Decoding Individual and Shared Experiences of Media Perception Using CNN Architectures

Johri, Riddhi; Pandey, Pankaj; Miyapuram, Krishna Prasad; Lomas, James Derek

DOI
10.1007/978-3-031-48593-0_14

Publication date
2024

Document Version
Final published version

Published in
Medical Image Understanding and Analysis - 27th Annual Conference, MIUA 2023, Proceedings

Citation (APA)
Johri, R., Pandey, P., Miyapuram, K. P., & Lomas, J. D. (2024). Decoding Individual and Shared Experiences of Media Perception Using CNN Architectures. In G. Waiter, G. Leontidis, T. Morris, T. Lambrou, N. Oren, & S. Gordon (Eds.), *Medical Image Understanding and Analysis - 27th Annual Conference, MIUA 2023, Proceedings* (pp. 182-196). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 14122 LNCS). Springer. https://doi.org/10.1007/978-3-031-48593-0_14

Decoding Individual and Shared Experiences of Media Perception Using CNN Architectures

Riddhi Johri[1], Pankaj Pandey[1], Krishna Prasad Miyapuram[1(✉)], and James Derek Lomas[2]

[1] IIT Gandhinagar, Gandhinagar, India
kprasad@iitgn.ac.in
[2] TU Delft, Delft, The Netherlands

Abstract. The brain is an incredibly complex organ capable of perceiving and interpreting a wide range of stimuli. Depending on individual brain chemistry and wiring, different people decipher the same stimuli differently, conditioned by their life experiences and environment. This study's objective is to decode how the CNN models capture and learn these differences and similarities in brain waves using three publicly available EEG datasets. While being exposed to a variety of media stimuli, each brain produces unique brain waves with some similarity to other neural signals to the same stimuli. However, to figure out whether our neural models are able to interpret and distinguish the common and unique signals correctly, we employed three widely used CNN architectures to interpret brain signals. We extracted the pre-processed versions of the EEG data and identified the dependency of time windows on feature learning for song and movie classification tasks, along with analyzing the performance of models on each dataset. While the minimum length snippet of 5 s was enough for the personalized model, the maximum length snippet of 30 s proved to be the most efficient in the case of the generalized model. The usage of a deeper architecture, i.e., DeepConvNet was found to be the best for extracting personalized and generalized features with the NMED-T and SEED datasets. However, EEGNet gave a better performance on the NMED-H dataset. Maximum accuracy of 69%, 100%, and 56% was achieved in the case of the personalized model on NMED-T, NMED-H, and SEED datasets, respectively. However, the maximum accuracies dropped to 18%, 37%, and 14% on NMED-T, NMED-H, and SEED datasets, respectively, in the generalized model. We achieved a 5% improvement over the state of the art while examining shared experiences on NMED-T. This marked the out-of-distribution generalization problem and signified the role of individual differences in media perception, thus emphasizing the development of personalized models along with generalized models with shared features at a certain level.

Keywords: EEG · Neural responses · Music and Movie perception · Subjective differences

G. Waiter et al. (Eds.): MIUA 2023, LNCS 14122, pp. 182–196, 2024.
https://doi.org/10.1007/978-3-031-48593-0_14

Table 1. Synonyms for Evaluation and Experience Terminology

Experience	Evaluation Key Term	Model
Individual	Within-Subject	Personalized
Shared	Cross-Subject	Generalized

1 Introduction

The words "digital transformation", "innovation", and "media experience" have been a lot common in recent years, and companies aim to translate these concepts into tangible results. Creating an environment that provides customers with the most incredible user experience by allowing them to receive information customized to their needs and preferences is essential. One such method is to regulate the media experience using the user's brainwaves by detecting the person's concentration and excitement levels using electroencephalography (EEG) in both educational and entertainment applications [8,13,14,18] (Table 1).

Neurotechnology is a branch of neuroscience that has already made significant contributions to our knowledge of the brain and nervous system. It entails the creation of novel sensors and wearable gadgets for measuring, stimulating, and modulating brain activity. An EEG measures and records electrical activity in the brain using non-invasive sensors. EEG headsets are being enhanced and developed for wearable consumer applications, such as cognitive state detection, mental and neurological disorder detection, consumer choice prediction [19], and understanding and predicting people's responses to different media experiences [1,2], emotion prediction [20] and practicing meditation [4,15–17]. There is a lot of potential for EEG data to be used to improve media experiences. For example, EEG could be used to track how engaging a particular piece of media is and make recommendations accordingly to maximize the impact of the content [22,25].

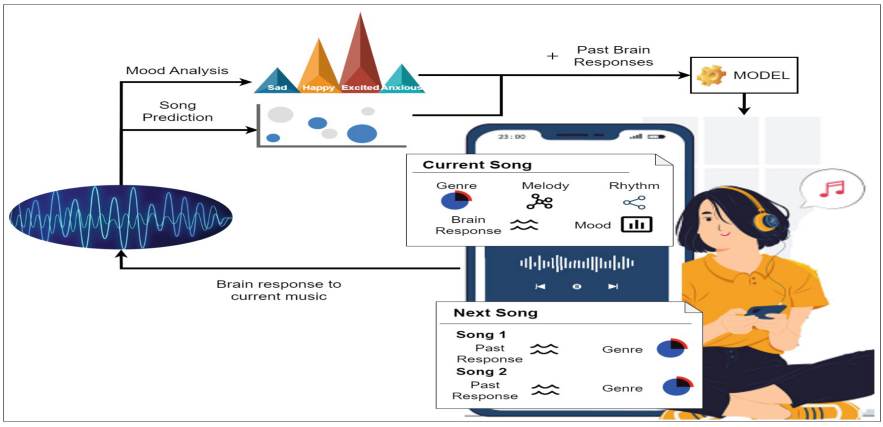


Fig. 1. Media Experience Brain Space: Each user's brain response and mood will be analyzed along with the prediction of the current song and its features. The model will then generate a playlist unique to the user, curated to evoke the emotion he/she is in

EEG is the most significant research area in the coming decade because EEG headsets can be pervasive as the fit-bit in our daily lives to monitor brain health. With the significant rise in the EEG domain, computation techniques should advance to capture intentional learning and reduce the unintentional factor of the learning. During the last decade, there has been a growing need for deep learning networks to be able to interpret learning in this field [3,21,23]. But to correctly elucidate these brain waves using neural architectures is the challenge. In this work, we predominantly discuss the disguise in feature learning that happens during the EEG classification task and the gap in the interpretations needed and interpretations done by these models.

The primary objective of this study is to evaluate Individual and Shared experiences in two different settings:

1. Song Classification
2. Movie Classification

and analyze the type and extent of feature learning done by deep learning networks. This is the first work which essentially presents the problem definition and proposes a direction for solving this problem.

2 Related Studies

Listening to music is a hobby, a tradition and a passion. The need for song recommendation systems arises as a result of system requirement that can recommend new songs to the users while being able to learn the user's past listening history, preferences, current listening habits, and mood. There are a few different ways that song recommender systems can operate. Some systems use collaborative filtering [6], which looks at the listening habits of a user's friends and followers to make recommendations. Other systems use content-based filtering [7], which looks at the attributes of a song (e.g. genre, artist, etc.) to make recommendations.

Yet another type is the EEG-based song identification which can be used to create playlist recommendations and improve song retrieval systems. The current state of the art [23] CNN-based model on NMED-T dataset for song identification is able to give an average test accuracy of 92.83% for within-subject but only 9%(less than chance level) for cross-subject classification. However, the cross-subject validation improved on retraining the model conditioned to classifying the EEG encodings into high and low-enjoyment classes. A research [18] has also shown that the first 20 s of a song segment can be used to train machine learning classifiers for accurate prediction from subsequent segments and that only β and γ band power spectra are enough to classify songs optimally. They achieved a maximum accuracy of 88% and 65% on Musin-G [12], and NMED-T datasets, respectively, using just one power band spectrum. However, as the testing shifted to cross-subject, the accuracy dropped to as low as 12% showing the out-of-generalization problem.

Listening to a song creates specific patterns in the brain, and these patterns are unique to each individual, which has been observed in the paper [24]. This is one reason why generalization is hard to get with EEG learning. While using a CNN architecture, the authors were able to predict the song using only 1 sec of EEG data and 10% of the data for training with an accuracy of 84.96% for within-participant. The results dropped to a 9.44% for cross-participant.

3 Data Description

3.1 NMED-T

This study analyzed NMED-T [11], a publicly available dataset containing behavioural responses and EEG from twenty participants engaged in a naturalistic song-listening. The EEG recordings were made with the electrode net attached, and behavioural ratings were obtained afterwards. Songs were presented in random order during the acquisition of the dataset. Each trial was followed by participants rating their familiarity and enjoyment of the music on a scale of 1–9. The EEG experiment was divided into two consecutive recording blocks to minimize participant tiredness and facilitate electrode impedance testing between the recording blocks. The preprocessed version of this dataset, which contains 125 channels of EEG data captured at 125 Hz, was primarily used in our study.

3.2 NMED-H

This publicly available naturalistic music EEG dataset [9] contains recorded brain signals of 48 adults listening to full-length Hindi pop songs. A total of sixteen stimuli, four songs, and four stimulus conditions per song are included in the dataset. A different version of the song was played twice to each group of twelve participants assigned to each stimulus. Each piece has four versions: Original, Reversed, Phase-scrambled, and Measure-shuffled, each lasting around 4.5 min. The results of this study are based on clean EEG Matlab files that had been cleaned, preprocessed, anonymized, and aggregated across participants into various stimulus matrices. Additionally, the dataset contains data structures listing the participants, stimulus and their behavioural ratings. Participants were only provided with one version of a song to listen to in the experiment.

3.3 SEED

SEED [28] contains EEG signals collected from fifteen subjects, seven males and eight females, who watched fifteen excerpts from Chinese films. The film clips lasted approximately four minutes and were well-edited to create coherent emotions that evoked and maximized three emotions: positive, negative, and neutral. Fifteen trials were conducted per subject, lasting 305 s, including a hint of starting for 5 s, a movie clip for 4 min, a self-assessment for 45 s, and a rest for

15 s. The data collected from the 62-electrode EEG cap was then downsampled to 200 Hz and processed using a bandpass filter ranging from 0 to 75 Hz [5]. The extracted differential entropy (DE) features of the EEG signals were smoothed further with conventional moving average and linear dynamic systems (LDS) methods.

4 Methodology

(See Fig. 2).

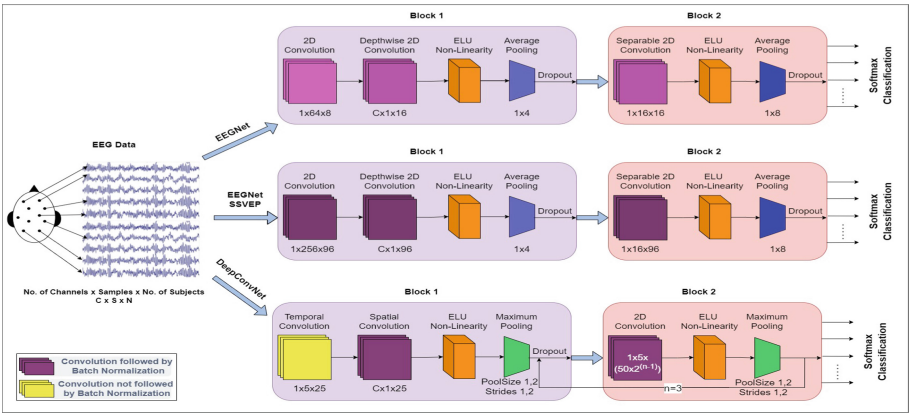


Fig. 2. The different architectures used for classification from EEG data

4.1 EEGNet

EEG data consisting of C channels, S time samples, and N subjects are passed for 150 epochs to the EEGNet model [10] composed of three convolutions in sequence. The input is routed through eight 2D convolution filters of size (1,64) to generate feature maps at various bandpass frequencies in the first block to obtain temporal information. Then, D*8 depthwise convolutions of size (C,1) are employed to learn spatial information within each temporal filter. The depth parameter D determines how many spatial filters should be learned for each feature map. With a dropout rate of 0.5, the model is regularized after applying exponential linear unit (ELU) Non-Linearity and Average Pooling layer of size (1, 4).

This is followed by a Separable 2D Convolution layer consisting of sixteen filters of size (1,16) in Block 2. This helps to combine spatial filters across temporal bands optimally. After applying ELU Non-Linearity, dimensionality reduction is achieved using an Average Pooling layer of size (1, 8). All the convolutions are followed by Batch Normalization. Finally, features after dropout are passed to the Softmax Classification layer.

4.2 EEGNet_SSVEP

The SSVEP variant of EEGNet [27] was explicitly designed for classifying Asynchronous Steady-State Visual Evoked Potentials signals. This differs from the above network in terms of size and number of kernels used in each convolution layer, as shown in Fig. 1. In block 1, ninety-six 2D Convolution and Depthwise 2D Convolution (D = 1) of size (1,256) and (C,1), respectively, are used to obtain frequency-specific spatial filters. Furthermore, depthwise convolutions reduce the number of free parameters to fit when compared to fully-connected convolutions.

In block 2, ninety-six separable convolutions of size (1,16) are used, which benefits by reducing the number of parameters to fit as well as explicitly decoupling the relationship between feature maps within and across them. In turn, a kernel summarizing each feature map is learned, followed by the optimal merging of the outputs. Each Convolution layer is followed by Batch Normalization. After convolutions in both blocks, the input passes through ELU non-linear activation, 2D average pooling, and dropout layers. Lastly, a dense layer and a softmax activation function are connected to the final layer.

4.3 DeepConvNet

The deep ConvNet architecture [26] to extract features and decode EEG signals is inspired by computer vision architectures. This architecture has four blocks, each consisting of a 2D convolution layer with max_norm constraint, batch normalization, ELU non-linearity activation, max pooling of size (1,2) with strides (1,2), and a dropout layer with a dropout rate of 0.5.

The convolution of the first block is split into two convolution layers of 25 filters each, one temporal layer (1,5) and one spatial layer (C,1). By using two layers, a linear transformation is forced into a combination of a temporal and a spatial filter, which implicitly regularizes the overall convolution. Finally, the fifth layer is a dense layer with a softmax activation function for classification.

4.4 Evaluation Strategy

To carry out this study, the datasets were divided into 5 s, 10 s, 20 s, and 30 s windows and an analysis of the performance of architectures was made to find the most efficient time window.

To decode how the models capture and comprehend the individuality and commonality in the perception of media by every individual, the experiments were performed in two settings. In the within-subject analysis, the data was split into the train, validate, and test datasets, thus having leakage of subjects' information. However, in the cross-subject analysis, the data was split such that the data of subjects present in the test dataset were not shown to the model while training. This resulted in a visible difference in the architecture's performance in both settings. They were inadequate to extract generalized features in the case of distribution shift. To verify these results, further experiments were done to plot t-SNE graphs showing song and subject classification. The t-SNE plots

displayed the groups formed in training and testing data according to songs in the case of the personalized model and subjects in the case of the generalized model.

5 Experimental Results

Table 2. Classification Accuracy

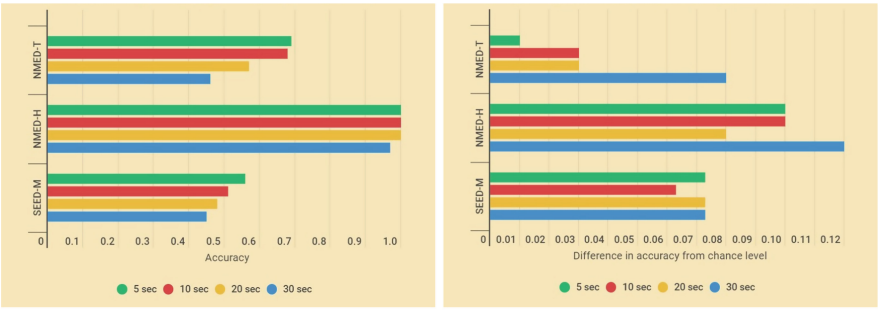
Within Subject			
Chunks	NMED-T	NMED-H	SEED-M
5	**0.69 ± 0.01**	1.0 ± 0.001	**0.56 ± 0.01**
10	0.68 ± 0.02	**1.0 ± 0.0**	0.51 ± 0.01
20	0.57 ± 0.04	**1.0 ± 0.0**	0.48 ± 0.02
30	0.46 ± 0.02	0.97 ± 0.04	0.45 ± 0.02
Cross Subject			
Chunks	NMED-T	NMED-H	SEED-M
5	0.11 ± 0.04	0.35 ± 0.15	**0.14 ± 0.04**
10	0.13 ± 0.04	0.35 ± 0.16	0.13 ± 0.04
20	0.13 ± 0.03	0.33 ± 0.16	**0.14 ± 0.04**
30	**0.18 ± 0.07**	**0.37 ± 0.18**	0.14 ± 0.05

5.1 Time Window for Personalized Model

With different-sized time windows of the same dataset, Fig. 3(a) compares the best performance of all the architectures. The results of NMED-T, NMED-H, and SEED datasets, also evident in Table 2, indicate that 5 s window size is adequate and give the best results with 69%, 100%, and 56% accuracy, respectively. This shows that the architectures are capable of identifying and classifying even from the smallest snippet of brain signals. Thus, the features can be learnt and identified accurately and efficiently from tiniest fragments of brain activity by the networks in the case of within-subject evaluation.

5.2 Time Window for Generalized Model

Figure 3(b) compares the best performance for cross-subject evaluation from all the architectures when fed with different-sized time windows of the same dataset. Here the results are contrary to that of the personalized model as the 30 s window size gives the best results. Moreover, the architectures achieved a maximum accuracy of 18%, 37%, and 14%, respectively, on NMED-T, NMED-H, and SEED datasets as shown in Table 2. This shows the disguise in feature learning done by the models resulting in such low accuracies even when fed by large chunks of data at a time. However, if the dataset size is increased, the model might be able to learn features from a higher number of 30-second windows, leading to better results.



(a) (b)

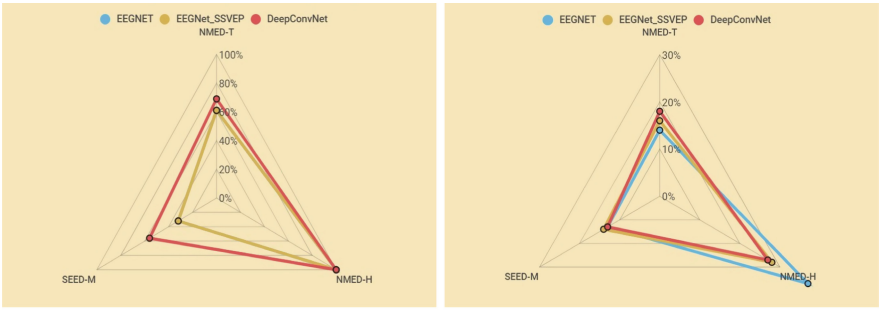
Fig. 3. (a) Within-Subject and (b) Cross-Subject Classification accuracies using different architectures on different-sized time windows

Table 3. Best accuracies achieved for each dataset

Within Subject			
Dataset	EEGNET	EEGNet_SSVEP	DeepConvNet
NMED-T	0.61 ± 0.05	0.61 ± 0.04	**0.69 ± 0.01**
NMED-H	**1.0 ± 0.001**	**1.0 ± 0.001**	1.0 ± 0.002
SEED-M	0.32 ± 0.02	0.32 ± 0.01	**0.56 ± 0.01**
Cross Subject			
Dataset	EEGNET	EEGNet_SSVEP	DeepConvNet
NMED-T	0.14 ± 0.04	0.16 ± 0.05	**0.18 ± 0.07**
NMED-H	**0.37 ± 0.18**	0.28 ± 0.09	0.27 ± 0.12
SEED-M	0.13 ± 0.05	**0.14 ± 0.05**	0.13 ± 0.05

5.3 Within-Subject Evaluation

We can see from the Table 3 and Fig. 4(a) that DeepConvNet outperformed the other two architectures on every dataset, whereas EEGNET and EEGNet_SSVEP performed similarly on the three datasets. Although the difference in accuracy on NMED-T is only 8%, the accuracy on SEED movie classification has increased significantly from 32% to 56%. In view of this, a deeper neural network can be said to do a better training on brain signal data for classifying individual experiences since it has more layers in its architecture, allowing it to learn the features more accurately.



(a) (b)

Fig. 4. Best (a) Within-Subject and (b) Cross-Subject Classification accuracies using different architectures on the datasets

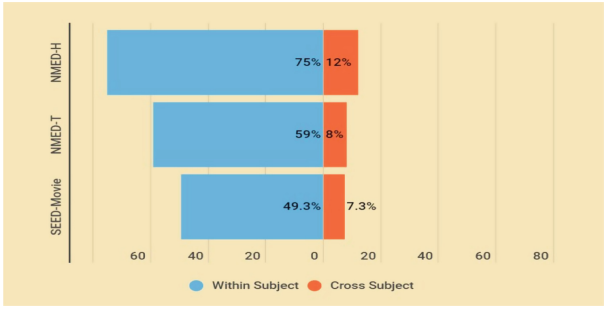


Fig. 5. The maximum increase in accuracy achieved from chance level on the datasets

5.4 Cross-Subject Evaluation

In contrast to the results of the personalized model, for the generalized model, no particular architecture performed well on all the datasets, as evident from Fig. 4(b). The architectures have showcased their inability to learn the appropriate feature to classify the data of unseen subjects. Table 3 shows that, on NMED-T data, DeepConvNet outperformed the other neural networks with a performance of 18%, EEGNet claimed a performance of 37% on NMED-H data, and EEGNet_SSVEP performed well on SEED data with a performance of 14%. However, recognizing the general characteristic of different brain signals for the same stimuli was not decoded by any architecture.

Table 4. Maximum classification accuracies achieved in the datasets considering all the architectures and various window sizes

Dataset	Chance Level	Within Subject	Increase	Cross Subject	Increase
NMED-T	0.1	**0.69**	0.59	**0.18**	0.08
NMED-H	0.25	**1**	0.75	**0.37**	0.12
SEED-Movie	0.067	**0.56**	0.493	**0.14**	0.073

5.5 Difference in Accuracies from Chance Level

A considerable difference between the increase in accuracies from chance level for within-subject and cross-subject classification can be seen in Fig. 5 and Table 4. In the personalized model, there was at least a 49% increase over chance levels, whereas, in the generalized model, it was barely 12%. This is due to the fact that the models did not learn relevant features of the song/movie to be able to classify when getting tested on distinct subjects indicating the distribution shift problem. However, the models performed exceedingly well when there was data leakage of the subjects. This suggests that models are influenced by the subjects' features and are learning properties specific to the song/movie as well as the subjects.

5.6 Network

Each of the experiments demonstrated that the DeepConvNet had shown consistent performance, with the model either providing the best accuracy or a similar level of precision to others. This is mainly due to its ability to capture the complex features of the data by increasing the depth of the network. The network uses a large number of filters and ReLU activations to increase the depth of the network and improve its performance. It also utilizes batch normalization and dropout layers to prevent overfitting, and the use of pooling layers has enabled the network to reduce the number of parameters used and thus reduce the computational complexity.

6 Discussion

6.1 Same Brain Perceives Different Stimuli Differently

The brain has a lot to say about who we are as individuals. Different parts of the brain may be activated depending on the type of stimuli and the person's individual response to it. Thus this difference is reflected in the EEG signals that help the models to identify the class of a new signal. Depending on the unique features found in EEG signals of different stimuli, the model gets trained to identify those features during testing to classify the song or movie. The t-SNE plots in Fig. 6(a) and (b) show how accurately the models were able to group the different songs of the NMED-T dataset.

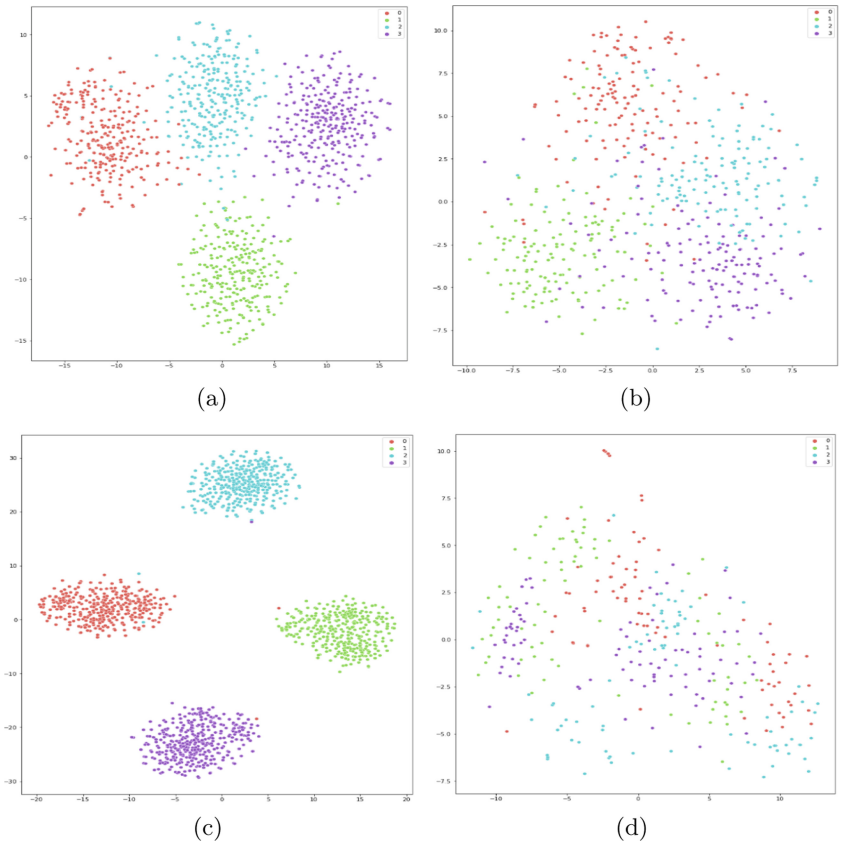


Fig. 6. t-SNE plots for Song classification on (a) train data, (b) test data after train-test split, and (c) train data, (d) test data after cross subject

6.2 Different Brains Perceive the Same Stimuli Differently

Variation in every individual's brain perception results in the models learning irrelevant features that might be specific to the subject rather than the media. This fact is visible while doing the cross-subject evaluation. The t-SNE plots in Fig. 6(c) and (d) show clear groups being formed for every song with the training dataset; however, on the test dataset, no such categorization is visible. On the other hand, for subject classification in Fig. 7(c) and (d), there is a clear group formation of subjects on the test dataset but not on the training dataset. Thus, the model has learnt features specific to the subjects and is categorizing based on that resulting in a low classification accuracy of only 18% for the generalized model. This is called the distribution shift or the out-of-distribution (OOD) generalization problem where the models are not able to accurately make predictions on data from the new, unknown distribution, which are new subjects here.

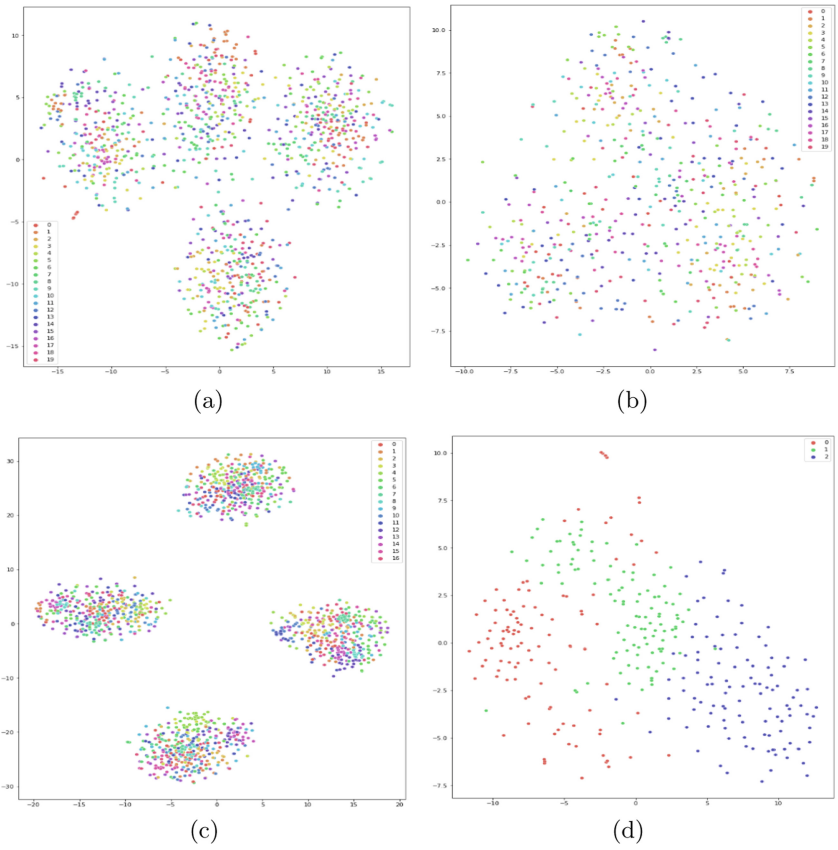


Fig. 7. t-SNE plots for Subject classification on (a) train data, (b) test data after train-test split, and (c) train data, (d) test data after cross subject

Since the same brain perceives the same stimuli invariably, the t-SNE plots in Fig. 6(a) and (b) of song classification in case of within-subject evaluation showcase precise group formation of tunes on both train and test datasets. This categorization is not achieved for subject classification, apparent in Fig. 7(a) and (b). Thus, when there was no distribution shift and the model was aware of all the subjects during the training and validation phase, it learned useful features of the media stimuli and achieved an accuracy of 69%.

6.3 Outperform State of the Art on NMED-T Music Identification

Recent works echoed in Table 5 have examined the song identification task using EEG signals to exhibit the subjective differences of neural responses in music perception. One of the studies [23] used a CNN architecture to identify songs of the NMED-T dataset with leave one subject out cross-validation and achieved a maximum precision of 9.9%. Another study [18] experimented with the relation

Table 5. Comparision with other works on Cross-Subject Song Classification

Article	Initial Input	Classifier	Accuracy
Dhananjay Sonawane et al. [24]	Time Frequency Plots	CNN architecture	9.44%
Gulshan Sharma et al. [23]	Topoplots	CNN architecture	9.9%
Pankaj Pandey et al. [18]	Frequency Bandpower	Random Forest	12.9%
Ours	Time Series	DeepConvNet	**18%**

between brain signals and unique and repetitive patterns present in the songs to classify the stimulus achieving a comparable maximum accuracy of 12.9% and 12.5% on NMED-T and MUSIN-G datasets respectively. Similarly, the authors of [24] showed the use of CNN architectures on the frequency domain dataset of MUSIN-G to gain a cross-subject song classification accuracy of 9.44%. Our work surpass these state-of-the-art works by achieving a maximum accuracy of 18% on the generalized model.

6.4 Why NMED-H Reflects the Highest Accuracy?

The models used in this research learned features not only unique to the media stimuli but to the subjects too. In the NMED-H dataset, a subject did not listen to a different version of the same song or an other song of the same version. Hence, no two songs considered for classification have the same subject. This results in corresponding groups of songs and subjects. This becomes more evident from the contrasting accuracies for within-subject and cross-subject evaluation of 100% and 37%, respectively.

7 Conclusion

In the future, EEG will empower the creation of large datasets that link people's brain activity to their responses to different media experiences. This could potentially be used to create personalized media experiences that are tailored to each individual's preferences and brain activity.

Our results show that different brains have different responses to the same experiences, and the same brain has different reactions to different experiences. The demand for new models that can adapt to this distribution shift between subjects is also evident from the t-SNE plots of song and subject classification for within-subject and cross-subject categories. It can be seen that while the test data t-SNE plot has clear and distinct groups of different songs in the case of the personalized model, the generalized model has the same for various subjects rather than the songs. The development of generalized models will elevate the media experiences of every individual using EEG wearables technology. The users will be able to better understand their cognitive and emotional processes, and thus make better decisions regarding their media consumption.

References

1. Building the world's most valuable brain data models. www.kernel.com/
2. Transforming music into medicine. www.lucidtherapeutics.com/
3. Bedmutha, P., Pandey, P., Ahmed, N., Miyapuram, K.P., Lomas, D.: Canonical correlation analysis (CCA) reveal neural entrainment for each song and similarity among genres (2022)
4. Chaudhary, S., Pandey, P., Miyapuram, K.P., Lomas, D.: Classifying EEG signals of mind-wandering across different styles of meditation. In: Brain Informatics: 15th International Conference, BI 2022, Padua, Italy, 15–17 July 2022, Proceedings, pp. 152–163. Springer, Heidelberg (2022). https://doi.org/10.1007/978-3-031-15037-1_13
5. Duan, R.N., Zhu, J.Y., Lu, B.L.: Differential entropy feature for EEG-based emotion classification. In: 6th International IEEE/EMBS Conference on Neural Engineering (NER), pp. 81–84. IEEE (2013)
6. Elahi, M., Ricci, F., Rubens, N.: A survey of active learning in collaborative filtering recommender systems. Comput. Sci. Rev. **20**, 29–50 (2016). https://doi.org/10.1016/j.cosrev.2016.05.002
7. Geetha, G., Safa, M., Fancy, C., Saranya, D.: A hybrid approach using collaborative filtering and content based filtering for recommender system. J. Phys. Conf. Ser. **1000**(1), 012101 (2018). https://doi.org/10.1088/1742-6596/1000/1/012101
8. Johri, R., Pandey, P., Miyapuram, K.P., Lomas, D.: Brain activity recognition using deep electroencephalography representation. In: 2023 IEEE Applied Sensing Conference (APSCON), pp. 1–3. IEEE (2023)
9. Kaneshiro, B., Nguyen, D.T., Dmochowski, J.P., Norcia, A.M., Berger, J.: Naturalistic music EEG dataset - hindi (nmed-h) (2014–2016). www.exhibits.stanford.edu/data/catalog/sd922db3535
10. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: EEGNET: a compact convolutional neural network for EEG-based brain-computer interfaces. J. Neural Eng. **15**(5), 056013 (2018). www.stacks.iop.org/1741-2552/15/i=5/a=056013
11. Losorelli, S., Nguyen, D.T.T., Dmochowski, J.P., Kaneshiro, B.: Naturalistic music EEG dataset - tempo (nmed-t) (2017). www.exhibits.stanford.edu/data/catalog/jn859kj8079
12. Miyapuram, K.P., Ahmad, N., Pandey, P., Lomas, J.D.: Electroencephalography (EEG) dataset during naturalistic music listening comprising different genres with familiarity and enjoyment ratings. Data Brief **45**, 108663 (2022). https://doi.org/10.1016/j.dib.2022.108663. www.sciencedirect.com/science/article/pii/S235234092200868X
13. Pandey, P., Ahmad, N., Miyapuram, K.P., Lomas, D.: Predicting dominant beat frequency from brain responses while listening to music. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 3058–3064 (2021). https://doi.org/10.1109/BIBM52615.2021.9669750
14. Pandey, P., Bedmutha, P.S., Miyapuram, K.P., Lomas, D.: Stronger correlation of music features with brain signals predicts increased levels of enjoyment. In: 2023 IEEE Applied Sensing Conference (APSCON), pp. 1–3. IEEE (2023)
15. Pandey, P., Gupta, P., Miyapuram, K.P.: Brain connectivity based classification of meditation expertise. In: Mahmud, M., Kaiser, M.S., Vassanelli, S., Dai, Q., Zhong, N. (eds.) BI 2021. LNCS (LNAI), vol. 12960, pp. 89–98. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86993-9_9

16. Pandey, P., Miyapuram, K.P.: Nonlinear EEG analysis of mindfulness training using interpretable machine learning. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 3051–3057. IEEE (2021)
17. Pandey, P., Rodriguez-Larios, J., Miyapuram, K.P., Lomas, D.: Detecting moments of distraction during meditation practice based on changes in the EEG signal. In: 2023 IEEE Applied Sensing Conference (APSCON), pp. 1–3. IEEE (2023)
18. Pandey, P., Sharma, G., Miyapuram, K.P., Subramanian, R., Lomas, D.: Music identification using brain responses to initial snippets. In: ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1246–1250 (2022). https://doi.org/10.1109/ICASSP43922.2022.9747332
19. Pandey, P., Swarnkar, R., Kakaria, S., Miyapuram, K.P.: Understanding consumer preferences for movie trailers from eeg using machine learning. arXiv preprint arXiv:2007.10756 (2020)
20. Pandey, P., Tripathi, R., Miyapuram, K.P.: Classifying oscillatory brain activity associated with Indian rasas using network metrics. Brain Inf. **9**(1), 1–20 (2022)
21. Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T.H., Faubert, J.: Deep learning-based electroencephalography analysis: a systematic review - iopscience (2019). www.iopscience.iop.org/article/10.1088/1741-2552/ab260c
22. Salehzadeh, A., Calitz, A.P., Greyling, J.: Human activity recognition using deep electroencephalography learning. Biomed. Signal Process. Control **62**, 102094 (2020). https://doi.org/10.1016/j.bspc.2020.102094. www.sciencedirect.com/science/article/pii/S1746809420302500
23. Sharma, G., Pandey, P., Subramanian, R., Miyapuram, K.P., Dhall, A.: Neural encoding of songs is modulated by their enjoyment (2022). https://doi.org/10.48550/ARXIV.2208.06679
24. Sonawane, D., Miyapuram, K.P., Rs, B., Lomas, D.J.: Guessthemusic: song identification from electroencephalography response (2020). https://doi.org/10.48550/ARXIV.2009.08793
25. Su, J., Wen, Z., Lin, T., Guan, Y.: Learning disentangled behaviour patterns for wearable-based human activity recognition. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 6, no. 1, pp. 1–19 (2022). https://doi.org/10.1145/3517252
26. Tibor, S.R., et al.: Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapp. **38**(11), 5391–5420 (2017). https://doi.org/10.1002/hbm.23730. www.onlinelibrary.wiley.com/doi/abs/10.1002/hbm.23730
27. Waytowich, N., et al.: Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. J. Neural Eng. **15**(6), 066031 (2018). www.stacks.iop.org/1741-2552/15/i=6/a=066031
28. Zheng, W.L., Lu, B.L.: Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans. Auton. Ment. Dev. **7**(3), 162–175 (2015). https://doi.org/10.1109/TAMD.2015.2431497