features, along with SAE. PCA-based covariate shift adaption boosted the accuracy. The accuracy obtained was 49.52 and 46.03% (valence and arousal). For arousal, the average and maximum classification rates were 55.7% and 67.0%, respectively, and for valence, 58.8 and 56.0%. References [8, 9] use the SVM technique as well. ASP uses data from both the left and right hemispheres of the brain. Valence and arousal had accuracy ratings of 55.0% and 60%, respectively. Artificial neural networks for

emotion recognition through EEG signals were used in [10–12] which use wavelet energy to predict emotions through EEG signals.

3 Methodology

Thirty-two subjects were exposed to 40 one-minute videos and their EEG signals were recorded for 63 s with 3 s of baseline in the DEAP dataset. During the 63 s period, 32 EEG channels and 8 physiological channels captured bio-physical data. Facial videos are also available for the first 22 subjects, but these were not used in this study. The electrode placement technique employed was the 10–20 systems. Subjects rated each video on a scale of 1–9 for valence, arousal, dominance, and familiarity. The signals were denoised and down sampled from 512 to 128 Hz.

In our work, we chose 14 channels: F3, FC5, AF3, F7, T7, P7, O1, O2, P8, T8, F8, AF4, FC6, and F4 that fit the Emotiv Epoch Plus model [4], which is a low-cost off-the-shelf EEG machine, in the hopes that the models developed in the future can be utilized directly in conjunction with the EEG device. To image neuronal activity in the subject’s frontal lobe, the electrodes F3, F4, AF3, AF4, F7, and F8 are used. Using the electrodes T7, T8, FC5, and FC6, the temporal lobes of the brain are imaged and electrodes P8, P7 are used to scan the parietal lobes. The electrodes O2 and O1 are used to acquire the neuronal activity of the occipital lobes. For each trial, we use a 4 s sliding window with a 0.5 s gap and split the 60 s EEG signal into 112 data points. As a result, each patient receives 4480 data points from 40 trails. Table 1 shows the encoding of valence arousal values into nine emotional states. Inter-personal emotional variation is reduced to some extent by encoding states based on range division.

Table 1 Emotional states based on valence arousal rating

Emotional state	Valence range	Arousal range
Depressed	1–3	1–3
Calm	4–6	1–3
Relaxed	7–9	1–3
Miserable	1–3	4–6
Neutral	4–6	4–6
Pleased	7–9	4–6
Distressed	1–3	7–9
Excited	4–6	7–9
Happy	7–9	7–9

3.1 Power Spectral Density

Power Spectral Density provides power in each of the bandwidths based on the frequency. A fundamental computational EEG analysis method that can provide information on power, spatial distribution, or event-related temporal change of a frequency of interest is spectra analysis. PSD of the channels is extracted followed by training of different machine learning models. PSD provides the power in each of the spectral bins by computing the Fast Fourier Transform (FFT) and calculating its complex conjugate.

Mathematically, it is represented in Eq. (1):

$$X(f) = F\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt, \tag{1}$$

where $x(t)$ is the time-domain signal, $X(f)$ is the FFT, and f is the frequency to analyze.

3.2 Discrete Wavelet Transform

Discrete Wavelet Transform helps in extracting features from the time and frequency domains of EEG signal. A wavelet is a wave-like oscillation that is localized in time. A wavelet in a signal is computed from the scale and location. Daubechies 4 (db4) is used as a smoothing feature for identifying the changes in the EEG signals. A low-pass scaling filter and a high-pass wavelet filter are used in DWT to create a filtering mechanism. The lower and higher frequency portions of the signals are separated using this transform decomposition. The lower frequency contents provide a good approximation of the signal, whereas the high-frequency contents contain the finer details of the fluctuation. A measurement criterion of the amount of information within the signal is entropy. It measures only the uncertainty in EEG signal which is similar to the possible configurations or the predictability. The mathematical formula to calculate entropy using DWT is given in Eq. (2):

$$ENT_j = - \sum_{k=1}^N (D_j(k^2)) \log(D_j(k^2)). \tag{2}$$

Wavelet energy reflects the distribution of the principle lines, wrinkles, and ridges in different resolutions. The square of the wavelet coefficients is summed over the temporal window to compute the energy for each frequency band in Eq. (3):

$$ENG_j = - \sum_{k=1}^N (D_j(k^2)) k = 1, 2, \dots, N. \tag{3}$$

Here, the wavelet decomposition level (frequency band) is implied as j and the number of wavelets coefficients within the j frequency band is referred to as k .

3.3 Experimental Setup

The proposed LSTM model has five LSTM layers with 512 nodes in layer1, 256 nodes in layer 2, 128 nodes in layer 3, 64 nodes in layer 4, 32 nodes in layer 5. The final layer is dense layer with nine nodes. The LSTM layers all have tanh activation function, whereas the dense layer has SoftMax activation. After each epoch, batch normalization and a dropout of 0.3 are done to avoid overfitting. RMSprop optimizer is used to avoid vanishing gradient problem (Fig. 3). The model is compiled with mean squared error loss. Figure 4 shows the proposed LSTM model architecture.

In the initial approach, the entire dataset consisting of 32 subjects was combined and PSD and DWT features were extracted. Cross-validation after training and testing gave inconsistent results. This was attributed to the interpersonal emotion variance which results in covariate shift in the dataset, i.e., the intensity at which one person feels for the same emotion is different from other subjects. Hence, the approach was changed to individual subjects' emotion recognition. As shown in Fig. 3, each subject's data were used to extract PSD and DWT features, trained on RBF SVM, RF, KNN, and LSTM models separately. The PSD and DWT were calculated in

Fig. 3 Experimental setup

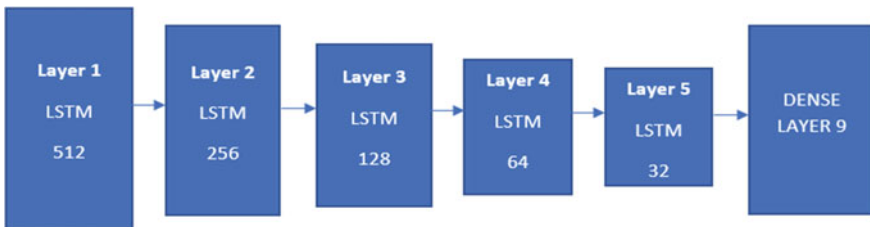
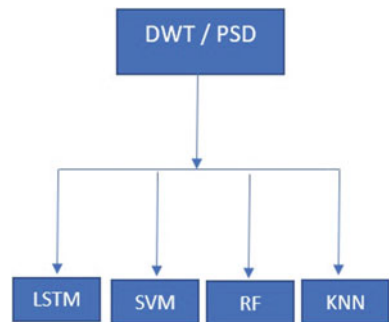


Fig. 4 LSTM model

five frequency bins, namely Delta (4–7 Hz), Theta (8–12 Hz), Alpha (13–16 Hz), Beta (17–25 Hz), and Gamma (26–45 Hz). These bins are associated with the most emotional activity. The window size is chosen as 512 [12]. The first three seconds, which were the baseline seconds, were removed for each subject, after which PSD and DWT were calculated. PSD technique gives a total of 70 features (14 channels \times 5 bandwidths), whereas DWT gives a total of 140 (70 entropy and 70 energy) features. Following feature extraction, a standard scalar was used to perform normalization on the data, and nine-state emotion encoding were done.

4 Experiment and Results

PSD and DWT are calculated and saved subject-by-subject using the parameters provided in Sect. 3. PSD and DWT data are loaded separately for each subject. The labels assigned to the data are then divided into nine emotional states. The data is normalized and divided into test and training groups using conventional methods. Grid search algorithms are used to optimize parameters such as number of estimators in random forest, C in RBF SVM, and number of neighbors in KNN. n estimators = 500, $C = 1e + 10$, and n neighbors = 3 are the best parameters found. With a batch size of 100, the LSTM model was run for 100 epochs. The testing accuracies for all 32 participants for all four models using PSD feature extraction are shown in Table 2. The DWT feature extraction approach is described in Table 3.

The standard deviations and mean accuracies are shown in Table 4. The best mean accuracy and lowest standard deviation are found in the LSTM model with DWT feature extraction. Because DWT includes both time and frequency features, LSTM outperforms other models. With the PSD feature extraction approach, the mean accuracies are 91.15 (random forest), 95.46 (RBF SVM), 92.03 (KNN), and 94.32 (KNN) (LSTM). The mean DWT accuracy is 90.96 (RF), 95.98 (RBF SVM), 93.65 (KNN), and 96.76 (KNN) (LSTM). A bar plot depicts the accuracies of the four different models. The PSD mean accuracies are shown in Fig. 5. The DWT mean accuracies are shown in Fig. 6.

5 Conclusion and Future Work

We present a state-of-the-art result for interpersonal emotion classification using EEG signals in this paper. With only 14 input channels, the LSTM model built can effectively classify the nine fundamental emotions. We may infer from this research that developing a Unified Emotion Recognition model for every human being is unrealistic, and that instead, each person's emotions must be trained separately in order to construct the model. Future research will focus on combining the models into a single algorithm and using soft voting to improve classification and identify the emotion activation mechanism that leads to emotion identification. Also, from

Table 2 Subject-wise PSD-based model accuracies

Subject	RF results	RBF SVM	KNN	Accuracy
1	98.32	99.22	99.11	98.44
2	88.39	95.42	91.85	93.22
3	87.95	97.77	97.1	95.87
4	89.06	92.97	87.61	89.62
5	85.94	88.5	79.13	88.67
6	94.53	98.21	96.54	96.65
7	98.44	99.78	99.11	97.99
8	94.75	95.2	91.52	95.17
9	92.63	95.76	90.63	97.01
10	90.63	93.75	86.5	92.05
11	91.63	95.09	85.6	90.26
12	90.18	95.31	92.97	94.5
13	86.72	92.75	88.95	92.83
14	85.27	95.42	94.64	92.91
15	97.32	98.66	95.54	97.38
16	97.66	98.88	97.99	97.71
17	88.73	94.53	88.73	94.67
18	91.52	97.54	96.21	96.6
19	90.29	94.64	89.96	96.29
20	88.73	94.64	93.3	96.34
21	91.41	96.76	94.98	94.22
22	87.83	91.41	84.26	83.48
23	97.43	98.88	98.33	98.3
24	87.61	91.74	86.16	92.75
25	83.93	92.86	85.27	87.19
26	86.38	92.41	90.29	92.61
27	94.42	96.32	95.76	96.9
28	84.82	94.53	90.85	92.77
29	93.86	96.76	93.42	96.93
30	94.2	94.53	92.63	95.51
31	90.74	96.54	94.98	96.4
32	95.65	97.88	95.31	97.04
Mean	91.15	95.46	92.03	94.32

Table 3 Subject-wise DWT-based model accuracies

Subject	RF result	RBF SVM	KNN	Accuracy
1	96.96	98.37	99.02	98.72
2	88.93	97.72	93.16	96.44
3	84.69	97.39	96.85	97.56
4	89.79	95.01	92.51	94.78
5	85.78	93.27	86.54	94.43
6	93.7	98.81	97.5	97.61
7	98.7	99.02	98.48	98.75
8	93.49	95.98	94.57	96.58
9	93.92	97.61	94.03	98.42
10	90.45	93.27	85.88	95.87
11	92.51	95.77	87.51	95.63
12	88.06	93.16	90.66	95.41
13	86.54	92.62	89.69	95.76
14	87.08	94.57	92.83	95.14
15	96.85	99.24	96.31	98.23
16	97.18	97.39	96.42	97.64
17	88.82	93.38	90.01	96.52
18	88.17	96.09	97.61	97.94
19	89.47	94.68	90.66	96.88
20	87.08	94.46	93.81	97.04
21	89.47	96.09	96.74	96.52
22	85.78	91.97	85.56	91.96
23	96.31	99.24	98.7	98.37
24	90.01	93.81	89.58	96.9
25	87.62	95.77	93.27	95.98
26	89.14	95.87	93.81	96.71
27	92.62	96.2	97.18	97.45
28	86.86	95.55	92.4	95.79
29	95.66	97.61	94.68	98.26
30	91.31	96.96	97.5	97.39
31	92.07	96.2	96.63	97.45
32	95.87	98.37	96.85	98.21
Mean	90.96	95.98	93.65	96.76

Table 4 Mean accuracies and standard deviations

	RF		SVM		KNN		LSTM	
	Accuracy	STD	Accuracy	STD	Accuracy	STD	Accuracy	STD
PSD	91.15	4.76	95.46	2.56	92.03	4.79	94.32	3.44
DWT	90.96	3.86	95.98	2.03	93.65	3.84	96.76	1.44

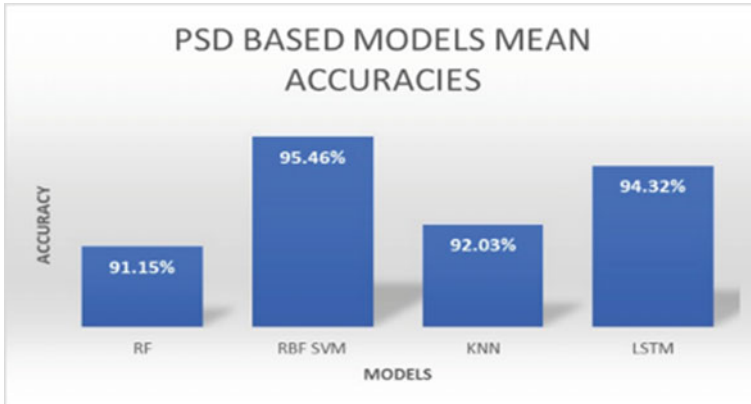


Fig. 5 PSD mean accuracies

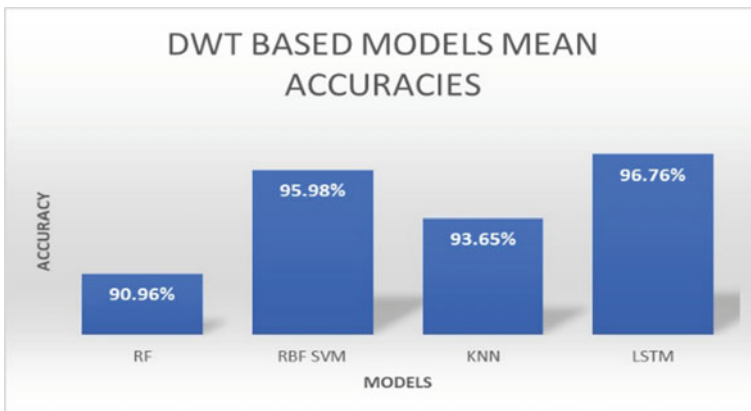


Fig. 6 DWT mean accuracies

collecting raw EEG signals through the Emotiv Epoch Plus machine through emotion recognition and classification, an end-to-end algorithm was developed.

References

1. Russell JA (1980) A circumplex model of affect. *J Pers Soc Psychol* 39(6):1161
2. Jirayucharensak S, Pan-Ngum S, Israsena P (2014) EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation. *Sci World J* 2014
3. Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Patras I (2011) Deap: a database for emotion analysis; using physiological signals. *IEEE Trans Affect Comput* 3(1)
4. Singh Y, A review paper on brain computer interface [Online]. Available: www.ijert.org
5. Katsigiannis S, Ramzan N (2017) DREAMER: a database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. *IEEE J Biomed Health Inform* 22(1):98–107
6. Roy S, Kiral-Kornek I, Harrer S (2019) ChronoNet: a deep recurrent neural network for abnormal EEG identification. In: *Conference on artificial intelligence in medicine in Europe*. Springer, Cham, pp 47–56
7. Huang D, Guan C, Ang KK, Zhang H, Pan Y (2012) Asymmetric spatial pattern for EEG-based emotion detection. In: *The 2012 International joint conference on neural networks (IJCNN)*. IEEE, pp 1–7
8. Koelstra S et al. Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos
9. Purnamasari PD, Ratna AAP, Kusumoputro B (2016) Artificial neural networks based emotion classification system through relative wavelet energy of EEG signal. In: *Proceedings of the Fifth International conference on network, communication and computing*, pp 135–139
10. Mohammadi Z, Frounchi J, Amiri M (2017) Wavelet-based emotion recognition system using EEG signal. *Neural Comput Appl* 28(8):1985–1990
11. Bousseta R, Tayeb S, El Ouakouak I, Gharbi M, Regragui F, Himmi MM (2016) EEG efficient classification of imagined hand movement using RBF kernel SVM. In: *2016 11th International conference on intelligent systems: theories and applications (SITA)*. IEEE, pp 1–6
12. Rozgić V, Vitaladevuni SN, Prasad R (2013) Robust EEG emotion classification using segment level decision fusion. In: *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, pp 1286–1290