Soft Computing
https://doi.org/10.1007/s00500-021-06578-4

FOCUS

# EmoPercept: EEG-based emotion classification through perceiver

**Aadam[1] · Abdallah Tubaishat[2] · Feras Al-Obeidat[2] · Zahid Halim[1] · Muhammad Waqas[1] · Fawad Qayum[3]**

Accepted: 11 November 2021


**Abstract**
Emotions play an important role in human cognition and are commonly associated with perception, logical decision making, human interaction, and intelligence. Emotion and stress detection is an emerging topic of interest and importance in the research community. With the availability of portable, cheap, and reliable sensor devices, researchers are opting to use physiological signals for emotion classification as they are more prone to human deception, as compared to audiovisual signals. In recent years, deep neural networks have gained popularity and have inspired new ideas for emotion recognition based on electroencephalogram (EEG) signals. Recently, widespread use of transformer-based architectures has been observed, providing state-of-the-art results in several domains, from natural language processing to computer vision, and object detection. In this work, we investigate the effectiveness and accuracy of a novel transformer-based architecture, called perceiver, which claims to be able to handle inputs from any modality, be it an image, audio, or video. We utilize the perceiver architecture on raw EEG signals taken from one of the most widely used publicly available EEG-based emotion recognition datasets, i.e., DEAP, and compare its results with some of the best performing models in the domain.

**Keywords** Deep learning · EEG data · Emotion identification · Perceiver

**Abbreviations**

| | |
|---|---|
| *EEG* | Electroencephalogram |
| *CNN* | Convolutional neural network |
| *NLP* | Natural language processing |
| *RNN* | Recurrent neural network |
| *DBN* | Deep Belief Network |
| *GCNN* | Graph Convolutional Neural Network |
| *CapsNet* | Capsule Network |
| *LSTM* | Long Short-Term Memory |
| *DT* | Decision Trees |

Communicated by Shah Nazir.

✉ Aadam
aadimator@gmail.com

✉ Zahid Halim
zahid.halim@giki.edu.pk

Abdallah Tubaishat
Abdallah.Tubaishat@zu.ac.ae

Feras Al-Obeidat
feras.al-obeidat@zu.ac.ae

Muhammad Waqas
muhammad.waqas@giki.edu.pk

Fawad Qayum
fawadqayum@uom.edu.pk

1 Faculty of Computer Science and Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Swabi, Khyber Pakhtunkhwa, Pakistan

2 College of Technological Innovation, Zayed University, Abu Dhabi, UAE

3 Department of Computer Science and IT, University of Malakand, Chakdara, Pakistan

## 1 Introduction

Emotion plays an important role in our everyday decision-making and social interaction as it influences the perception of human surroundings (Wu et al. 2020). In general, there are two different ways to recognize emotion. One is through behavioral signals, such as speech (Petrushin 2000), facial expressions (Anderson and McOwan 2006), gestures (Soleymani et al. 2012), and body posture, to name a few, to construct models. This approach collects data in a noninvasive way but it is challenging to obtain the correct emotion if the person conceals her true emotion. The other method is to use physiological signals such as skin conductivity, heart rate, respiration, and EEG, to construct models and classify emotions. As compared to behavioral signals, physiological signals are more spontaneous and difficult to conceal (Yang et al. 2018).

Out of all the physiological signals, the most widely used physiological signal in brain imaging technologies is the electroencephalography (EEG) (Liu et al. 2021; Tao et al. 2020; Yin et al. 2021; Halim and Rehan 2020), which measures human brain activity directly. EEG signals are collected by placing several electrodes on the surface of the human head. Recently, many researchers have used EEG signals for human emotion recognition, achieving very convincing results and proving the effectiveness of EEG signals for the purposes of emotion recognition, among other things.

Because of the powerful ability of automatic feature extraction, deep learning algorithms (Muhammad and Halim 2016; Uzma and Halim 2021; Halim et al. 2017) have achieved noteworthy performance in the field of computer vision (Lecun et al. 1998; He et al. 2016), natural language processing (Liu et al. 2019; Collobert and Weston 2008), speech recognition (Yao et al. 2020), object detection (Liu et al. 2021), as well as EEG-based emotion recognition (Zhang et al. 2021; Yin et al. 2021; Xiao et al. 2021; Ding et al. 2021; Deng et al. 2021). Many models have been applied in the past for EEG-based emotion recognition, including but not limited to, convolutional neural networks (CNNs) (Yang et al. 2018; Tripathi et al. 2017), recurrent neural networks (RNNs) (Zhang et al. 2019), deep belief networks (DBNs) (Zheng et al. 2014), graph convolutional neural networks (GCNNs) (Song et al. 2020; Zhang et al. 2021), and capsule networks (CapsNet) (Chao et al. 2019; Liu et al. 2020), to name a few. While many of them use feature extraction techniques (Nawaz et al. 2020) to extract statistical, power, frequency domain, entropy, or wavelet energy-based features to train their models, some have also utilized raw EEG signals as an input to their deep learning models.

Convolutional neural networks and recurrent neural networks have seen the most widespread use in the field of EEG emotion recognition in the past decade; however, recently, capsule networks (Chao et al. 2019; Liu et al. 2020), and graph neural networks (Song et al. 2020; Zhang et al. 2021) are also being used and providing state-of-the-art results. An increasing trend in the use of attention mechanisms in spatial, temporal, and spectral domains to extract more relevant information from the EEG signals is also observed.

Recently, transformers have emerged in the field of deep learning with their utility in natural language processing and computer vision. Models like VisionTransformers (Yuan et al. 2021), GPT-3 (Brown et al. 2020), and DALLE (Ramesh et al. 2021) are outperforming previous state-of-the-art methods and attaining better results. For EEG emotion classification, the use of transformers has been overlooked in the past, mainly because the previous transformer-based models were designed specifically for their respective modalities.

In the current work, we evaluate a recently proposed transformer-based architecture, perceiver (Jaegle et al. 2021), on a widely used publicly available dataset in the domain of EEG-based emotion recognition, DEAP. Perceiver can take inputs from different modalities, i.e., images, video, audio, 3D mesh points, etc. We preprocess EEG signals and map them to 2D matrix representation and then use perceiver to classify emotions from the raw EEG signals. Additionally, we compare our results with other baseline and state-of-the-art methods.

This paper is organized as follows. In Sect. 2, we provide the literature review, and in Sect. 3, the perceiver model is described. Section 4 illustrates experiment settings, training strategy, and preprocessing on DEAP datasets. Section 5 gives the discussion of the experimental results. Finally, we conclude this work in Sect. 7.

## 2 Literature review

In this section, we will briefly take a look at some of the methods and studies that are used for emotion classification. We will only focus on those studies that use raw EEG signals for training their models, as opposed to those that employ manual feature extraction techniques to train their models. Table 1 shows the different models used in different studies and their reported accuracies.

Alhagry et al. (2017) used LSTM-RNN to learn features from raw EEG signals and then used a dense layer in the end for classification, achieving 85.45%, 85.65%, and 87.99% accuracies using a fourfold cross-validation strategy on valence, arousal, and liking, respectively.

Wang et al. (2018) proposed EmotioNet, a 3D CNN-based architecture capable of recognizing emotions using raw EEG signals. It had an accuracy of 72.1% and 73.1% for valence and arousal on the DEAP dataset, respectively.

Zhang et al. (2019) used a specifically designed convolutional network to extract spatiotemporal information and then extracted the attentive temporal dynamics from raw EEG temporal slices for emotion classification. Using the same preprocessing and training strategy as ourselves, CRAM achieved an accuracy of 87.09% for valence and 84.46% for arousal on DEAP.

Chen et al. (2019) mirrored the hierarchical structure of EEG signals by introducing the attention mechanism combined with the hierarchical bidirectional gated recurrent unit (GRU) network and reported accuracies of 67.9% for valence and 66.5% for arousal on the DEAP dataset.

Tao et al. (2020) proposed ACRNN, which integrates channel-wise attention into a CNN that extracts spatial attentive features and channel attentive features. They also integrated extended self-attention into RNN to extract temporal attentive information, resulting in the reported accuracies

**Table 1** Details of several reported studies on DEAP dataset

| Studies | Models | Year | Accuracy (%) | | | |
|---|---|---|---|---|---|---|
| | | | Valence | Arousal | Dominance | Liking |
| Tao et al. (2020) | DT | 2020 | 75.95 | 78.18 | – | – |
| Wang et al. (2018) | EmotioNet | 2018 | 72.1 | 73.1 | – | – |
| Tao et al. (2020) | SVM | 2020 | 89.33 | 89.99 | – | – |
| Zhang et al. (2019) | CRAM | 2019 | 87.09 | 84.46 | – | – |
| Alhagry et al. (2017) | LSTM-RNN | 2017 | 85.45 | 85.65 | – | 87.99 |
| Chen et al. (2019) | H-ATT-BGRU | 2019 | 67.9 | 66.5 | – | – |
| Tao et al. (2020) | ACRNN | 2020 | 93.72 | 93.38 | – | – |
| Liu et al. (2020) | MLF-CapsNet | 2020 | 97.97 | 98.31 | 98.32 | – |

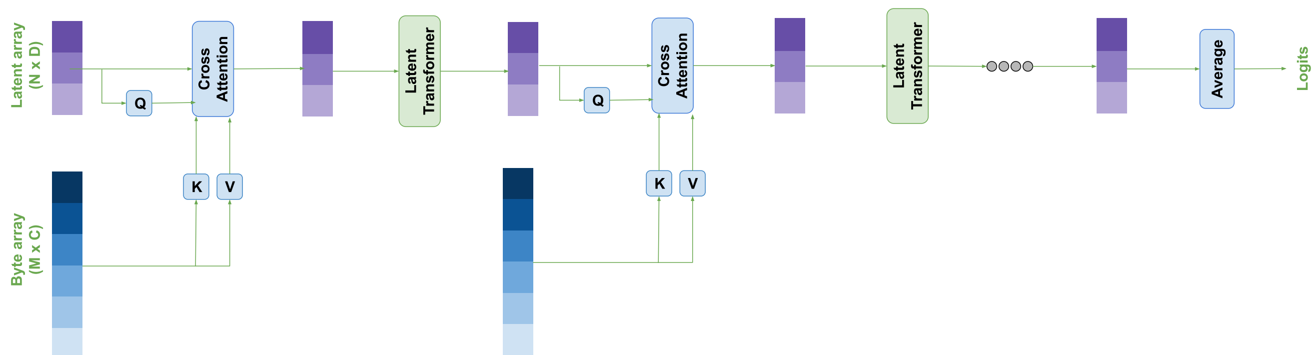


**Fig. 1** Without making any domain-specific assumptions, the perceiver architecture can scale to high-dimensional inputs such as images, videos, and audio. Using cross-attention module, the perceiver projects a high-dimensional input byte array to a fixed-dimensional latent array where $M \gg N$ and then processes it using a stack of transformer modules in low-dimensional latent space

of 93.72% and 93.38% for valence and arousal dimensions. Using the same preprocessing and training strategy as outlined in the present work, they reported the accuracies of 75.95% and 78.18% for valence and arousal using Decision Trees (DT), and 89.33% and 89.99% for valence and arousal using Support Vector Machines (SVM) on raw EEG signals in the DEAP dataset.

Liu et al. (2020) introduced MLF-CapsNet, which uses multi-level features extracted from different convolution layers to form primary capsules and reduced the number of parameters required by adding a bottleneck layer, resulting in reduced computation time. They reported accuracy of 97.97%, 98.31%, and 98.32% on valence, arousal, and dominance, respectively, using raw EEG signals from the DEAP dataset.

## 3 Method

For the past decade, ConvNets (Lecun et al. 1998) have been the dominant family of architectures in the field of deep learning because of their good performance and scalability. Due to their local type of computation — convolutions, they can conveniently handle high-resolution images. However, we are seeing widespread usage of the self-attention-based model, e.g., transformers, in language processing, image classification, and object detection. Although transformers are quite flexible and have shown amazing results, they scale poorly with the input size.

*Perceiver.* Perceiver, introduced in Jaegle et al. (2021), tries to make the transformers more scalable. Their model is built on two main architectural components: a cross-attention module and a transformer block, as shown in Fig. 1. The cross-attention module maps a byte array (input array) and a latent array, which is chosen to be much smaller than the byte array. The model alternates the application of cross-attention and transformer modules. The scalability issue of transformers is solved by projecting a high-dimensional input byte array through a lower-dimensional attentional bottleneck, before processing it with a stack of transformer modules.

To mitigate the potential information loss caused by the mapping of the byte array into a latent array, an iterative attentional approach is followed. The model is structured with multiple byte-attend layers, allowing the latent array to iteratively extract the required information from the byte array.

**Table 2** Description of DEAP dataset

| Materials | Setup | | |
|---|---|---|---|
| Number of participants | 32 | | |
| Number of videos | 40 | | |
| Recording signals | 32 EEG channels + 8 other peripheral channels | | |
| | Valence | Indicator of pleasantness | Float between 1 and 9 |
| Rating scales | Arousal | Intensity of the emotion | Float between 1 and 9 |
| | Dominance | Feeling of being in control of the emotion | Float between 1 and 9 |
| | Liking | Liking of the video | Float between 1 and 9 |

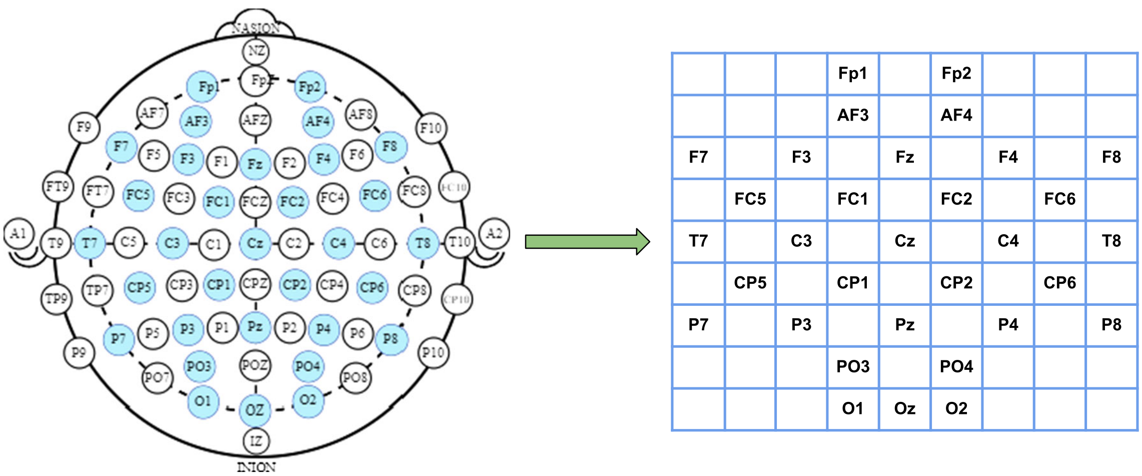


**Fig. 2** 2D Spatial mapping

## 4 Experiment

This paper focuses on the application of transformer architecture, namely the perceiver architecture, on raw EEG signals for emotion classification. In this section, we firstly introduce a widely used EEG dataset for emotion classification. Then, we describe the preprocessing, and experiment settings. We also describe the baseline models that were used for the comparison. Finally, the results on the datasets are reported and discussed.

### 4.1 DEAP dataset

In this work, we use the DEAP dataset (Koelstra et al. 2012), a multimodal dataset created by Koelstra et al., which is publicly available and many researchers have performed their analysis on it. Table 2 gives a brief description of the dataset.

The DEAP dataset collected EEG and peripheral signals from 32 subjects (16 males and 16 females, age ranged from 19 to 37, age mean=26.9). The EEG data were captured at a sampling rate of 512 Hz using 32 electrodes, while the subjects were watching 41 minutes long music videos carefully selected to elicit specific emotions. After watching each video, participants assessed the videos at different levels ranging from 1 (low) to 9 (high), along four dimensions, valence, arousal, dominance, and liking. Valence indicates pleasantness, while arousal is a measure of the intensity of the emotion varying from unexcited to excited and dominance represents the feeling of being in control of the emotion (Koelstra et al. 2012). Liking indicates the participant's likeness of the video. Each EEG signal contains a 3s baseline signal which was recorded in a relaxed state and a 60s experimental signal which was recorded under-stimulation.

Different researchers have used different threshold values for valence, arousal, dominance, and liking. In this work, we use the middle of the nine-point rating to generate two classes. When the rating is less than 5, we label it as low, and when the label is greater than or equal to 5, we label it as high. Another thing to keep in consideration is that the dominance scores of all the 40 experimental signals of the 27th subject are greater than 5, which results in the dominance labels with only one category, i.e., high. Therefore, we exclude the samples of the 27th subject to conduct experiments on the dominance as the model trained by such samples would be invalid.

The DEAP dataset also provides a preprocessed version, and we used the preprocessed version in the article. In the preprocessed version, EEG signals are down-sampled to 128Hz,

**Table 3** Accuracy of subjects in DEAP

| Subject | Valence | Arousal | Dominance | Liking |
|---|---|---|---|---|
| 1 | 91.19 | 90.81 | 92.79 | 99.52 |
| 2 | 86.93 | 89.42 | 90.91 | 87.77 |
| 3 | 94.98 | 96.78 | 94.44 | 98.39 |
| 4 | 85.41 | 85.63 | 89.42 | 85.97 |
| 5 | 90.80 | 91.93 | 91.64 | 90.56 |
| 6 | 90.01 | 86.71 | 87.61 | 91.93 |
| 7 | 91.22 | 88.18 | 86.63 | 98.19 |
| 8 | 91.42 | 91.91 | 92.15 | 97.27 |
| 9 | 88.93 | 87.65 | 89.11 | 93.63 |
| 10 | 94.54 | 93.08 | 93.97 | 95.56 |
| 11 | 82.93 | 85.09 | 86.91 | 83.95 |
| 12 | 88.29 | 94.78 | 88.42 | 87.72 |
| 13 | 84.40 | 94.27 | 83.37 | 83.39 |
| 14 | 86.62 | 89.97 | 86.59 | 90.32 |
| 15 | 91.52 | 94.47 | 91.82 | 93.67 |
| 16 | 93.97 | 93.35 | 92.37 | 93.40 |
| 17 | 83.39 | 84.44 | 95.14 | 85.07 |
| 18 | 92.65 | 92.20 | 96.11 | 90.85 |
| 19 | 91.53 | 91.21 | 89.64 | 92.53 |
| 20 | 92.69 | 95.51 | 91.45 | 95.64 |
| 21 | 92.00 | 94.82 | 91.01 | 91.69 |
| 22 | 94.31 | 95.89 | 95.79 | 94.03 |
| 23 | 92.49 | 94.11 | 95.32 | 92.43 |
| 24 | 90.61 | 96.01 | 90.05 | 88.71 |
| 25 | 89.71 | 91.89 | 98.92 | 89.65 |
| 26 | 91.72 | 89.09 | 89.06 | 95.52 |
| 27 | 94.99 | 92.95 | – | 97.39 |
| 28 | 88.67 | 90.13 | 90.87 | 89.66 |
| 29 | 92.01 | 93.25 | 91.28 | 90.91 |
| 30 | 94.56 | 92.88 | 94.61 | 93.83 |
| 31 | 88.55 | 87.15 | 89.13 | 93.08 |
| 32 | 90.19 | 92.01 | 97.96 | 90.49 |
| Average | 90.41 | 91.49 | 91.43 | 91.96 |
| Median | 91.20 | 91.97 | 91.28 | 92.18 |
| Std. | 3.30 | 3.36 | 3.56 | 4.17 |
| Min | 82.93 | 84.44 | 83.37 | 83.39 |
| Max | 94.99 | 96.78 | 98.92 | 99.52 |

and a band-pass frequency filter from 4.0−45.0 Hz is applied to remove the artifacts.

### 4.2 Preprocessing

As deep learning models require sufficient data to obtain meaningful results, we segment each experimental signal along the temporal dimension. A 1s sliding window is used to segment an experimental signal into 60 non-overlapping segments, each containing 128 sampling points. As a result, we obtain 2400 (40 trails $\times$ 60 segments) EEG samples for each subject, and each sample is a 32 $\times$ 128 matrix.

Although many researchers have used carefully extracted features along spatial, spectral, temporal, and statistical dimensions, in this analysis, we will only use raw EEG signals. For the sake of fairness, we also exclude those papers from the comparison which use manual feature extraction strategies to train their models.

As the perceiver model can take input in any modality, we experimented with two input sizes, one where each sample out of 2400 samples is of shape 32 $\times$ 128, meaning all the electrode values were given as a 1D vector, and one where we mapped the electrodes according to their spatial locations in the international 10-20 system, as shown in Fig. 2.

### 4.3 Experiment settings

We use a tenfold cross-validation strategy to evaluate the performance of the perceiver on raw EEG signals from the DEAP dataset, as has been done in many of the previous studies (Liu et al. 2020; Tao et al. 2020; Wang et al. 2018; Zhang et al. 2019). Even though this results in an increase in the training time, this strategy makes use of all the available dataset for training the model, and it also gives a more reliable accuracy. Typically, tenfold cross-validation divides data into 10 equal data subsets where one subset is used as the test set, and the other nine subsets form the training set. This process is repeated 10 times. We take the average accuracy of the tenfold cross-validation as the result of one subject, and then the average accuracy of all the subjects as the final accuracy. We used the Adam optimizer to minimize the margin loss function. We set the learning rate, batch size, and the number of epochs to $10^{-4}$, 16, and 8, respectively. For the perceiver model, we set its depth to 6, the number of latent dimensions to 512, the drop-out value of attention and feed forward layer to 0.25. We did not use any weight sharing between cross-attention and transformer block modules.

## 5 Results

After training the perceiver model on raw EEG signals from the DEAP dataset using a tenfold training strategy, for each label dimension, i.e., valence, arousal, dominance, and liking, we obtained the average accuracy of 90.41%, 91.49%, 91.43% and, 91.96% respectively. These results are subject-dependent, meaning that the model was trained on a single subject where some trials were used as training set and the remaining trials as test set, as compared to subject-independent or cross-subject, where the model is trained on a subset of subjects and evaluated on the remaining subjects. Table 3 shows the individual accuracies for

**Table 4** Comparison with several reported studies on DEAP dataset

| Studies | Models | Year | Accuracy (%) | | | |
|---|---|---|---|---|---|---|
| | | | Valence | Arousal | Dominance | Liking |
| Tao et al. (2020) | DT | 2020 | 75.95 | 78.18 | – | – |
| Wang et al. (2018) | EmotioNet | 2018 | 72.1 | 73.1 | – | – |
| Tao et al. (2020) | SVM | 2020 | 89.33 | 89.99 | – | – |
| Zhang et al. (2019) | CRAM | 2019 | 87.09 | 84.46 | – | – |
| Alhagry et al. (2017) | LSTM-RNN | 2017 | 85.45 | 85.65 | – | 87.99 |
| Chen et al. (2019) | H-ATT-BGRU | 2019 | 67.9 | 66.5 | – | – |
| Tao et al. (2020) | ACRNN | 2020 | 93.72 | 93.38 | – | – |
| Liu et al. (2020) | MLF-CapsNet | 2020 | 97.97 | 98.31 | 98.32 | – |
| Perceiver (proposed) | | 2021 | 90.41 | 91.49 | 91.43 | 91.96 |

valence, arousal, dominance, and liking for all 32 subjects in the DEAP dataset. The results shown here are obtained from the DEAP preprocessed dataset where the EEG signals were mapped into their respective 2D spatial representations. We also tried the 1D signal representation but it performed poorly as compared to the 2D representation, signifying the importance of 2D mapping and the ability of the model to learn from spatial dimensions of the EEG signal.

For a fair comparison of the perceiver model with other baseline and state-of-the-art methods for EEG emotion recognition, we chose the models where the preprocessing and training strategy was similar to ours, and the model used raw EEG signals as input, instead of manually extracted statistical, temporal, or frequency-based features.

As given in Table 4, perceiver performs better than all the other models with an exception of ACRNN and MLF-CapsNet.

## 6 Discussion

Our proposed model, perceiver, performs better than most of the previous methods, but it also reports lower accuracies as compared to the ACRNN(Tao et al. 2020) and MLF-CapsNet(Liu et al. 2020) models.

A major reason behind the better performance of our model as compared to previous state-of-the-art approaches like CNN, LSTM, SVM, DT, etc., is that our model uses transformer architecture which has proven to be more generalizable and which is able to learn relevant features across long sequences, as evident from their success in computer vision, natural language processing, and many other domains. As EEG signals can be treated as a long sequence of numerical values, a transformer-based architecture, which utilizes self-attention mechanism, is a more suitable choice to attend to those features that are relevant and responsible for a certain emotional state. This increased generalizability and the ability to learn long-term dependencies result in higher accuracies for the perceiver model.

Even though our model gives better accuracies as compared to the previous baseline and state-of-the-art methods, it is not able to beat ACRNN (Tao et al. 2020) and MLF-CapsNet (Liu et al. 2020) models. One of the main reasons behind this is that the ACRNN and MLF-CapsNet models are specifically designed for EEG dataset and emotion classification, while perceiver is a general architecture that can be used for images, audio, video, and further modalities. Moreover, capsule networks (Sabour et al. 2017) have been shown to work really well with EEG data (Chao et al. 2019; Liu et al. 2020; Zhang and Etemad 2021) because of the small data size and improved representational capacity of capsule networks.

## 7 Conclusion

This work presented an analysis of using a transformer-based architecture, perceiver, for emotion classification using raw EEG signals. We performed experimentation on the DEAP dataset, which is a publicly available EEG dataset for emotion classification, and compared its results with other baseline and state-of-the-art methods. Because of its generalizability and multimodal input accommodation, perceiver performed fairly well-compared to widely used baseline methods from previous years. However, it was not able to beat two methods specifically designed to work with EEG and emotion classification. This study shows the potential impact of using transformers in the domain of EEG emotion recognition. In the future, more specialized transformer-based architectures can be specifically designed to work with EEG data for emotion recognition.

**Acknowledgements** The authors would like to thank GIK Institute for providing research facilities. This work was sponsored by the Higher Education Commission (HEC) under the National Research Programme for Universities (NRPU) having project number 8898.

**Author Contributions** The authors contributed to each part of this paper equally. The authors read and approved the final manuscript.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Human and animal rights** This article does not contain any studies with human participants performed by any of the authors.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References


Alhagry S, Fahmy AA, El-Khoribi RA (2017) Emotion Recognition based on EEG using LSTM Recurrent Neural Network. Int J Adv Comput Sci Appl (IJACSA). 8(10). https://doi.org/10.14569/IJACSA.2017.081046

Anderson K, McOwan P (2006) A real-time automated system for the recognition of human facial expressions. IEEE Trans Syst Man Cybernet Part B (Cybernet) 36(1):96–105. https://doi.org/10.1109/TSMCB.2005.854502

Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler DM, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D (2020) Language Models are Few-Shot Learners. arXiv:2005.14165 [cs]

Chao H, Dong L, Liu Y, Lu B (2019) Emotion recognition from multiband EEG signals using capsnet. Sensors 19(9):2212. https://doi.org/10.3390/s19092212

Chen JX, Jiang DM, Zhang YN (2019) A hierarchical bidirectional GRU model with attention for EEG-based emotion classification. IEEE Access 7:118530–118540. https://doi.org/10.1109/ACCESS.2019.2936817

Collobert R, Weston J (2008) A unified architecture for natural language processing: deep neural networks with multitask learning. In Proceedings of the 25th international conference on Machine learning, ICML '08, New York, NY, USA. Association for Computing Machinery, pp 160–167

Deng X, Zhu J, Yang S (2021) SFE-Net: EEG-based Emotion Recognition with Symmetrical Spatial Feature Extraction. arXiv:2104.06308 [cs, eess]

Ding Y, Robinson N, Zeng Q, Guan C (2021) April. TSception: Capturing Temporal Dynamics and Spatial Asymmetry from EEG for Emotion Recognition. arXiv:2104.02935 [cs]

Halim Z, Atif M, Rashid A, Edwin CA (2017) Profiling players using real-world datasets: clustering the data and correlating the results with the big-five personality traits. IEEE Trans Affect Comput. https://doi.org/10.1109/TAFFC.2017.2751602

Halim Z, Rehan M (2020) On identification of driving-induced stress using electroencephalogram signals: a framework based on wearable safety-critical scheme and machine learning. Inf Fusion 53:66–79. https://doi.org/10.1016/j.inffus.2019.06.006

He K, Zhang X, Ren S, Sun J (2016) Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 770–778. ISSN: 1063-6919

Jaegle A, Gimeno F, Brock A, Zisserman A, Vinyals O, Carreira J (2021)Perceiver: general perception with iterative attention. arXiv:2103.03206 [cs, eess]

Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I (2012) DEAP: a database for emotion analysis; using physiological signals. IEEE Trans Affect Comput 3(1):18–31. https://doi.org/10.1109/T-AFFC.2011.15

Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324. https://doi.org/10.1109/5.726791

Liu H, Guo H, Hu W (2021) EEG-based Emotion Classification Using Joint Adaptation Networks. In *2021 IEEE international symposium on circuits and systems (ISCAS)*, pp 1–5. ISSN: 2158-1525

Liu X, He P, Chen W, Gao J (2019) Multi-Task Deep Neural Networks for Natural Language Understanding. arXiv:1901.11504 [cs]

Liu Y, Ding Y, Li C, Cheng J, Song R, Wan F, Chen X (2020) Multichannel EEG-based emotion recognition via a multi-level features guided capsule network. Comput Biol Med 123:103927. https://doi.org/10.1016/j.compbiomed.2020.103927

Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B (2021) Swin Transformer: hierarchical vision transformer using shifted windows. arXiv:2103.14030 [cs]

Muhammad T, Halim Z (2016) Employing artificial neural networks for constructing metadata-based model to automatically select an appropriate data visualization technique. Appl Soft Comput 49(C):365–384. https://doi.org/10.1016/j.asoc.2016.08.039

Nawaz R, Cheah KH, Nisar H, Yap VV (2020) Comparison of different feature extraction methods for EEG-based emotion recognition. Biocybernet Biomed Eng 40(3):910–926. https://doi.org/10.1016/j.bbe.2020.04.005

Petrushin V (2000) Emotion in speech: recognition and application to call centers. Proceedings of artificial neural networks in engineering

Ramesh A, Pavlov M, Goh G, Gray S, Voss C, Radford A, Chen M, Sutskever I (2021) Zero-Shot Text-to-Image Generation. arXiv:2102.12092 [cs]

Sabour S, Frosst N, Hinton GE (2017) Dynamic Routing Between Capsules. arXiv:1710.09829 [cs]

Soleymani M, Pantic M, Pun T (2012) Multimodal emotion recognition in response to videos. IEEE Trans Affect Comput 3(2):211–223. https://doi.org/10.1109/T-AFFC.2011.37

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. https://doi.org/10.1109/TAFFC.2018.2817622

Tao W, Li C, Song R, Cheng J, Liu Y, Wan F, Chen X (2020) EEG-based emotion recognition via channel-wise attention and self attention. IEEE Trans Affect Comput. https://doi.org/10.1109/TAFFC.2020.3025777

Tripathi S, Acharya S., Sharma RD, Mittal S, Bhattacharya S (2017)Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset. In twenty-ninth IAAI conference

Uzma, Halim Z (2021) An ensemble filter-based heuristic approach for cancerous gene expression classification. Knowl-Based Syst 234:107560. https://doi.org/10.1016/j.knosys.2021.107560

Wang Y, Huang Z, McCane B, Neo P (2018) EmotioNet: a 3-D convolutional neural network for EEG-based emotion recognition. In 2018 international joint conference on neural networks (IJCNN), pp 1–7. ISSN: 2161-4407

Wu X, Zheng WL, Lu BL (2020) Investigating EEG-based functional connectivity patterns for multimodal emotion recognition. arXiv:2004.01973 [cs]

Xiao G, Ye M, Xu B, Chen Z, Ren Quansheng (2021) 4D attention-based neural network for EEG emotion recognition. arXiv:2101.05484 [cs]

Yang Y, Wu Q, Fu Y, Chen X (2018) Continuous convolutional neural network with 3D input for EEG-based emotion recognition. In: Cheng L, Leung ACS, Ozawa S (eds) Neural information processing. Lecture notes in computer science. Springer International Publishing, Cham, pp 433–443

Yao Z, Wang Z, Liu W, Liu Y, Pan J (2020) Speech emotion recognition using fusion of three multi-task learning-based classifiers: HSF-DNN. MS-CNN and LLD-RNN. Speech Commun 120:11–19. https://doi.org/10.1016/j.specom.2020.03.005

Yin Y, Zheng X, Hu B, Zhang Y, Cui X (2021) EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM. Appl Soft Comput 100:106954. https://doi.org/10.1016/j.asoc.2020.106954

Yuan L, Chen Y, Wang T, Yu W, Shi Y, Jiang Z, Tay FE, Feng J, Yan S (2021) Tokens-to-Token ViT: training vision transformers from scratch on imagenet. arXiv:2101.11986 [cs]

Zhang D, Yao L, Chen K, Monaghan J (2019) A convolutional recurrent attention model for subject-independent EEG signal analysis. IEEE Signal Process Lett 26(5):715–719. https://doi.org/10.1109/LSP.2019.2906824

Zhang G, Etemad A (2021) Distilling EEG Representations via Capsules for Affective Computing. arXiv:2105.00104 [cs]

Zhang G, Yu M, Liu YJ, Zhao G, Zhang D, Zheng W (2021) SparseDGCNN: recognizing emotion from multichannel EEG signals. IEEE Trans Affect Comput. https://doi.org/10.1109/TAFFC.2021.3051332

Zheng WL, Zhu JY, Peng Y, Lu BL (2014) EEG-based emotion classification using deep belief networks. In 2014 IEEE international conference on multimedia and expo (ICME), pp 1–6. ISSN: 1945-788X