

EEG-based Emotion Recognition using Transfer Learning Based Feature Extraction and Convolutional Neural Network

Vaibhav Jadhav^{1*}, Namita Tiwari¹ and Meenu Chawla¹

¹Maulana Azad National Institute of Technology, Bhopal, India

Abstract. In this paper, a novel method for EEG(Electroencephalography) based emotion recognition is introduced. This method uses transfer learning to extract features from multichannel EEG signals, these features are then arranged in an 8×9 map to represent their spatial location on scalp and then we introduce a CNN model which takes in the spatial feature map and extracts spatial relations between EEG channel and finally classify the emotions. First, EEG signals are converted to spectrogram and passed through a pre-trained image classification model to get a feature vector from spectrogram of EEG. Then, feature vectors of different channels are rearranged and are presented as input to a CNN model which extracts spatial features or dependencies of channels as part of training. Finally, CNN outputs are flattened and passed through dense layer to classify between emotion classes. In this study, SEED, SEED-IV and SEED-V EEG emotion data-sets are used for classification and our method achieves best classification accuracy of 97.09% on SEED, 89.81% on SEED-IV and 88.23% on SEED-V data-set with fivefold cross validation.

1 Introduction

Emotion Recognition has its applications in Brain-Computer interface, mental health diagnosis, fatigue and mental workload detection and disease evaluation. Human emotions can be recognized either by other human interaction or it can be automated and can be done with the help of tools such as facial expression detection, Electrocardiograph(ECG) recordings or EEG signals of a subject. Out of all these methods for human emotion detection EEG based emotion recognition is more popular in recent research. EEG based emotion recognition can be automated with high accuracy with the help of classification methods such as Machine Learning and Deep Learning techniques. As EEG data reflects true brain activity, it is more reliable data to work with. EEG data acquisition devices are relatively cheap, so it is research friendly. EEG has high temporal resolution. Because of these advantages of EEG, EEG based emotion recognition is very popular in researchers.

* Corresponding author: vsjadhav.98@gmail.com

There are specific areas of brain which gets more activated than others for specific tasks, which are also called Brodmann areas. A multichannel EEG can capture various positions on scalp, to get information about which brain areas are relatively more active and we can also extract dependencies between electrodes, which is also called as extracting spatial features. Also, there are known frequency bands associated with brain state, which can be used for spectral feature extraction. And finally, we can extract a temporal feature which tells what activity happened at what time and relation between their sequences. Most of the recent research on emotion recognition using EEG is based on efficiently extracting spectral, spatial and temporal relations to make classification easier. Some of the popular EEG emotion data-sets are compared in Table 1 and classification techniques on those data-sets found in literature is discussed below. Paper [1] has extensively reviewed open access EEG emotion datasets and classification techniques on them.

In study [2], authors have used a novel method, called 4D-CRNN for EEG based emotion recognition. Here, 4D refers to the input dimension to their classifier. This method extracts spectral (frequency bands), spatial and temporal dependencies in data to efficiently classify on SEED and DEEP datasets. Authors of [2] first apply band-pass filters to EEG signals to extract spectral information and then differential entropy of those band passed signals is extracted, differential entropy is found to be very effective for EEG based classification by [3] and many other studies. Then these differential entropy of different channels are arranged in a novel compact 8×9 map to form a 2D image, such 2D images are created for different frequency bands and stacked to form a 3D input for a CNN (convolutional neural network) to extract spatial and spectral dependency. Further to make use of temporal dependency outputs of such CNNs are stacked and passed to a RNN (recurrent neural network), making it a 4D input of data. This method achieves intra-subject accuracy of $94.74\% \pm 2.32$ on SEED dataset and $94.22\% \pm 2.61$ on valance and $94.58\% \pm 3.69$ on arousal classification on DEAP dataset.

In paper [4], Authors used 4D-aNN with same strategy of 4D data input as [2], but to effectively create spatial spectral and temporal representation of data, Authors adopt attention mechanism to assign weights to bands, channels and temporal slices to further increase accuracy of classifier. Here they use attention based CNN and attention based bidirectional LSTM on top of it. This method achieves accuracy of 96.25% on SEED, 86.77% on SEED-IV, 96.90% and 97.39% on valance and arousal classification of DEAP dataset.

Hierarchical CNN (HCNN) structure is used in [5] for EEG based emotion recognition on SEED dataset. This paper also uses 2D map for channel representation with sparse arrangement of channels in the map. In this paper differential entropy of short time Fourier transform of EEG signals is used as feature on four frequency bands. Best accuracy is achieved by this method is $86.2\% \pm 6.6$ on beta band of frequency.

A novel architecture, Emotion-Net is used in [6] to classify on SEED and SEEDIV datasets. In this method, 2D map of channels is crated with differential entropy features for five different frequency bands. Emotion-Net contains two separate streams with same architecture for spatial-spectral and spatial-temporal feature learning, each stream with several 3D attention blocks and transition layers. It uses pseudo 3D layer to decrease number of trainable parameters. Emotion-Net achieves accuracy of $96.02\% \pm 2017$ on SEED and $84.92\% \pm 6.66$ on SEED-IV datasets.

Table 1. Popular EEG emotion datasets found in literature

Ref	Dataset	Number of Classes	Class description	Number of channels	Number of Subjects	Sampling frequency (Hz)
[3, 11]	SEED	3	• negative • positive • neutral	62	15 (7male, 8Female)	1000 (down-sampled to 200)
[12]	SEED-IV	4	• happy • sad • neutral • fear	62	15 (7male, 8Female)	1000 (down-sampled to 200)
[7]	SEED-V	5	• happy • sad • disgust • neutral • fear	62	16 (6male, 10Female)	1000 (down-sampled to 200)
[13]	DEAP	3	• arousal • valance • dominance (on scale [1-9])	2	20 (10male, 10Female)	512
[14]	Mahnob-HCI	3	• arousal • valance • dominance (on scale [1-9])	32	27 (11male, 6Female)	256
[15]	DREAMER	9 with 3 scales	• arousal • valance • dominance (on scale [1-5])	14	23 (14male, 11Female)	128

In [7], Authors compare DCCA (deep canonical correlation analysis) and BDAE (bimodal deep autoencoder) on SEED, SEED-IV, SEED-V, DEAP and DREAMER datasets. They propose extension of DCCA with weighted sum fusion and attention based fusion within DCCA. Accuracy achieved by extended DCCA is 94.6% on SEED, 87.5% on SEED-IV, 85.3% on SEED-V datasets. On DEAP dataset they reported 84.3% and 85.6% for arousal and valance tasks respectively and 89.0%, 90.6%, 90.7% on arousal, valance and dominance tasks of DREAMER dataset.

A dynamical graph convolutional neural network (DGCNN) is used by [8] to classify on SEED and DREAMER datasets. Relations between different EEG channels are represented as an adjacency matrix graph, which is learned in neural network training. DGCNN achieves 90.4% and 79.95% accuracy in intra-subject and inter-subject classification task on SEED dataset respectively. It also achieves 86.23%, 84.54% and 85.02% accuracy on valance, arousal and dominance tasks on DREAMER dataset. This study reviews only limited paper in EEG emotion recognition which are related to proposed method and

dataset, but a state of the art and extensive review on EEG based emotion recognition can be found in study [9]. Also a review on various BCI datasets and classification methods can be found in [10]

2 Method

In this section our classification method is discussed in detail. This section has subsections as Pre-processing data (section 2.1), Feature extraction using transfer learning (section 2.2), CNN model (section 2.3).

2.1 Pre-processing data

Data is acquired from SJTU[†]. In this EEG based emotion recognition dataset, carefully chosen video clips are shown to participants to elicit particular emotion. In each experiment all the distinct videos are shown to participant and there EEG recordings are measured; each video watching event is called a trial. Each trial or video corresponds to exactly one emotional state. Section 3.1 explains details of SEED, SEED-IV and SEED-V datasets.

First, EEG signals are down sampled to 200Hz. Then band-pass filter is applied to get signal separated in three different bands with low and high cut-off as [1-14]Hz, [14-31]Hz, [31-51]Hz which captures delta-theta-alpha, beta and gamma spectrum respectively. Also, band passed signals are further normalized to values between 0 and 1 to compute effective spectrograms. Then these three spectral signals are converted to three spectrograms which is then passed to a pre-trained image classification model to extract transfer learning features. Pre-trained model used here is Inception V3[16]. Therefore, after feeding 2D spectrograms as images we finally get a 1-dimensional output vector from Inception Model, which will be fed to a CNN model. Figure 1 explains data pre-processing steps.

Spectrogram creation:

To crate spectrogram first input EEG signal is sliced into *num_samples* number of non-overlapping samples without padding. This *num_samples* variable depends on *spectrogram_width* and *hop_size* parameters, where *hop_size* is user defined parameter and *spectrogram_width* depends on input dimension requirements of transfer learning model. General idea about *hop_size* is that, as *hop_size* increases spectrogram will tend to be more discrete and with less temporal information, therefore, minimum value of *hop_size* is preferred for obtaining a continuous spectrogram and to get high temporal resolution, with downside of low values of *hop_size* being computationally expensive to calculate spectrogram.

A flow chart of all the data preprocessing is illustrated in figure1 Transfer learning model expects input image size to be close to the image sizes that it was trained on. To achieve spectrograms of desired size following formulae can be used to find out *frame_size* and *sample_size* parameters for user defined *hop_size*.

$$\begin{aligned} \text{frame_size} &= ((\text{spectrogram_height} - 1) \times 2) \\ \text{sample_size} &= ((\text{spectrogram_width} - 1) \times \text{hop_size}) \end{aligned}$$

[†] <https://bcmi.sjtu.edu.cn/home/seed/index.html>

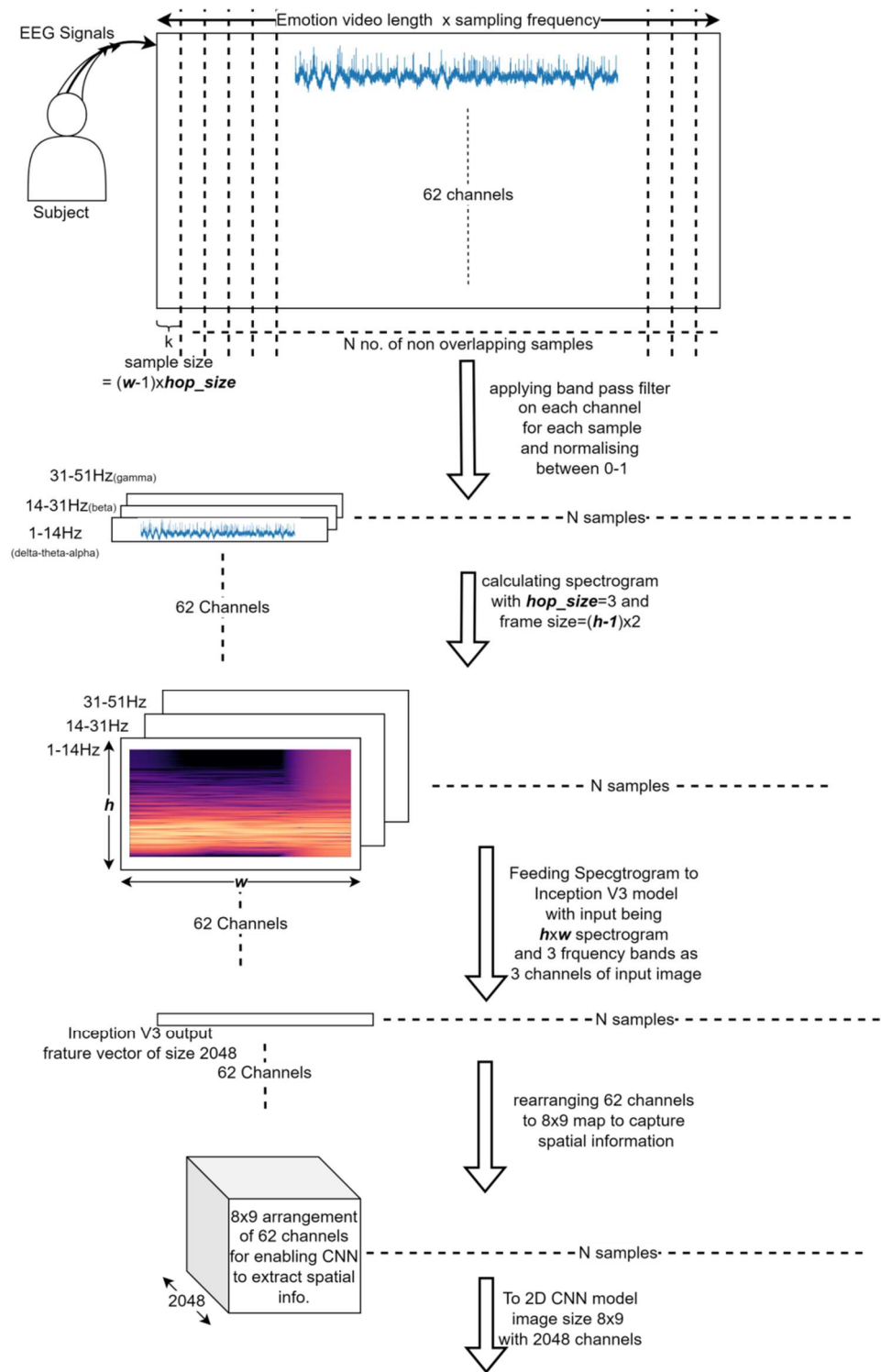


Fig. 1. Data pre-processing

2.2 feature extraction using Transfer Learning

Transfer learning is a deep learning technique by which we can use large pretrained models for classification and feature extraction without need to do compute hungry training. In transfer learning model weights and parameters are either used as it is or trained with a small dataset to tune them. In this study, popular transfer learning model Inception is used to extract features for EEG based emotion recognition. Particularly, we use Inception V3 model to obtain feature vector of EEG, the reason for using Inception model is discussed in discussion section. Inception model is trained on images with input size $229 \times 229 \times 3$, so to create spectrograms of size 229×229 we can use above formulae to find out *sample_size* and *frame_size*. Passing these *sample_size* and *frame_size* values to short time Fourier transform algorithm along with user defined *hop_size* we will get a 229×229 dimensional spectrogram for each band. Inception model without top layer outputs 2048 dimensional feature vector which can be utilised for further classification.

Therefore, for Inception model with *hop_size* of 3, we will have *frame_size* and *sample_size* as:

frame_size = ((229 - 1) × 2) = 456

sample_size = ((229 - 1) × 3) = 684

Creating 8x9 map of channels

For every event of EEG we have 62 transfer learnt feature vectors as SEED dataset is acquired with 62 channel device. Therefore, for each training example it is ideal to make all 62 channel's data available for classifier model to predict on. Positions of these 62 channels on scalp in SEED dataset experiment setup is known and therefore 62 feature vectors can be arranged in a compact map like an image. This compact map reflects channel positions as a top view of scalp. Studies like [2, 4, 5] uses this idea of arranging features as a compact map like an image. In this study 8x9 map of channels is created as used in [2] and as shown in figure2.

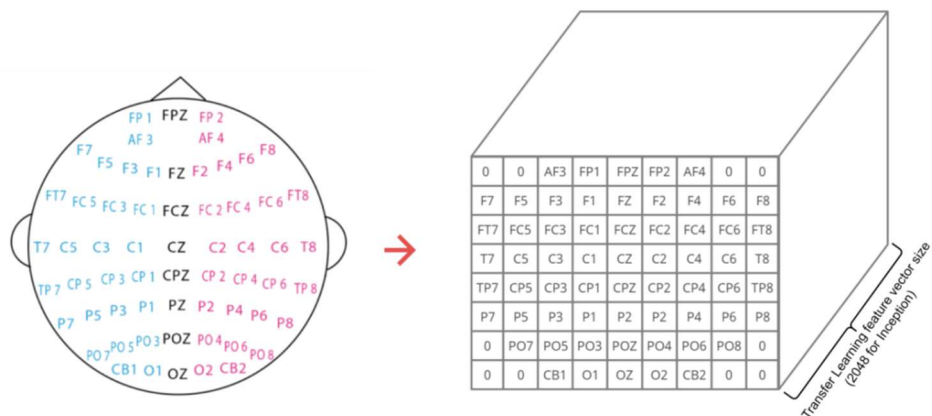


Fig. 2. 62 channels to 8×9 map conversion

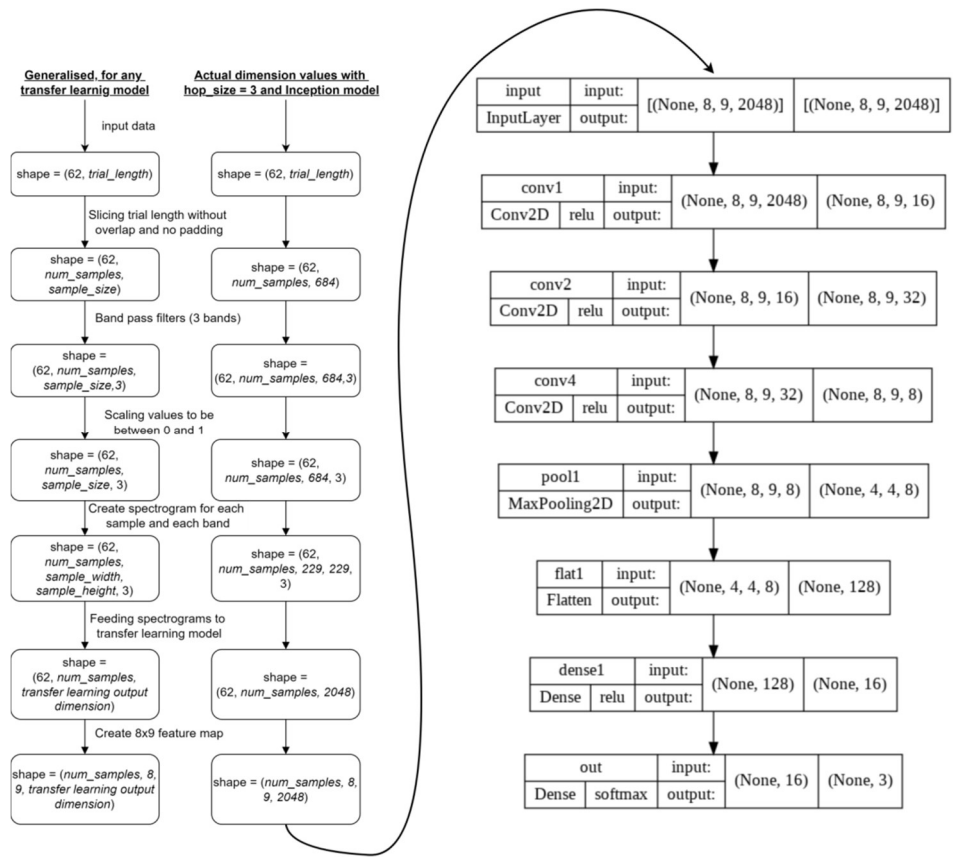


Fig. 3. Feature extraction and 2D CNN model architecture used in this paper

2.3 CNN Model for spatial dependency

As shown in figure2, we have a compact map of channels arranged in 8×9. This map can be viewed as an 8×9 dimensional image, where instead of RGB channels we have 2048 number of channels for Inception model. This image can be fed to 2D CNN model which will then extract spatial relations between neighbor channels. It is also possible to apply 3D CNN model if we don’t want look at all 2048 dimensional feature vector at once, and feed data with custom depth of CNN filter. After extracting spatial feature form CNN model, those features are passed through a dense layer and then finally a classifier layer with softmax activation. A 2D CNN model used has architecture as shown in figure3

3 Experiment

In this section, three open access EEG emotion datasets are introduced. Then, experimental settings, like hyper-parameters are explained and finally, results on these three dataset are presented with discussion on those results.

3.1 Datasets

Three publicly available EEG based emotion recognition datasets are used in this study. Datasets used are SEED[3, 11], SEED-IV[12] and SEED-V[7] provided by SJTU[‡]. Details of these datasets are as in table4. In each SEED dataset carefully chosen video clips are shown to participants to elicit particular emotion. In each experiment all the distinct videos are shown to participant and there EEG recordings are measured; each video watching event is called a trial. Each trial or video corresponds to exactly one emotional state. There are three such experiments done on each participant.

Table 2. Details of SEED, SEED-IV and SEED-V dataset

specification type	SEED	SEED-IV	SEED-V
Number of classes	3	4	5
Number of subjects	15	15	16
emotion stimulus	Video	Video	Video
Total distinct video clips	15	72	15
Number of videos per class per experiment	5	6	3
video clip length	4 min.	2 min.	2-4 min.
Number of experiments per subject	3	3	3
Eye movement tracking	No	Yes	Yes

3.2 Experimental Setup

After extracting transfer learning features, those features are passed to a CNN model described above with adam optimiser and learning rate of 0.001. Epochs are set to 150 with early stopping callback of patience=30 and val accuracy as monitor to trigger early stopping. This model is implemented in Keras[§] api of Google Tensorflow^{**} and trained on Google’s cloud with Tesla T4 and P100 GPUs. Codes will be available at https://github.com/vsjadhav/SEED_emotion_recognition_with_transfer_learning after publication of paper.

3.3 Results

Proposed method uses Inception as transfer learning model and has hyper-parameters as sample size and hop size. Model takes input from original EEG data with sample size non-overlapping segments. This sample size is mostly defined by required output dimension of spectrogram which is defined by input shape of image required by transfer learning model. This leaves only hop size parameter to be optimized. Effect of increasing or decreasing hop size is experimented and reported in following section.

[‡] <https://bcmi.sjtu.edu.cn/home/seed/index.html>
[§] <https://keras.io/>
^{**} <https://www.tensorflow.org/>

3.3.1 The effect of hop size

Effect of various hop sizes used for computing spectrogram on model performance is shown in table 3. It can be seen that with lower values of hop size model performs better, this is due to higher temporal information present in spectrogram corresponding to lower hop size values.

Table 3. Effect of hop size on accuracy on SEED and SEED-IV data. accuracy in five fold cross validation

<i>hop_size</i>	SEED	SEED-IV
24	91.04%	68.49%
12	95.78%	79.53%
6	97.01%	85.05%
3	97.09%	89.81%

3.4 Overall Performance

Model is trained with five-fold cross validation for intra subject emotion recognition task. Training and testing for each subject is done separately and results are generated for each subject. Average accuracy is considered as average of all subject’s accuracy, where subject’s accuracy is average of each subject’s k-fold five results. Figure 4, 5 and 6 shows average accuracy and average standard deviation for each subject on SEED and SEED-IV and SEED-V datasets respectively.

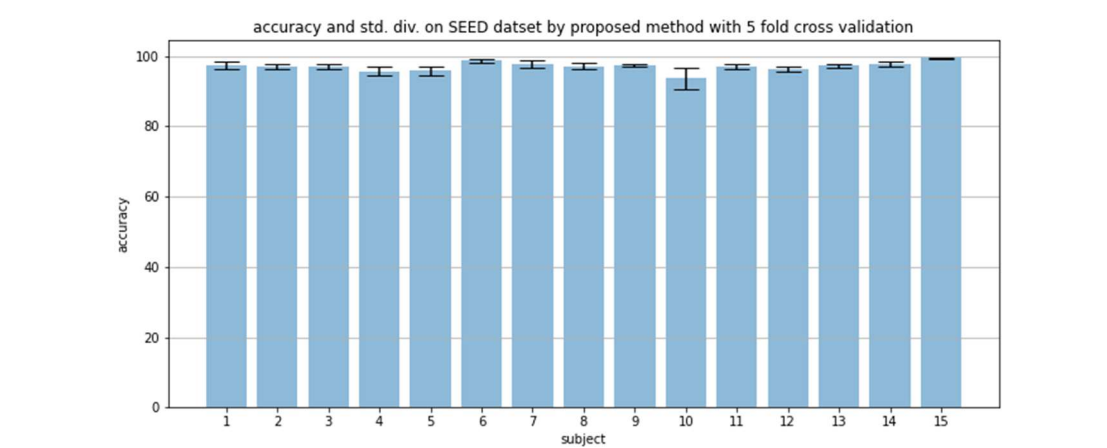


Fig. 4. Accuracy and std. div. with hop size=3 and SEED dataset on five fold cross validation

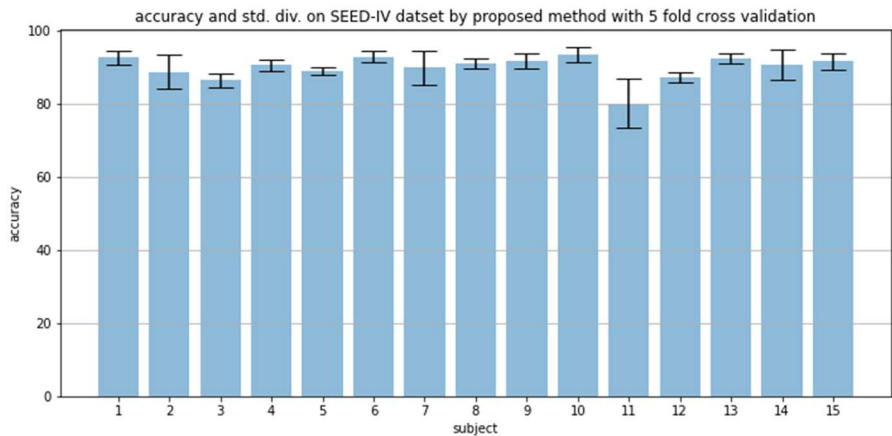


Fig. 5. Accuracy and std. div. with hop size=3 and SEED-IV dataset on five fold cross validation

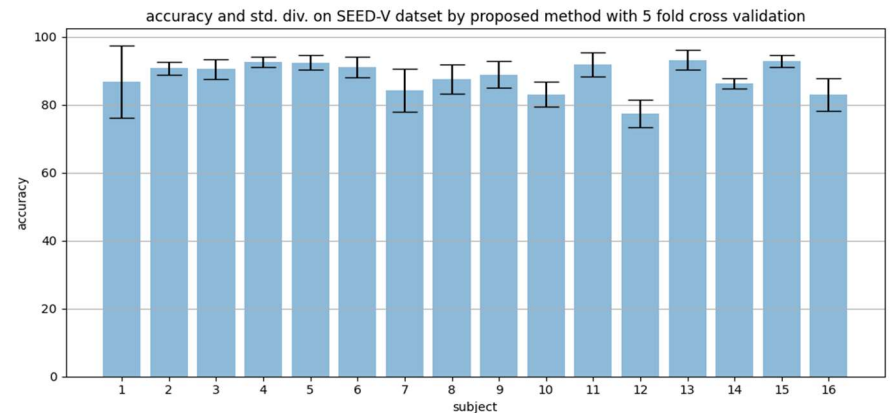


Fig. 6. Accuracy and std. div. with hop size=3 and SEED-V dataset on five fold cross validation

3.4.1 Comparison with other methods

This section compares performance of other studies with proposed method on SEED and SEED-IV dataset. Other methods compared here also uses five-fold cross validation as performance measure. In table 4 proposed method is compared with other methods found in literature on accuracy scale.

Table 4. Performance comparison

specification type	SEED	SEED-IV	SEED-V
HCNN (2018) [5]	86.2%	-	-
DGCNN (2018) [8]	90.4%	-	-

4D-CRNN (2020) [2]	94.74%	-	-
Emotion-Net (2020) [6]	96.02%	84.92%	-
DCCA (2021) [7]	94.6%	87.5%	85.3%
4D-aNN (2022) [4]	96.25%	86.77%	-
Proposed Method	97.09%	89.81%	88.23%

3.5 Discussion

Above comparison analysis shows that our Transfer Learning feature extraction and CNN architecture performs better than other methods. In this study Inception[16] model is chosen for feature extraction, but Resnet[17] is also a good choice. In Study [18] authors claimed with their experiment that both Inception and Resnet extracts almost similar features. They also reported that Inception extracts all the features extracted by Resnet along with extra features which were not present in Resnet’s features. Both Inception and Resnet are almost similar but Inception is slightly better and robust in feature extraction as found in [18]. For this reason Inception is chosen in this study for feature extraction.

Part of the reason that our proposed model achieves good accuracy is due to arranging the transfer learnt features of different EEG channels in 8×9 compact map as did in [2] which help CNN to learn spatial features.

Also in our method EEG signals converted to spectrogram which is log scaled short time Fourier transform (STFT). STFTs are known to have combination of both, good frequency resolution and also good temporal resolution. Therefore, with having spectrogram, Transfer Learning model gets input image with information about frequency components present in time interval of frame size of STFT, which helps efficiently learn frequency and temporal features.

Also by varying hop size temporal resolution of STFTs can be varied, which affects model performance as shown in table 3. With low value of hop size, spectrogram captures more temporal information, which increases model performance.

While conducting experiments, there was following normalizing (scaling values between 0 and 1) methods studied as shown in table 5, refer figure 3 for understanding normalizing experiment in table 5

Table 5. data Normalizing in different experiments

normalizing done at what step, refer fig.3	<i>exp1</i>	<i>exp2</i>	<i>exp3</i>	<i>exp4</i>
between band pass and spectrogram	No	No	Yes	Yes
between spectrogram and transfer learning	No	Yes	No	Yes

In reference to table 5, it was observed that, *exp1* and *exp2* resulted in very poor performance. Best performance was observed in normalizing as in *exp3*, which is also reported in result section3. It was expected that, *exp4* will give best performance as it does normalizing on spectrograms which are then fed to Inception model, because Inception expects images to be normalized, but it resulted in moderate performance. Best accuracy, as reported in result section3 was given by normalization as in *exp3*, although it does not normalize before transfer learning, but it was also observed that it is quite sensitive to initial conditions of training model i.e. initialized weights of CNN model after transfer leaning step. Model architecture is shown in figure 3

4 Conclusion

In this study, we present a novel method to classify EEG-based emotion. Our method extracts frequency, temporal and spatial feature from EEG data. Our proposed method achieves state-of-the-art performance on SEED and SEED-IV and SEED-V datasets. Transfer learning model extracts frequency and temporal features from EEG data and CNN with channel map learns spatial feature in multichannel EEG. Performance is greatly improved by making use of transfer learning to extract features. Method used in this study can be applied to other BCI applications as well with minor changes as per experimental conditions while acquiring the EEG data. Applications for the method used in this study can be Medical Diagnosis, BCI based UI design, Motor Imagery, Sleep Stage detection and any other EEG based classification application.

In future more studies with transfer learning model can be done, as it increases classification accuracy with minimum data for training. For this purpose it is required that a generalised deep neural network to be trained on various EEG datasets to extract features similar to generalised pre-trained image classification models.

References

1. E. P. Torres, E. A. Torres, M. Hernandez- ´ Alvarez, and S. G. Yoo, “Eeg-based ´ bci emotion recognition: A survey,” *Sensors*, vol. 20, no. 18, p. 5083,(2020).
2. F. Shen, G. Dai, G. Lin, J. Zhang, W. Kong, and H. Zeng, “Eeg-based emotion recognition using 4d convolutional recurrent neural network,” *Cognitive Neurodynamics*, vol. 14, no. 6, pp. 815–828, (2020).
3. R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, “Differential entropy feature for eegbased emotion classification,” in 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), IEEE, pp. 81–84, (2013).
4. G. Xiao, M. Shi, M. Ye, B. Xu, Z. Chen, and Q. Ren, “4d attention-based neural network for eeg emotion recognition,” *Cognitive Neurodynamics*, pp. 1– 14, (2022).
5. J. Li, Z. Zhang, and H. He, “Hierarchical convolutional neural networks for eeg-based emotion recognition,” *Cognitive Computation*, vol. 10, no. 2, pp. 368–380, (2018).
6. Z. Jia, Y. Lin, X. Cai, H. Chen, H. Gou, and J. Wang, “Sst-emotionnet: Spatial-spectral-temporal based attention 3d dense network for eeg emotion recognition,” in *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 2909–2917, (2020).
7. W. Liu, J.-L. Qiu, W.-L. Zheng, and B.-L. Lu, “Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition,” *IEEE Transactions on Cognitive and Developmental Systems*,(2021).
8. T. Song, W. Zheng, P. Song, and Z. Cui, “Eeg emotion recognition using dynamical graph convolutional neural networks,” *IEEE Transactions on Affective Computing*, vol. 11, no. 3, pp. 532–541, (2018).
9. N. S. Suhaimi, J. Mountstephens, and J. Teo, “Eeg-based emotion recognition: A state-of-the-art review of current trends and opportunities,” *Computational intelligence and neuroscience*, vol. 2020, (2020).
10. V. Jadhav, N. Tiwari, and M. Chawla, “A review on eeg data classification methods for brain–computer interface,” in *International Conference on Innovative Computing and Communications*, Springer, pp. 747–760, (2023).
11. W.-L. Zheng and B.-L. Lu, “Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks,” *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162– 175, (2015).

12. W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE transactions on cybernetics*, vol. 49, no. 3, pp. 1110–1122, (2018).
13. F. Fahimi, Z. Zhang, W. B. Goh, T.-S. Lee, K. K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for eeg-based bci," *Journal of neural engineering*, vol. 16, no. 2, p. 026 007, (2019).
14. M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 42–55, (2011).
15. S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 98–107, (2017).
16. C. Szegedy, W. Liu, Y. Jia, et al., "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, (2015).
17. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, (2016).
18. D. McNeely-White, J. R. Beveridge, and B. A. Draper, "Inception and resnet features are (almost) equivalent," *Cognitive Systems Research*, vol. 59, pp. 312–318, (2020).