




CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings

Abhishek Iyer¹ · Srimit Sritik Das¹ · Reva Teotia¹ · Shishir Maheshwari² · Rishi Raj Sharma³ 

Received: 31 March 2021 / Revised: 25 August 2021 / Accepted: 17 January 2022 /
Published online: 11 April 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Emotion is a significant parameter in daily life and is considered an important factor for human interactions. The human-machine interactions and their advanced stages like humanoid robots essentially require emotional investigation. This paper proposes a novel method for human emotion recognition using electroencephalogram (EEG) signals. We have considered three emotions namely neutral, positive, and negative. These EEG signals are separated into five frequency bands according to EEG rhythms and the differential entropy is computed over the different frequency band components. The convolution neural network (CNN) and long short-term memory (LSTM) based hybrid model is developed for accurate emotion detection. Further, the extracted features are fed to all three models for emotion recognition. Finally, an ensemble model combines the predictions of all three models. The proposed approach is validated on two datasets namely SEED and DEAP for EEG based emotion analysis. The developed method achieved 97.16% accuracy on SEED dataset for emotion classification. The experimental results indicate that the proposed approach is effective and yields better performance than the compared methods for EEG-based emotion analysis.

Keywords Emotion recognition · EEG · Hybrid model · Differential entropy · LSTM

1 Introduction

Emotion is a pivotal factor in human life as it affects the working ability, mental state, and judgment of human beings. Numerous experts worked on this topic in different disciplines like psychology, cognitive science, neuroscience, computer technology, brain-computer interfacing (BCI) [38], and others. The electroencephalogram (EEG) based emotion recognition has created a lot of scope in these disciplines using the actual affective state [20]. BCI has numerous applications of EEG-based emotion recognition including humanoid robots. Most of the humanoid robots lack in terms of emotion and this field is not much explored. In

✉ Rishi Raj Sharma
dr.rsrss@gmail.com

effective BCI, emotion recognition is a major parameter and it becomes complex due to its fuzzy property. Human emotion correlates with context, language, time, space, culture, and other components. Therefore, the absolute true labels are also not possible with different emotions with EEG recording, which creates the issue [3].

Many authors proposed facial expression [51], gesture [9, 30], posture, speech [27], and other physical signal-based emotion recognition methods. These types of data are easy to record but can be easily controlled and falsify the true emotion [15]. The controlling or mimicking of nervous system related signals is very tough as they are involuntarily activated [39] and only subject experts can control these signals. Therefore, true emotion signature can be observed in nervous system related recordings. Several physiological recordings such as EEG, electrocardiogram (ECG) [34], temperature, electromyogram (EMG) [5], respiratory, galvanic skin response (GSR) can be used to study the human emotion [39]. The minute investigation of brain activity with various emotions can assist the accurate and computation efficient emotion recognition models. Recent research in dry electrode implementation [11, 13] and wearable devices promote the EEG recording based emotion identification in real-time accomplishment for mental state monitoring [23, 35]. EEG based emotion recognition is one of the key feature required in human-machine interaction (HMI) and a humanoid robot. This study is focused on EEG based human emotion analysis in which the electrical activity of the brain is investigated during different emotions (neutral, positive, and negative emotions).

Many studies are performed for EEG based human emotion diagnosis and tried to form a definitive relationship of EEG signals with different emotions [28, 33]. EEG signal analysis is a very challenging task as it is non-stationary [36]. In a real-time scenario, other signals are added into EEG recording and signal to noise ratio (SNR) becomes low. Matrix decomposition-based EEG signal analysis methods are proposed but due to high complexity, real-time implementation is tough [7, 37, 41]. The emotion related stable patterns of EEG recordings are observed in [55] which uses the DEAP dataset. The critical frequencies of emotion and significant channel selection in EEG recordings are micro-observed in [54]. This channel selection is good to find the position of the electrode for emotion analysis. The time-frequency based and various non-linear features are studied for EEG based emotion recognition and achieve 59.06% accuracy (ACC) with DEAP data [20]. It is suggested that the gamma-band in EEG recordings is more correlated with the emotion function [18].

Many machine learning-based architectures are proposed for EEG based emotion examination. The bi-hemisphere based neural network is designed for EEG emotion detection, and the experiment is performed on SEED dataset with 63.50% ACC [21]. The graph neural network-based emotion recognition is performed on the same dataset with 89.23% ACC using gamma band, and 94.24% ACC with all bands [56]. A regional asymmetric convolution neural network (CNN) based study is carried on DEAP data and acquire 95% ACC for arousal and valence emotion detection [6]. In most of the methods, existing models are improved to achieve a good classification of human emotion. The proposed approach employs multiple models and develops a hybrid approach to attain better ACC than the developed methods. The two models, based on CNN and long short term memory (LSTM), are hybridized to improve the final prediction using the ensemble model.

Rest of the article is organized in the following manner. Section 2, presents the dataset. Proposed approach with features is explained in Section 3. The proposed hybrid model along with the CNN and LSTM based models are presented in Section 4. This also includes the implementation of ensemble learning. Results are explained in Section 5. Finally, the article is concluded in Section 6.

2 Dataset

We have used two datasets for EEG based emotion recognition. The detailed explanation of both the datasets is given next to this section.

2.1 SEED data

The database employed in the proposed approach has been obtained from the brain like computing and machine learning (BCMI) methods. We employed the SJTU emotion EEG dataset (SEED) [8, 54]. The dataset contains EEG data of 15 subjects (7 males and 8 females) recorded in three separate sessions, each session having 15 trials. In each trial, the EEG signal is recorded when the subject is watching Chinese film clips with three types of emotions, namely positive, neutral, and negative. The duration of each film clip is about 4 minutes and two film clips targeting the same emotion are not shown consecutively. The participants reported their emotional reactions to each film clip by completing the questionnaire immediately after watching each film clip. The EEG signals are recorded using a 62-channel electrode cap according to the international 10-20 system. The data is then down-sampled to 200Hz to make system faster and a band-pass frequency filter from 0-75Hz is applied which contains all the EEG rhythm information.

2.2 DEAP data

The DEAP data has been recorded for the analysis of human emotion using EEG signals. It is recorded for 32 healthy participants aged between 19 years to 37 years and out of 32 participants 16 were female. Each participant has been exposed to 40 music videos each of which has a duration of 1-min with the same emotion throughout the video length. The data comprises 40 channels out of which 32 EEG channels have been investigated in this paper. The data is recorded with Biosem ActiveTwo devices at a sampling rate of 512 Hz. It is further downsampled to 128 Hz to reduce the system complexity. The DEAP data provides 32 files, where each file contains the 40-channel EEG recording of 40 videos of one minute duration each.

3 Proposed approach

Block diagram of the proposed approach is shown in Fig. 1. All the subjects were made to sit on a chair in the resting state and are asked to watch the videos portraying different emotions. Simultaneously, EEG signals are recorded and pre-processing is done. The differential entropy (DE) based features are computed in five EEG rhythms. The DE based features are explained in the next sub-section. Further, CNN and LSTM models are employed and combined to obtain the hybrid model. Thereafter, the ensemble model is proposed based on these models.

3.1 Features extraction

We have employed DE as a feature in the proposed approach. DE extends the idea of Shannon entropy and is used to measure the complexity of a continuous random variable. DE as a feature was first introduced to EEG-based emotion recognition by Duan et al. [8]. It has been found to be more suited for emotion recognition than the traditional feature. DE has

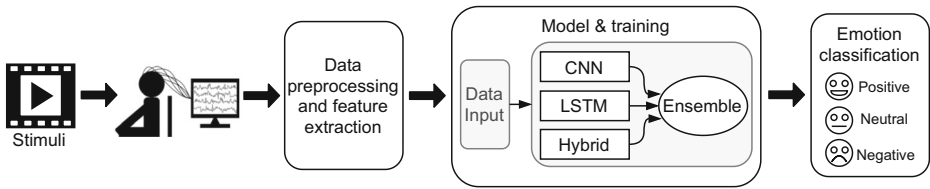


Fig. 1 Block diagram of the proposed system for EEG-based emotion recognition

the balanced ability to discriminate EEG patterns between low and high-frequency energy. DE feature extracted from EEG data provides stable and accurate information for emotion classification [53]. The differential entropy feature is as defined below:

$$\begin{aligned}
 h(Y) &= - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}\right) dy \\
 &= \frac{1}{2} \log(2\pi e\sigma^2)
 \end{aligned} \quad (1)$$

where the time series Y obeys the Gaussian distribution $N(\mu, \sigma^2)$. stop DE was employed to construct features in the five frequency bands: delta (1–3Hz), theta (4–7Hz), alpha (8–13Hz), beta (14–30Hz), and gamma (31–50Hz). For the SEED dataset, the extracted DE feature for a sample EEG signal has 310 dimensions as there are 62 channels for each frequency band [54]. Similarly, the 32 channels are considered for DEAP dataset in five EEG sub-bands which leads to total 160 DE features.

Various models along with the hybrid model and ensemble model are explained in next section.

4 Model employed for emotion recognition in the proposed system

In the proposed work, initially, the CNN and LSTM based models are employed for emotion recognition. Thereafter, a hybrid model is proposed which is a combination of CNN and LSTM. Finally, an ensemble model of these three proposed models is taken into consideration. All these models are explained in this section.

4.1 CNN-based model

The idea behind CNNs bears a resemblance to traditional artificial neural networks (ANNs), consisting of neurons that self-optimize through learning. CNN's are powerful performers on large sequential data represented by matrices such as images broken down to their pixel values [45]. A smaller $n \times n$ kernel slides over the entire feature matrix performing convolutions over the superposed space [12]. The feature map size can be kept consistent across multiple convolutions using padding of 0s. However, functions like Max Pooling are employed to reduce the amount of computational data and still retain the important information [26]. As the feature maps pass through the different convolutional layers, the filters learn to detect patterns and more abstract features.

EEG based emotion classification using the CNN method was also explored in the approaches of [46]. Cascade and parallel convolutional recurrent neural networks have been

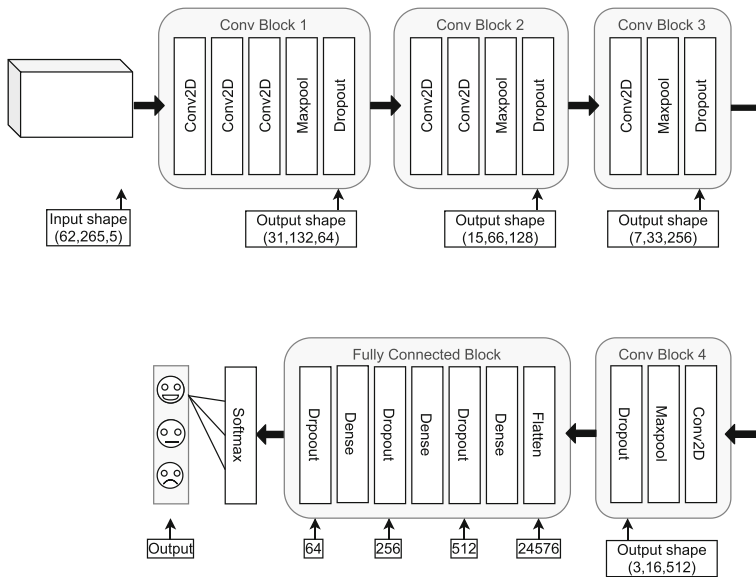


Fig. 2 CNN-based model architecture employed in the proposed approach for EEG-based emotion recognition

used for EEG human-intended movement classification tasks [52]. Additionally, before applying the CNN, EEG data could be converted to image representation after feature extraction [42]. However, the accuracy of emotion recognition by using only CNN is not high.

The details of the CNN architecture employed in the proposed approach are shown in Fig. 2: The CNN model consists of four convolutional (conv) blocks with 64, 128, 256, 512 filters, respectively. The kernel size of conv filters is 5×5 and 3×3 . All the layers use padding, followed by maximum sub-sampling layers, which operate over 2×2 sub-windows at each conv layer, known as the Max Pooling layers. The network ends with three fully connected dense layers fed to the c-way softmax [24] classification layer. Relu activation is employed due to its unity gradient, where the maximum amount of error is passed during back-propagation [1]. Dropout regularization is used after every layer which improves the performance of the model via a modest regularization effect [29]. Thereafter, the predictions of the CNN model are fed to the proposed ensemble model for emotion recognition.

4.2 LSTM-based model

The LSTM networks are modified recurrent neural networks (RNN), capable of learning long-term dependencies. LSTM network is parametrized by weight matrices from the input and the previous state for each of the gates, in addition to the memory cell, which overcomes the issue of vanishing/exploding gradient [10].

We use the standard formulation of LSTMs with the logistic function (σ) [4] on the gates and the hyperbolic tangent [2] on the activations. The input is of the shape 1325×62 . The model has 4 LSTM layers with dropouts in between, and then the output is passed to the fully connected network. SoftMax activation function [24] is used to predict the final output. The block diagram of the LSTM architecture is shown in Fig. 3.

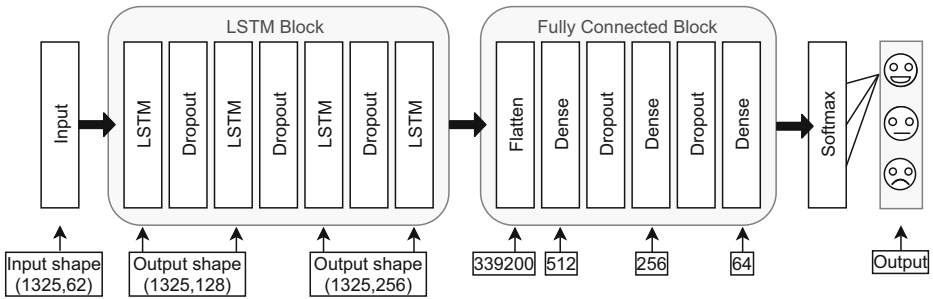


Fig. 3 LSTM-based architecture employed in the proposed approach for EEG-based emotion recognition

4.3 Hybrid model

The hybrid model combines more than one base model in series. Figure 4 shows the structure of the hybrid model employed in the proposed approach. The hybrid model improves the performance by capturing more information that is left undetected previously.

The first three blocks of the hybrid model consist of convolutional (conv) blocks. The conv block consists of max pool layers and the Dropout regularisation to avoid overfitting [29]. The output shape of the third conv block is $15 \times 66 \times 512$. On the other hand, the input shape to the LSTM block is 66×7680 . The reshape layer is employed between the conv and LSTM block to facilitate this dimensional mismatch. In general, 2D conv block work on inputs which are \mathbb{R}^3 , while LSTM inputs are in \mathbb{R}^2 . The LSTM network uses the Tanh activation function [2] and batch norm regularization [47]. The output of the LSTM block is passed to a fully connected network that uses softMax [24] to calculate the probabilities of the output.

4.4 Ensemble learning-based model

Ensemble learning is mainly of two types, namely, homogeneous and heterogeneous. It combines the prediction from multiple models and integrates the individual strengths of the base models. This results in the robustness and the improved performance of the overall approach [50]. Ensemble learning is homogeneous when the base models are of the same type. In the proposed approach, ensemble learning is heterogeneous as the base models are different.

Once these models are trained, a statistical method is used to combine the predictions of the different models. The statistical method involves the methods of bagging, boosting, and stacking. We have employed stacking as it is suitable for heterogeneous ensemble models [43]. Stacking is the process in which separate models learn parallelly on the dataset and a small meta model, usually a feed-forward neural network (FNN) is used to combine individual predictions and come up with the final outputs. Stacking introduces a meta-model that receives the different predictions of the base models as its input. The meta-model [48] learns to maximize the output prediction, and this becomes our final output. In addition to stacking, we have also investigated the max function as a statistical method to combine the predictions. Figure 5 shows the block diagram of ensemble model. The meta model used in the stacking method consists of 4 fully connected (FC) layers followed by a softmax classifier [24].

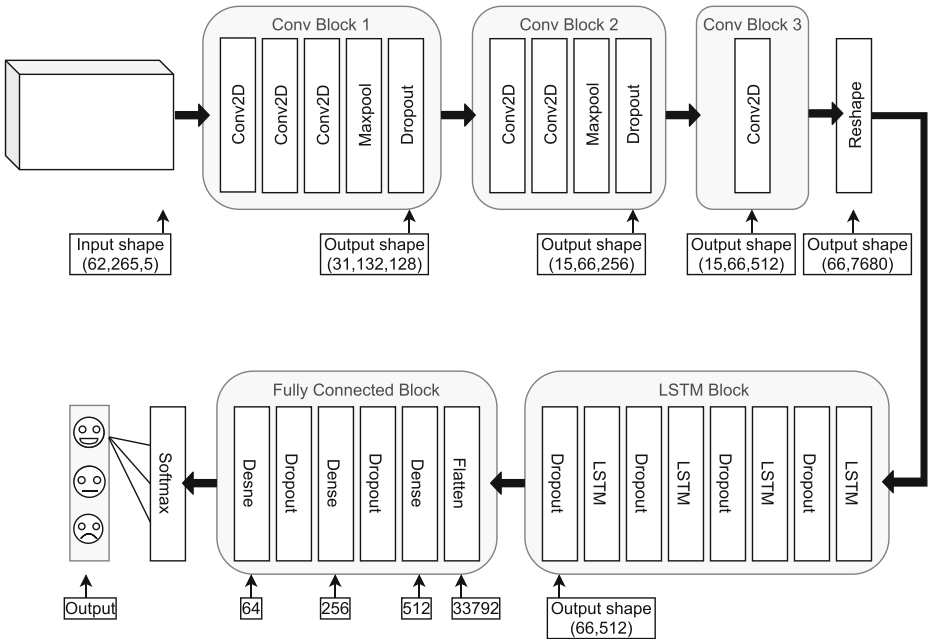


Fig. 4 Hybrid model employed in the proposed approach for EEG-based emotion recognition

5 Results & discussion

The proposed approach has been evaluated on two datasets namely SEED and DEAP. In the proposed approach, the performance of various models has been investigated using the k-fold cross-validation test [32] with $k = 10$. The individual performances of the CNN, LSTM and the hybrid model have been obtained. Further, we have also obtained the performance of the ensemble model. Each model is trained for 60 epochs with a batch size

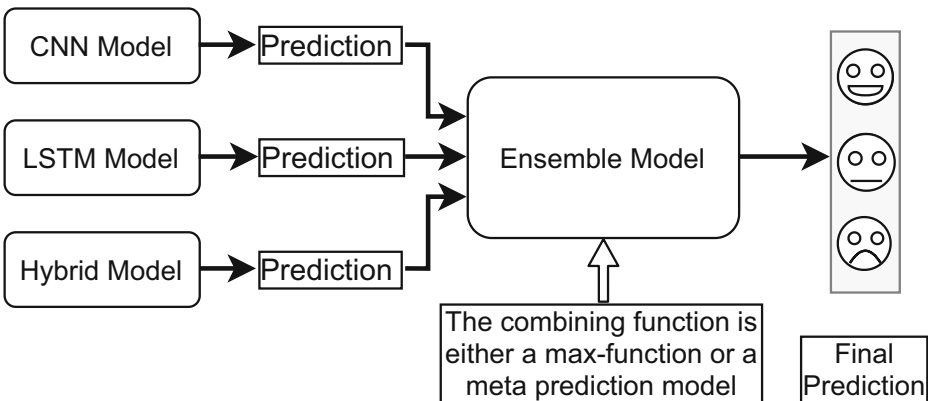


Fig. 5 Ensemble model employed in the proposed approach for EEG-based emotion recognition

Table 1 Classification performance of individual and ensemble model for EEG-based emotion recognition on SEED data

Model	MCC	WAP(%)	WAS(%)	WAF1(%)	Avg ACC (%)	STD (%)
CNN-based	0.819	89	88	88.49	89.53	1.78
LSTM-based	0.821	89	89	89.00	89.99	1.92
Hybrid	0.867	94	93	93.49	93.46	1.36
Ensemble	0.90	97	97	97.00	97.16	1.08

of 64. The learning rate (LR) has not been fixed due to saturation in loss which results in no further improvement in the performance of the model. To overcome this limitation, we have employed LR annealer which makes the learning rate a variable parameter. It should be noted that we have used same feature and experimental setup for both the datasets.

5.1 Experimental results for SEED data

The performance of individual models are measured by evaluating certain parameters such as weighted average precision (WAP), weighted average sensitivity (WAS), and weighted average F1 score (WAF1). F1 score is a good metric to check stability of the model. Table 1 tabulates the performance parameters of individual models, hybrid model and ensemble model for EEG emotion recognition.

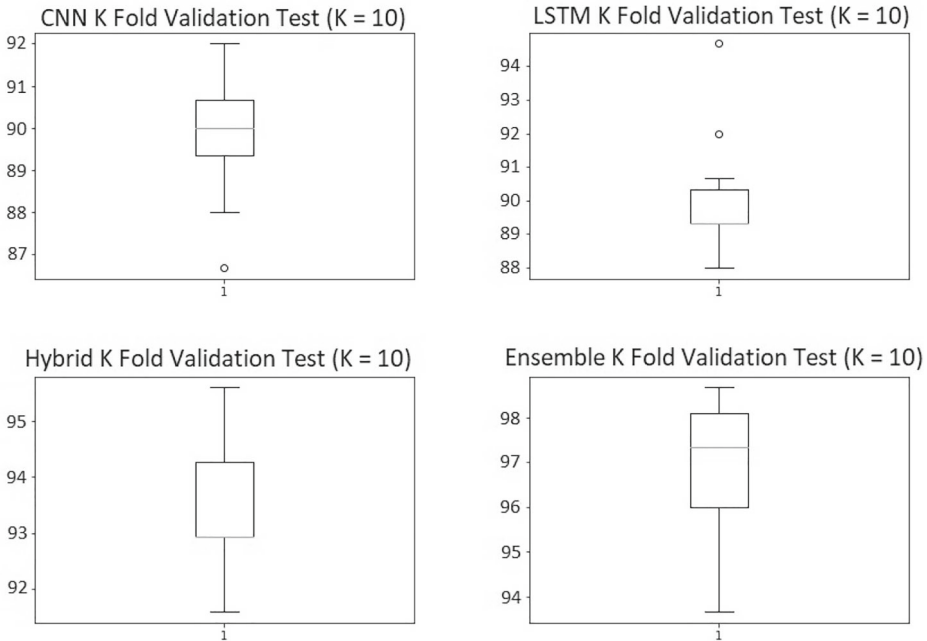


Fig. 6 The box and Whisker plots for ACC achieved by proposed (top-left) CNN-based model, (top-right) LSTM-based model, (bottom-left) Hybrid-model, (bottom-right) Ensemble stack model for EEG-based emotion recognition

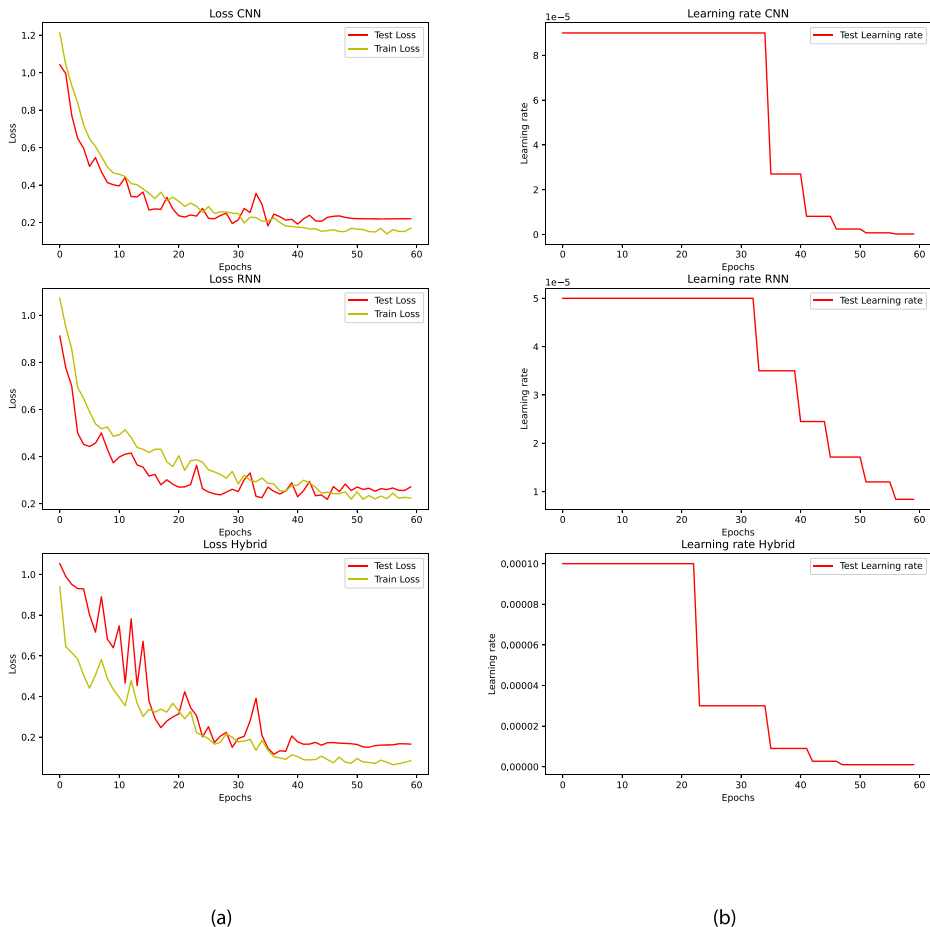


Fig. 7 (a) Loss (b) Learning rate of the proposed CNN based model, RNN based model, and Hybrid model on SEED data

The experimental results suggest that the CNN and LSTM model individually achieves the classification accuracy (ACC) of 89.53% and 89.99%, respectively. The hybrid model achieves an ACC of 93.46%. On the other hand, the ensemble model achieves the ACC of 97.16% for the stack-based ensemble learning. The results for SEED data are tabulated in Table 1. From Table 1, it can be noticed that the ensemble-based method provides improved performance over other models. We believe that this is because the base models are not weak and provide good accuracy by themselves (Fig. 6).

Figure 7 shows the plot of loss function and LR with respect to epoch. When LR saturates after some epochs, there is no significant decreases in loss which results in poor model performance. On the other hand, as we decrease the RL when loss saturates, the loss tends to settle more quickly and improves system performance.

We have also shown Box and Whisker plots of ACC to shed some more light on the results in Fig. 6. The inter-quartile range (IQR) is indicated with the box and the orange line showing where the median lies. This includes all the results from 25 percentile to 75 percentile. The minimum and maximum values are marked by the solid black line at the top

Table 2 Comparison of previous benchmark methodologies for EEG-based emotion recognition on SEED data

Paper	Model	Feature	ACC (%)	STD (%)
He Li et al. [17]	DAN	DE	83.81	8.56
Wang et al. [49]	PGCNN	RASM	84.35	10.28
T. Song et al. [40]	DGCNN	DE	90.40	8.49
Hwang et al. [14]	CNN	TP-DE (cubic)	90.41	8.71
Wei Liu et al. [25]	BDAE	DE	91.01	8.91
W Zheng et al. [53]	GELM	DE	91.07	7.54
Hao Tang et al. [44]	Bimodal-LSTM	PSD+DE	93.97	–
JL Qiu et al. [31]	DCCA	–	94.58	6.16
Liu J et al. [22]	CNN+SAE+DNN	PCC	96.77	–
Proposed approach	Ensemble(stack)	DE	97.16	1.08

and bottom of the box and whisker plot. The outliers, marked by circles, are results that did not fall in the whisker range, which contains results in the range of $1.5 \times \text{IQR}$.

We further compare our experimental results of the proposed approach with some of the past benchmark methodologies on emotion recognition on the SEED dataset. Table 2 tabulates the comparison of the proposed approach with other past benchmark results. It can be observed that the proposed approach outperforms the previous methodologies. It can also be noticed that the standard deviation (STD) of the proposed approach is very less as compared with other approaches tabulated in Table 2. This also reflects the repeatability and reproducibility of the proposed approach.

5.2 Experimental results for DEAP data

We have also employed DEAP dataset for evaluating the performance of proposed approach for EEG-based emotion analysis with same feature and experimental setup. The performance of the proposed approach on DEAP dataset has been tabulated in Table 3. It can be observed from Table 3 that the ensemble obtains maximum performance as compared to other individual models. The CNN-based, LSTM-based and hybrid models achieve classification performances of 63.50%, 63.89%, 64.02%, respectively. Tables demonstrate that the ensemble model achieves better performance than individual models. The performance of other existing works on DEAP data with same DE feature has been compared in Table 4. It can be observed from Table 4 that the proposed system attains better performance than the existing methods for EEG-based emotion recognition.

Table 3 Classification performance of individual and ensemble model for EEG-based emotion recognition on DEAP data

Classification Methods	WAP(%)	WAS(%)	WAF1(%)	Avg ACC (%)	STD (%)
CNN	64	74	68	63.50	3.89
LSTM	64.3	74	68	63.89	3.78
Hybrid	64	75	69	64.02	3.84
Ensemble	65	75	70	65.00	3.57

Table 4 Comparison of previous benchmark methodologies for EEG-based emotion recognition on DEAP data

Paper	Classification Methods	Feature	Avg ACC (%)	STD (%)
Lan et al. [16]	logistic regression	DE	48.93	15.50
Li et al. [19]	SVM	DE	57.00	0.3
Proposed approach	Ensemble(stack)	DE	65.00	3.57

For the future work, we planned to extent our work to propose new feature for the effective emotion recognition from EEG signal. Also, the SEED and DEAP datasets will be evaluated with new features to further improve the existing performances. We also intend to test the proposed model for other EEG-based neuronal system development.

6 Conclusion

This paper proposes the ensemble learning-based EEG emotion recognition system. Firstly, the differential entropy was extracted from different frequency bands of EEG signals. Thereafter, these features are fed to CNN and LSTM based models. The hybrid model is developed by combining the sub-blocks of CNN and LSTM models. The ensemble model is proposed based on the CNN, LSTM, and hybrid model. The experimental results suggest that the ensemble model achieves better classification performance than the other models employed in the proposed approach. The proposed ensemble model outperforms the compared methodologies with 97.16% ACC for EEG-based emotion recognition on SEED dataset. The proposed method is also evaluated on DEAP dataset and obtains 65% ACC using same features and model parameters. All the models provided impressive accuracy individually and showed a much lower standard deviation.

BCI is an upcoming field that is highly reliant on the accurate, repeatable, and efficient classification of our brain waves frequently recorded by EEG methods. The experimental results suggest that the proposed approach is suitable for this purpose and paves the way for upcoming research fields of such as humanoid robots, sophisticated prosthetics, and AI-assisted healthcare and recovery. In future, a hardware implementation can be done for the proposed model.

References


1. Agarap AF (2019) Deep learning using rectified linear units (relu)
2. Anastassiou GA (2011) Multivariate hyperbolic tangent neural network approximation. *Comput Math Applic* 61(4):809–821
3. Bos DO et al (2006) Eeg-based emotion recognition. *The Influence of Visual and Auditory Stimuli* 56(3):1–17
4. Chen Z, Cao F, Hu J (2015) Approximation by network operators with logistic activation functions. *Appl Math Comput* 256:565–571
5. Cheng B, Liu G (2008) Emotion recognition from surface emg signal using wavelet transform and neural network. In: *Proceedings of the 2nd international conference on bioinformatics and biomedical engineering (ICBBE)*, pp 1363–1366
6. Cui H, Liu A, Zhang X, Chen X, Wang K, Chen X (2020) Eeg-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. *Knowl-Based Syst* 205:106243
7. Dash M, Liu H (1997) Feature selection for classification. *Intell Data Anal* 1(1–4):131–156

8. Duan RN, Zhu J, Lu B (2013) Differential entropy feature for EEG-based emotion classification. In: 6th International IEEE/EMBS conference on neural engineering (NER), pp 81–84. IEEE
9. Glowinski D, Camurri A, Volpe G, Dael N, Scherer K (2008) Technique for automatic emotion recognition by body gesture analysis. In: 2008 IEEE Computer society conference on computer vision and pattern recognition workshops, pp 1–6. IEEE
10. Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J (2017) Lstm: a search space odyssey. *IEEE Trans Neural Netw Learn Syst* 28(10):2222–2232
11. Grozea C, Voinescu CD, Fazli S (2011) Bristle-sensors—low-cost flexible passive dry eeg electrodes for neurofeedback and bci applications. *J Neur Eng* 8(2):025008
12. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, Chen T (2018) Recent advances in convolutional neural networks. *Pattern Recogn* 77:354–377
13. Huang YJ, Wu CY, Wong AMK, Lin BS (2014) Novel active comb-shaped dry electrode for eeg measurement in hairy site. *IEEE Trans Biomed Eng* 62(1):256–263
14. Hwang S, Hong K, Son G, Byun H (2020) Learning cnn features from de features for eeg-based emotion recognition. *Pattern Anal Applic* 23(3):1323–1335
15. Kurbalija V, Ivanović M, Radovanović M, Geler Z, Dai W, Zhao W (2018) Emotion perception and recognition: an exploration of cultural differences and similarities. *Cogn Syst Res* 52:103–116
16. Lan Z, Sourina O, Wang L, Scherer R, Müller-Putz GR (2019) Domain adaptation techniques for eeg-based emotion recognition: a comparative study on two public datasets. *IEEE Trans Cogn Develop Syst* 11(1):85–94
17. Li H, Jin YM, Zheng W, Lu B (2018) Cross-subject emotion recognition using deep adaptation networks. In: *Neural information processing*, pp 403–413. Springer International Publishing
18. Li M, Lu B (2009) Emotion classification based on gamma-band eeg. In: 2009 Annual international conference of the IEEE engineering in medicine and biology society, pp 1223–1226. IEEE
19. Li P, Liu H, Si Y, Li C, Li F, Zhu X, Huang X, Zeng Y, Yao D, Zhang Y, Xu P (2019) Eeg based emotion recognition by combining functional connectivity network and local activations. *IEEE Trans Biomed Eng* 66(10):2869–2881
20. Li X, Song D, Zhang P, Zhang Y, Hou Y, Hu B (2018) Exploring eeg features in cross-subject emotion recognition. *Front Neurosci* 12:162
21. Li Y, Zheng W, Zong Y, Cui Z, Zhang T, Zhou X (2018) A bi-hemisphere domain adversarial neural network model for eeg emotion recognition. *IEEE Transactions on Affective Computing*
22. Liu J, Wu G, Luo Y, Qiu S, Yang S, Li W, Bi Y (2020) Eeg-based emotion classification using a deep neural network and sparse autoencoder. *Front Syst Neurosci* 14:43
23. Liu NH, Chiang CY, Hsu HM (2013) Improving driver alertness through music selection using a mobile eeg to detect brainwaves. *Sensors* 13(7):8199–8221
24. Liu W, Wen Y, Yu Z, Yang M (2017) Large-margin softmax loss for convolutional neural networks
25. Liu W, Zheng W, Lu BL (2016) Multimodal emotion recognition using multimodal deep learning
26. Manli S, Song Z, Jiang X, Pan J, Pang Y (2016) Learning pooling for convolutional neural network. *Neurocomputing*, 224
27. Mariooryad S, Busso C (2014) Compensating for speaker or lexical variabilities in speech for emotion recognition. *Speech Comm* 57:1–12
28. Mathersul D, Williams LM, Hopkinson PJ, Kemp AH (2008) Investigating models of affect: relationships among eeg alpha asymmetry, depression, and anxiety. *Emotion* 8(4):560
29. Park S, Kwak N (2017) Analysis on the dropout effect in convolutional neural networks, pp 189–204
30. Piana S, Staglianò A, Odone F, Camurri A (2016) Adaptive body gesture representation for automatic emotion recognition. *ACM Trans Interact Intell Syst (TiIS)* 6(1):1–31
31. Qiu JL, Liu W, Lu B (2018) Multi-view emotion recognition using deep canonical correlation analysis. In: *Neural information processing*, pp 221–231. Springer International Publishing
32. Rodriguez JD, Perez A, Lozano JA (2010) Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Trans Pattern Anal Mach Intell* 32(3):569–575
33. Sammler D, Grigutsch M, Fritz T, Koelsch S (2007) Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology* 44(2):293–304
34. Sarkar P, Etemad A (2020) Self-supervised eeg representation learning for emotion recognition. *IEEE Transactions on Affective Computing*
35. Sauvet F, Bougard C, Coroenne M, Lely L, Van Beers P, Elbaz M, Guillard M, Leger D, Chennaoui M (2014) In-flight automatic detection of vigilance states using a single eeg channel. *IEEE Trans Biomed Eng* 61(12):2840–2847
36. Sharma R, Sahu SS, Upadhyay A, Sharma RR, Sahoo AK (2021) Sleep stage classification using DWT and dispersion entropy applied on EEG signals. In: *Computer-aided design and diagnosis methods for biomedical applications*, pp 35–56. CRC Press

37. Sharma RR, Pachori RB (2017) Time-frequency representation using IEVDHM-HT with application to classification of epileptic EEG signals. *IET Sci Measur Technol* 12(1):72–82
38. Sharma S, Sharma RR (2022) Variational mode decomposition based finger flexion movement detection using ECoG signals. In: *Artificial intelligence-based brain-computer interface*, pp 101–119. Elsevier
39. Shu L, Xie J, Yang M, Li Z, Li Z, Liao D, Xu X, Yang X (2018) A review of emotion recognition using physiological signals. *Sensors* 18(7):2074
40. Song T, Zheng W, Song P, Cui Z (2020) Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Trans Affect Comput* 11(3):532–541
41. Subasi A, Gursoy MI (2010) Eeg signal classification using pca, ica, lda and support vector machines. *Exp Syst Applic* 37(12):8659–8666
42. Tabar YR, Halici U (2016) A novel deep learning approach for classification of EEG motor imagery signals. *J Neur Eng* 14(1):016003
43. Tahir MA, Kittler J, Bouridane A (2012) Multilabel classification using heterogeneous ensemble of multi-label classifiers. *Pattern Recogn Lett* 33(5):513–523
44. Tang H, Liu W, Zheng W, Lu B (2017) Multimodal emotion recognition using deep neural networks. In: *Neural information processing*, pp 811–819. Springer International Publishing
45. Teuwen J, Moriakov N (2020) Chapter 20 - convolutional neural networks. In: *Handbook of medical image computing and computer assisted intervention*, the Elsevier and MICCAI society book series, pp 481–501. Academic Press
46. Tripathi S, Acharya S, Sharma RD, Mittal S, Bhattacharya S (2017) Using deep and convolutional neural networks for accurate emotion classification on deep dataset. In: *Proceedings of the thirty-first AAAI conference on artificial intelligence*, pp 4746–4752. AAAI Press
47. van Laarhoven T (2017) L2 regularization versus batch and weight normalization
48. Wang YX, Ramanan D, Hebert M (2019) Meta-learning to detect rare objects. In: *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*
49. Wang Z, Tong Y, Heng X (2019) Phase-locking value based graph convolutional neural networks for emotion recognition. *IEEE Access* 7:93711–93722
50. Webb GI, Zheng Z (2004) Multistrategy ensemble learning: reducing error by combining ensemble learning techniques. *IEEE Trans Knowl Data Eng* 16(8):980–991
51. Young AW, Rowland D, Calder AJ, Etcoff NL, Seth A, Perrett DI (1997) Facial expression megamix: tests of dimensional and category accounts of emotion recognition. *Cognition* 63(3):271–313
52. Zhang D, Yao L, Zhang X, Wang S, Chen W, Boots R, Benatallah B (2018) Cascade and parallel convolutional recurrent neural networks on eeg-based intention recognition for brain computer interface. In: *Proceedings of the thirty-second AAAI conference on artificial intelligence*, pp 1703–1710. AAAI Press
53. Zheng W, Zhu J, Peng Y, Lu B (2014) EEG-based emotion classification using deep belief networks. In: *2014 IEEE International conference on multimedia and expo (ICME)*, pp 1–6
54. Zheng W, Lu B (2015) Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Trans Auton Ment Dev* 7(3):162–175
55. Zheng W, Zhu J, Lu B (2017) Identifying stable patterns over time for emotion recognition from eeg. *IEEE Trans Affect Comput* 10(3):417–429
56. Zhong P, Wang D, Miao C (2020) EEG-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Abhishek Iyer¹ · Srimit Sritik Das¹ · Reva Teotia¹ · Shishir Maheshwari² · Rishi Raj Sharma³ 

Abhishek Iyer
f20181105@pilani.bits-pilani.ac.in

Srimit Sritik Das
f20180527@pilani.bits-pilani.ac.in

Reva Teotia
f20190268@pilani.bits-pilani.ac.in

Shishir Maheshwari
shishir.maheshwari@vit.ac.in

¹ Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, 333031, India

² School of Electronics Engineering (SENSE), Vellore Institute of Technology (VIT), Chennai, 600127, India

³ Department of Electronics Engineering, Defence Institute of Advanced Technology, Pune, 411025, India