



ScienceDirect®

Journal of Neuroscience Methods

Volume 410, October 2024, 110223

Lightweight attention mechanisms for EEG emotion recognition for brain computer interface

Naresh Kumar Gunda ^a✉, Mohammed I. Khalaf ^b✉, Shaleen Bhatnagar ^c✉, Adam Quraishi ^d✉,
Leeladhar Gudala ^e✉, Ashok Kumar Pamidi Venkata ^f✉, Faisal Yousef Alghayadh ^g✉,
Shtwai Alsubai ^h✉, Vaibhav Bhatnagar ⁱ✉

Show more ▾

☰ Outline | Share Cite

<https://doi.org/10.1016/j.jneumeth.2024.110223> ↗

[Get rights and content](#) ↗

Highlights

- Identifying emotions from EEG data is a difficult due to signal complexities.
- A lightweight network is provided to improve accuracy of emotion identification.
- Differential entropy features of EEG signals are used as inputs.
- Proposed model achieved 95.18 % accuracy on the SEED dataset.
- Proposed method enhances the model's ability to aggregate features.

Abstract

Background

In the realm of brain-computer interfaces (BCI), identifying emotions from electroencephalogram (EEG) data is a difficult endeavor because of the volume of data, the intricacy of the signals, and the several channels that make up the signals.

New methods

Using dual-stream structure scaling and multiple attention mechanisms (LDMGEEG), a lightweight network is provided to maximize the accuracy and performance of EEG-based emotion identification. Reducing the number of computational parameters while maintaining the current level of classification accuracy is the aim. This network employs a symmetric dual-stream architecture to assess separately time-domain and frequency-domain spatio-temporal maps constructed using differential entropy features of EEG signals as inputs.

Result

The experimental results show that after significantly lowering the number of parameters, the model achieved the best possible performance in the field, with a 95.18 % accuracy on the SEED dataset.

Comparison with existing methods

Moreover, it reduced the number of parameters by 98 % when compared to existing models.

Conclusion

The proposed method distinct channel-time/frequency-space multiple attention and post-attention methods enhance the model's ability to aggregate features and result in lightweight performance.



Previous



Next

Keywords

Brain-computer interface; EEG emotion; Attention mechanisms; Deep learning

1. Introduction

The notion that "computation is related to emotions, generating and acting upon emotions" ([Pan et al., 2023](#)) was initially proposed by Professor Picard at the MIT Media Lab in 1997, giving rise to the field of affective computing as a subfield of artificial intelligence. Enabling robots to recognize and respond to human emotional states is the goal. Healthcare, education, transportation, the military, and entertainment are just a few of the industries where affective computing has seen tremendous growth in recent years. Behavioral and physiological signs can be used to express emotions. In contrast to behavioral indications like posture, facial expressions, and speech, physiological signals like breathing, blood pressure, and EEG are increasingly being used in emotional computing. Emotions may be represented by physiological and behavioral signals. Physiological signals, including respiration, blood pressure, and EEG, are progressively

being employed in affective computing as opposed to behavioral signals such as posture ([Peng et al., 2022](#)), facial expressions ([Zheng, Sept. 2017](#), [Liu et al., Nov. 2022](#)), and voice. This is primarily due to the latter's low cost, resistance to subjective disguise, and high temporal resolution. EEG, comprising multi-channel time-series data obtained directly from the brain, provides a comprehensive representation of the electrophysiological activity of nerve cells located on the cortex or cranium of the brain. Moreover, it is replete with emotion-related insights. Research has demonstrated that EEG signals associated with distinct emotions vary considerably with respect to temporal, frequency, and spatial (channel) characteristics. A technique that enables communication between two or more network stack layers, either through information sharing or the translation of messages into instructions to satisfy vertical optimization requirements, is referred to as a cross-layer approach, also known as interlayer processing. The Network Model's layered architecture breaks down a single network process into smaller tasks. Every small task is then allocated to a specific layer, which processes it exclusively.

As an illustration, it can be observed that positive emotions elicit greater activation in the frontal and temporal lobes when compared to negative emotions, while negative emotions elicit greater activity in the low-frequency bands than the high-frequency bands ([Asif et al., 2023](#)). EEG is therefore capable of objectively reflecting distinct individual emotions. As a result, researchers have become increasingly intrigued by emotion recognition based on EEG, which has evolved into a significant representation of emotions.

Since AlexNet ([Gao et al., 2022](#)) won the ImageNet challenge, deep learning algorithms have gradually gained widespread application. In recent years, leveraging their excellent automatic feature extraction

capabilities, deep learning algorithms have also been applied to EEG affective computing. The general trend is to design deeper and more complex networks to achieve higher accuracy, such as DBN ([Asif et al., 2023](#)), DGCNN ([Feng et al., 2022](#)), RGNN ([Xu et al., 2023](#)), and SST-EmotionNet ([Zhang, 2020](#)). Using various attention processes and dual-stream structure scaling, LDM-EEG is a unique lightweight EEG emotion identification model that aims to balance classification accuracy and parameter volume in neural networks for EEG emotion recognition. Higher accuracy, however, does not always equate to efficiency in terms of network speed and size. Deep learning models necessitate high-performance hardware, take longer to train and infer, and have limited applicability on embedded systems and mobile platforms. To achieve a balance between parameter volume and classification accuracy in neural networks for EEG emotion recognition, this paper proposes LDM-EEG, a novel lightweight EEG emotion recognition model based on multiple attention mechanisms and dual-stream structure scaling. An essential part of the dual-stream neural network is the Residual Learning Module. One convolution layer and four stacked residual learning modules make up each branch network. Only the final two residual learning modules have numerous attention modules based on the post-attention process. These residual learning modules have the same structure as those without attention mechanisms, apart from the attention modules. To increase the performance and accuracy of EEG-based emotion recognition, a lightweight network is given using dual-stream structure scaling and multiple attention mechanisms (LDMGEEG). The objective is to minimize the number of computational parameters while retaining the present degree of classification accuracy. This model addresses the issue of EEG emotion recognition in an effective manner. These are the primary contributions of this paper:

- (1) Tested on public datasets for accuracy and parameter volume, this model surpasses existing mainstream models in precision and achieves the lowest parameter volume. By scaling the model, our model reduces parameter volume by 98 %, slightly lower in accuracy than the current best results, and enhances training efficiency.
- (2) Proposed a novel parameter-saving dual-stream scalable residual module, using depth wise separable convolution and the same dual-stream structure to streamline network design, effectively avoiding overfitting and significantly reducing parameter volume.
- (3) Proposed a novel multiple attention mechanism and explored the impact of the number and position of attention modules on classification accuracy for the first time in the EEG field, effectively demonstrating the efficacy of multiple attention mechanisms and posterior attention mechanisms.

2. Related work

2.1. EEG emotion recognition

EEG emotion recognition typically involves two fundamental steps: feature extraction and classification. Frequent characteristics encompass both time-domain and frequency-domain attributes. Event Related Potentials (ERP) ([Peng et al., 2021](#)), Hjorth features ([Kim et al., 2022](#)), Non-Stationary Index (NSI)

([Zhang et al., 2023](#)), and so forth, are time-domain characteristics. Higher Order Spectra (HOS) ([Peng et al., 2022](#)), Band Power (BP) ([Peng et al., 2021](#), [Liu et al., 2018](#)), and other characteristics exist in the frequency domain. Differential Entropy (DE) is the most frequently employed feature for emotion classification in EEG ([Tian et al., 2021](#), [Hou et al., 2023](#)). Zheng et al. initially established the accuracy and stability of DE features as classification features through the development of a deep belief network ([Hou et al., 2023](#)). Since the electrode channels are directly correlated with the cranium electrodes of the subjects, a spatial positional relationship exists between the channels. Building the topological structure between channels or transforming it into a two-dimensional form can more successfully capture the relationships and spatial positioning information between channels than the oversimplified method of stacking channel information. Constructed the topological link among EEG channels to extract the spatial positioning information between channels.

The spatial positional information between channels was extracted by Song et al. through the construction of the topological relationship among EEG channels ([Feng et al., 2022](#)). EEG signals were converted to image-based representations by Jung et al., which led to successful classification outcomes ([Pereira et al., 2021](#)).

Many traditional machine learning methods, such as K-Nearest Neighbor (KNN) ([Song et al., 2020](#)) and Support Vector Machine (SVM) ([Wu et al., 2023](#)), were used extensively in the early days of EEG emotion categorization. Traditional machine learning algorithms, on the other hand, rely mostly on human expertise in feature selection and design, which significantly limits their accuracy. Deep learning has recently supplanted other methods for EEG emotion categorization, because to the proliferation of large-

scale EEG datasets and the lightning-fast development of GPUs. Some examples of such algorithms are Graph Neural Networks (GNN) ([Peng et al., 2023](#)), Recurrent Neural Networks (RNN) ([Shen et al., 2023](#)), and Convolutional Neural Networks (CNN) ([Li et al., 2023](#)). The time-domain information can be effectively extracted by RNN and RNN, whilst the spatial information can be effectively extracted by CNN and GNN. The task of recognizing emotions from electroencephalogram (EEG) data is challenging in the field of brain-computer interfaces (BCI) due to the large amount of data, the complex nature of the signals, and the several channels that comprise the signals. A brain-computer interface (BCI) is a computer system that receives, processes, and converts brain signals into commands that are sent to an output device so that a desired action can be performed. A direct communication channel between the electrical activity of the brain and an external device—most often, a computer—is known as a brain–computer interface (BCI), also occasionally referred to as a brain–machine interface (BMI). To get generalizations from electroencephalogram (EEG) data, for example, Zheng et al. used a deep belief network (DBN) ([Asif et al., 2023](#), [Hou et al., 2023](#)). Graph Convolutional Neural Networks (DGCNNs) were introduced in the literature as a means of emotion recognition in electroencephalogram (EEG) data ([Feng et al., 2022](#)). The most commonly used feature in EEG for the classification of emotions is differential entropy (DE), which was first shown to be accurate and stable as classification features through the construction of a deep belief network. Since the electrode channels are directly correlated with the subjects' cranium electrodes, there is a spatial positional relationship between the channels; this means that, as opposed to the oversimplified method of layering channel information, the correlations and spatial positional information between channels can be more efficiently captured by building the topological arrangement among channels or translating it into the two-dimensional format. To learn about global and regional

spatiotemporal features for EEG emotion identification, Literature. created a Hierarchical Spatio-Temporal Neural Network (R2G-STNN) ([Cheng et al., 2021](#)).

The above features and models often consider only one or two combinations of the three domains of time, frequency, and space, failing to fully extract the information from these three features. Jia et al.'s work SST-EmotionNet ([Zhang, 2020](#)) used DE as an EEG feature and utilized the spatial positional information of channels to construct a 3D representation input into a spatiotemporal-frequency dual-stream network, making full use of information from the time, frequency, and spatial domains, achieving good emotion classification results. This paper is an improvement based on this work.

2.2. Attention mechanism

The attention mechanism, which is frequently employed in speech recognition ([Li et al., 2022](#)), image recognition ([Gu et al., 2023](#)), natural language processing ([Apicella et al., 2022](#)), and other applications, is a technique that improves neural network performance by emphasizing relevant information and suppressing irrelevant information. In essence, the attention technique is similar to how humans observe outside objects. When humans look at an object, they often concentrate on a few key local details at first, combining information from several areas to create a general sense of the object. Comparably, the attention mechanism aids in the model's ability to give various weights to every component of the input X, extracting more crucial and significant data and enabling the model to reach more correct conclusions.

Most studies conducted in the last few years on the attention mechanism have formed the attention mechanism using masks. It entails learning weight distributions from pertinent feature maps and appending the weights that have been learned to the original feature map. The topological link among EEG channels in order to extract the spatial positioning information between channels. Several conventional machines learning techniques, including Support Vector Machine (SVM) and K-Nearest Neighbour (KNN), were widely employed in the initial stages of EEG emotion classification. However, the accuracy of traditional machine learning algorithms is severely limited because they mostly rely on human skill in feature selection and design. For instance, CBAM ([Wang et al., 2024](#)) combined feature channel and feature spatial dimensions for attention mechanisms, SKNet ([Li et al., 2021](#)) added attention mechanisms for varying convolution kernel sizes, and FcaNet ([Sarma and Barma, 2022](#)) proposed a multi-spectral channel attention mechanism to investigate the effects of using various frequency domain components and their combinations. SENet ([Wang et al., 2023](#)) added the attention mechanism to the feature channels through the Squeeze and Excitation module. Measuring and analyzing changes in different human bodily characteristics or electrical impulses in the nervous system is the usual method used for automated emotion identification. Electroencephalography, skin resistance testing, blood pressure, heart rate, eye activity, and motion analysis are the most often used methods. While you look at a signal in the frequency domain representation, you can see several aspects of the signal that are either difficult to see or not visible at all while looking at the signal in the time domain. For example, frequency-domain analysis is helpful when attempting to identify a signal's cyclical behavior.

A sensible number of attention mechanisms and their locations must be chosen, as the attention mechanism itself has significant computational and storage overhead. In PPG-LCNet ([Song et al., 2019](#)), the effects of including SE attention mechanisms ([Wang et al., 2023](#)) at the start, middle, and end of the model were compared. The results showed that, when the delay circumstances were the same, the attention module at the end of the model obtained higher accuracy than those at the beginning or midway. On this, BotNet ([Li et al., 2022](#)) and offTANet ([Zheng, 2017](#)) carried out few trials. Multi-channel, multi-frequency, and time-series data make up EEG data. When it comes to emotion identification, different timestamps, frequency bands, and channels have different levels of significance. This makes attention research essential to the field of EEG emotion recognition. In database management systems, timestamps are essential for monitoring data changes. Databases can effectively facilitate data retrieval and auditing by employing timestamps to record the date and time that records were added or changed. They are also used in temporal databases to implement data versioning and time-based searches.

Jia ([Zhang, 2020](#)) improved the model's classification performance by incorporating time/frequency-space attention processes. Chai ([Lin, 2020](#)) added channel and frequency attention processes to the model, achieving good classification results on the DEAP and DREAMER datasets. Tao ([Gilakjani and Osman, 2024](#)) added channel attention mechanisms to the model, enhancing the classification accuracy of EEG emotion recognition. The model may combine features more effectively and perform with less weight when channel-time/frequency-space multiple attention and post-attention techniques are used. By utilizing attention processes that are time/frequency-space based. Improved the model by including channel and frequency attention processes, which produced accurate classification results on the

DREAMER and DEAP datasets. Improved the model's channel attention mechanisms, which improved the EEG emotion recognition's classification accuracy. Nevertheless, the multi-attention mechanism of channel-time/frequency-space has not been thoroughly investigated in previous work, and all attention mechanisms are applied evenly to every module of the network. This research aims to investigate the number and location of its modules as well as the multi-attention mechanism of channel-time/frequency-space.

2.3. Network lightweighting

Although the classification accuracy of deep learning models has increased, these models frequently include enormous parameter volumes, which increases the time and resource requirements for training and inference. Methods to minimize model parameters usually sacrifice efficiency for accuracy to achieve a balance between network accuracy and efficiency. EEG-based emotion identification is more objective and precise than human movements and speech. Neural network models have been successfully applied to EEG emotion identification with good results because to the advancements in deep learning technology. The brain serves as the body's information processing and control center. The central nervous system directly produces electroencephalogram (EEG) physiological signals, which are strongly associated with human emotions. As a result, human emotional states can now be objectively and instantly reflected in EEG data. Model pruning ([Lu et al., 2023](#)) and distillation ([Wang et al., 2022](#)) are two of these techniques. One method is to manually build lightweight networks, as demonstrated by the ShuffleNet series, which uses the concept of grouped convolutions, and the MobileNet series ([Li et al., 2024](#), [Li et al., 2022](#), [Du et al., 2023](#)), which mainly uses depth wise separable convolutions for lightweight networks. Furthermore,

neural architecture search (NAS) has gained popularity as a method for creating lightweight, effective networks. By carefully adjusting convolution kernel types and sizes, network width, and depth, NAS can achieve higher efficiency than manually created networks. Use an Emotive EEG device to record data to increase the precision and effectiveness of EEG-based emotion identification. The participants score the EEG data according to levels of dominance, valence, and arousal. It is investigated how the affective ratings and the powers of various EEG bands correlate. Your EEG may not always indicate that you did not experience a seizure. EEGs performed on people experiencing seizures are considered normal. A normal EEG test can be obtained even by someone who experiences seizures once a week. This is so because the EEG only records brain activity while the test is being administered.

Although they have achieved good accuracy, existing deep learning models for EEG emotion categorization rarely take parameter volume into account. Deep learning models have become more accurate in classifying data; nevertheless, they often involve large volumes of parameters, increasing the time and resources needed for training and inference. To strike a compromise between network accuracy and efficiency, methods to minimize model parameters typically trade off efficiency for accuracy. The length of time it takes a machine learning model to digest fresh input and produce a prediction is known as the inference time in deep learning. Big datasets are used to train deep learning models. The model can swiftly make predictions on new data once it has been trained. To balance accuracy and parameter volume, this study scales the network parameters with inspiration from EfficientNet and applies the lightweight network design ideas of MobileNet to the backbone network. In addition to speeding up training and inference, using a lightweight network for EEG emotion classification also makes it easier to install models

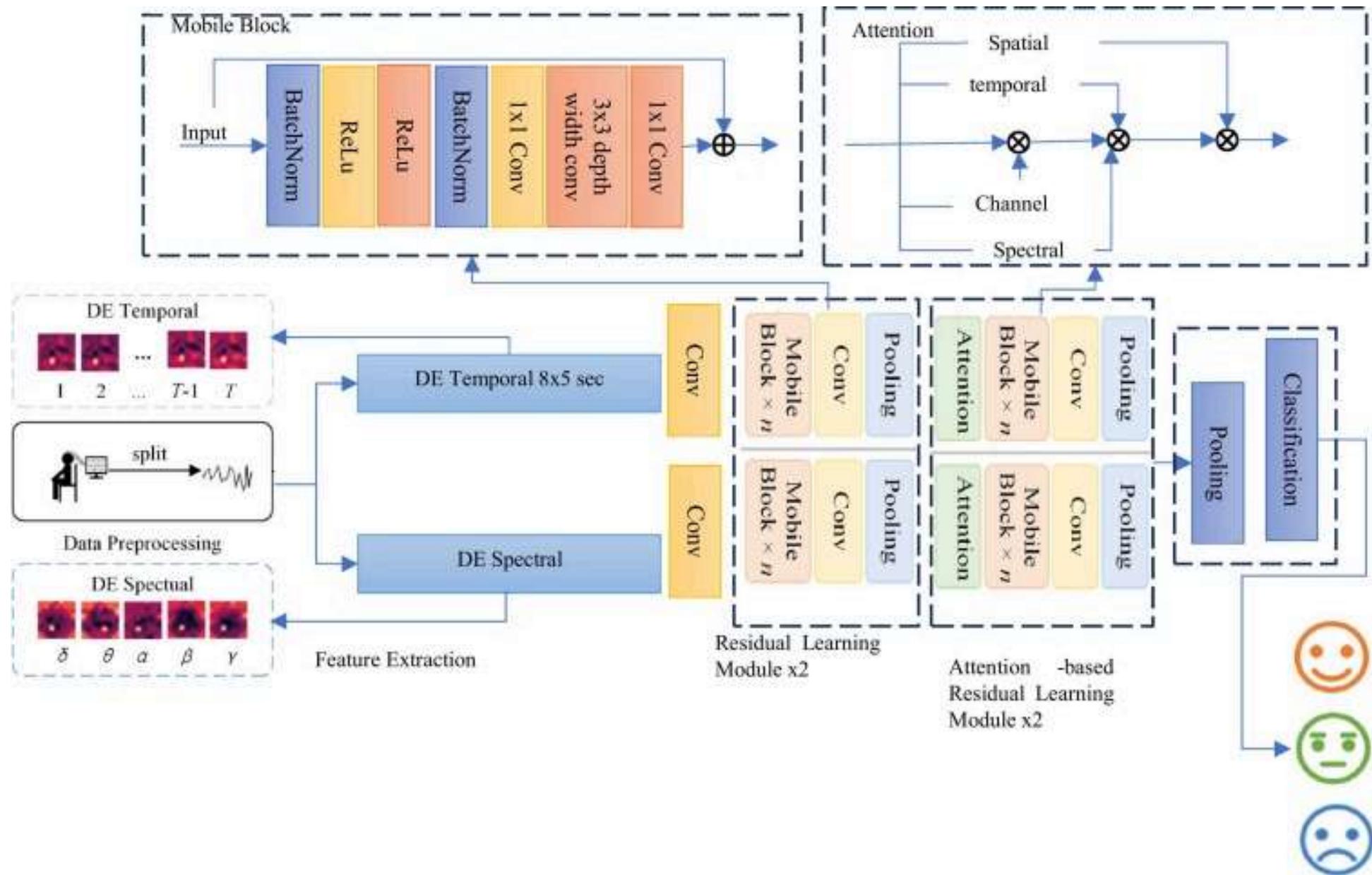
on-device, improving real-time performance. [Zheng et al. \(2017\)](#) developed a multi-attention convolutional neural network with local models and fine-grained feature representations for collaborative learning. The network includes convolutional, channel grouping, and part classification subnetworks. This promotes mutual reinforcement between part generation and feature learning. A Subspace Alignment Auto-Encoder (SAAE) with non-linear transformation and consistency constraints is used by [Chai et al. \(2016\)](#). This study compared the findings to TCA. Compared to TCA's 73.82 % accuracy on the SEED dataset, it obtains a leave-one-out mean of 77.88 %. The mean classification accuracy for session-to-session evaluation is 81.81%, a 1.62 % improvement over the best baseline TCA. Another study ([Zheng et al., 2015](#)) used CNN with the deep domain confusion technique to achieve 90.59 % and 82.16 mean accuracy for traditional (subject-dependent) EEG emotion recognition and "leave one out cross validation" using the SEED dataset. Other studies using domain adaptation strategies on the SEED dataset include ([Zheng and Lu, 2016](#), [Chai et al., 2017](#)). Domain adaptation strategies involve exploring an invariant feature subspace to eliminate EEG data discrepancies between people or sessions.

3. Method

3.1. Model overview

The overall framework of the method proposed in this paper is shown in [Fig. 1](#). This framework is mainly divided into two parts: one part is the EEG signal preprocessing and feature extraction module, and the other part is the attention mechanism-based time-frequency dual-stream residual learning emotion

classification module. The network's dual-stream architecture extracts the EEG's spatiotemporal and spatiofrequency properties, respectively. We blend these features for classification to obtain improved classification accuracy. DualRAN is a dual-stream architecture made up mostly of modules that are aware of both the local and global environment, simulating a discussion from several angles at the same time. Moreover, we create two DualRAN single-stream network variations.



Download: [Download high-res image \(448KB\)](#)

Download: [Download full-size image](#)

Fig. 1. Time-frequency dual-stream residual learning emotion classification model framework based on attention mechanism.

The EEG signal preprocessing part includes raw data segmentation and the construction of a 3D representation of the EEG in the time-frequency domain. The emotion classification module consists of a symmetric time-frequency dual-stream network structure, with inputs being the 3D time-domain representation and the 3D frequency-domain representation of the EEG signals. The brain serves as the body's information processing and control center. The central nervous system directly produces electroencephalogram (EEG) physiological signals, which are strongly associated with human emotions. As a result, human emotional states can now be objectively and instantly reflected in EEG data. Increased right hemisphere beta asymmetry is suggestive of stress or anxiety. Stressful situations can result from excessive beta waves. Rapid gamma waves and highly concentrated attention are linked to certain gamma activity. Like the design of ResNeXt [47], the time-frequency dual-stream in this network adopts the same topology. This design aims to prevent model overfitting and improve the model's adaptability to different tasks and datasets. Using depth-wise separable convolution and the same dual-stream structure, a unique parameter-saving dual-stream scalable residual module streamlines network construction, efficiently preventing overfitting and drastically lowering parameter volume. To enhance the model's flexibility to various tasks and datasets and avoid overfitting. There are four phases of residual learning modules in each branch; the attention mechanism is included in the last two modules while it is absent in the first two. In order to categorize emotions, the network finally combines the time-frequency dual-stream properties.

3.2. 3D time-frequency representation of EEG signals

The number of channels (62) and the sampling frequency (200) of the EEG signal determine the dimension of a data sample after data preprocessing and segmentation. The data are further passed via time domain and frequency domain feature extraction algorithms before being fed into the time domain and frequency domain network branches, respectively. Automated feature extraction eliminates the need for human interaction by extracting features from signals or images using deep networks or specialized algorithms. When you want to construct machine learning algorithms quickly from raw data, this technique might be quite helpful. Image files are a common use case for feature extraction, which converts unprocessed data into numerical features that may be used with machine learning techniques. By identifying an object's shape or an image's redness value, data scientists can provide new features that are appropriate for machine learning applications. What follows is a description of the particulars.

Making complete use of EEG data from various frequency bands is essential for the frequency domain branch's feature extraction process. The EEG signal is initially separated into the δ , θ , α , β , and γ frequency bands, as described in references [48–52]. You can express the b -th frequency band signal as

$F_b = (f_b^1, f_b^2, \dots, f_b^{E-1}, f_b^E)$, where b is an element of the set $b \in \{\delta, \theta, \alpha, \beta, \gamma\}$, E is the channel count, and f_b is a single-channel EEG signal with 200 points of sampling. Then, for every frequency band, we compute the DE characteristics of every channel. Eq. (1) shows that the DE characteristic is obtained from the Shannon entropy.

$$DE = - \int_{-\infty}^{+\infty} f(x) \log(f(x)) dx \quad (1)$$

The probability density function of continuous information, $f(x)$, is one of these functions. If the length of an EEG signal x roughly follows a normal distribution $N(\mu, \sigma^2)$, then $f(x)$ can be written as:

$$f(x) = \frac{1}{\sqrt[2]{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

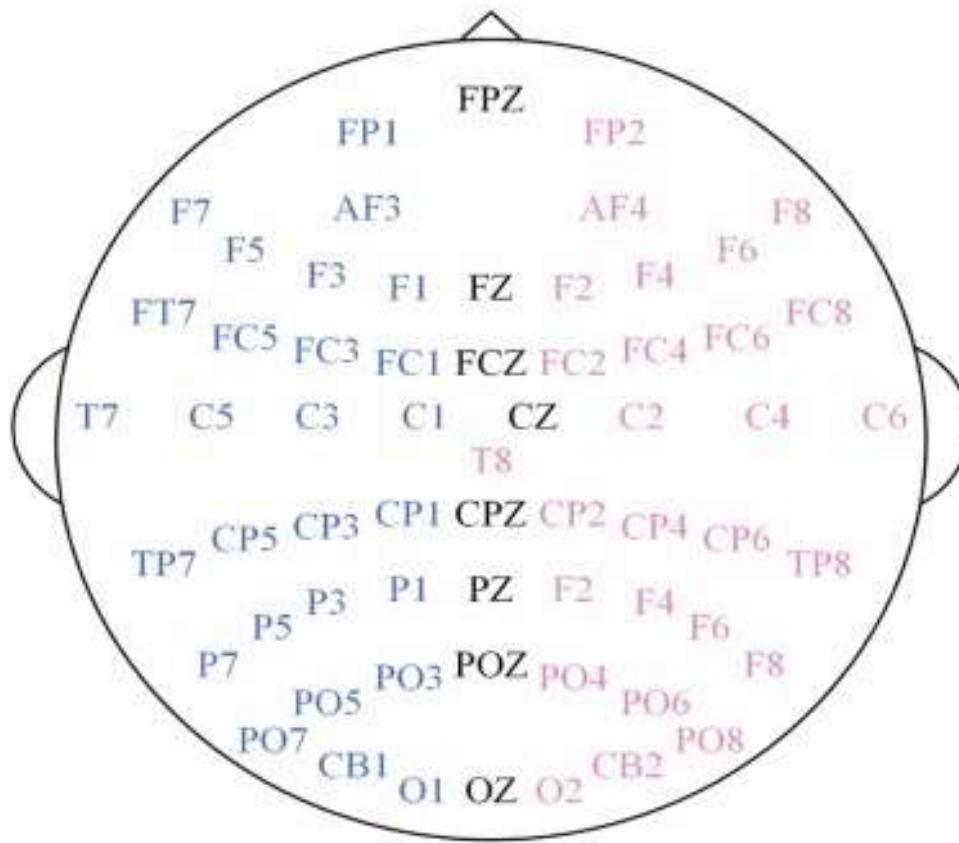
Substituting Eq. (2) into Eq. (1), we can obtain the simplified result of the DE feature formula, as shown in Eq. (3).

$$DE = \frac{1}{2} \log 2\pi\sigma^2 \quad (3)$$

Currently, we have transformed the time-series data of each channel into DE features. An extensive depiction of the electrical activity of nerve cells on the cortex or skull of the brain is provided by electroencephalography (EEG), which consists of multi-channel time-series data that is directly acquired from the brain. It is also filled with insights about emotions. Studies have indicated that there are significant differences in the temporal, frequency, and spatial (channel) properties of EEG signals linked to different emotions.

Then, for the DE feature data of 62 channels in each frequency band, they are converted into a 9×9 2D data format (as shown in Fig. 3) according to the relative spatial position relationship between the channels (as shown in Fig. 2), where positions without data are processed by zero-padding. To refine the information, cubic spline interpolation (INTER_CUBIC) is used to convert the 9×9 2D single-band EEG

signals into a 32×32 2D format. Since there are a total of 5 bands, 3D data representation is finally obtained, as shown in Fig. 4.



[Download: Download high-res image \(117KB\)](#)

[Download: Download full-size image](#)

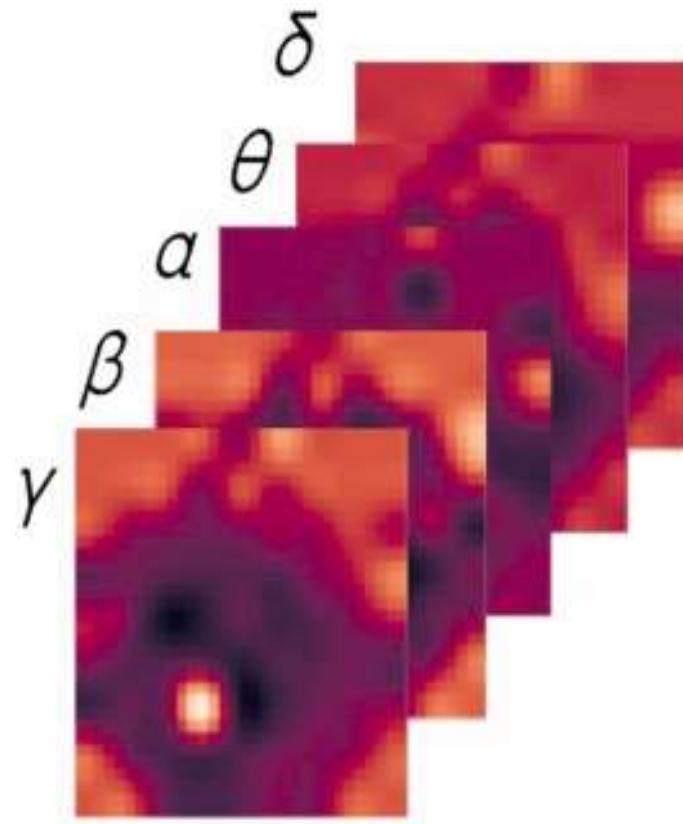
Fig. 2. Schematic diagram of the 62-channel EEG cap of the international 10G20 system 1).

			FP1	FPZ	FP2			
			AF3		AF4			
F7	F5	F3	F1	FZ	F2	F4	F6	F8
FT7	FC5	FC3	FC1	FCZ	FC2	FC4	FC6	FT8
T7	C5	C3	C1	CZ	C2	C4	C6	T8
TP7	CP5	CP3	CP1	CPZ	CP2	CP4	CP6	TP8
P7	P5	P3	P1	PZ	P2	P4	P6	P8
	PO7	PO5	PO3	POZ	PO4	PO6	PO8	
		CB1	O1	OZ	O2	CB2		

Download: [Download high-res image \(155KB\)](#)

Download: [Download full-size image](#)

Fig. 3. Schematic diagram of converted 2-dimensional EEG representation.



[Download: Download high-res image \(70KB\)](#)

[Download: Download full-size image](#)

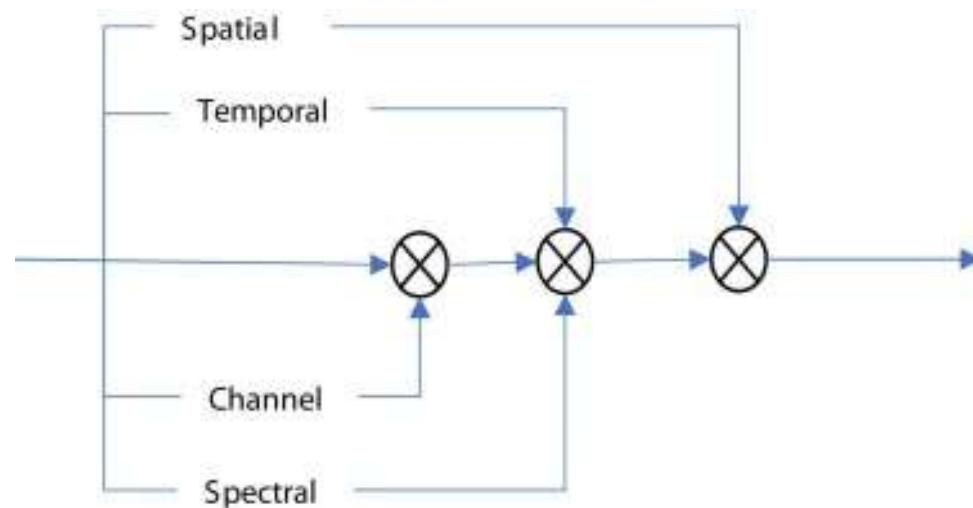
Fig. 4. 3D characteristic map of EEG signals in the frequency domain.

Similarly, in the feature extraction process of the time domain branch, it is necessary to fully utilize the EEG information at different timestamps. First, the EEG signals are divided into T groups according to a certain number of sampling points. The signal at the t -th timestamp can be expressed as

$F_t = (f_t^1, f_t^2, \dots, f_t^{E-1}, f_t^E)$, where $t \in \{1, 2, \dots, T-1, T\}$, E represents the number of channels, and f_t represents the single-channel EEG signal with $200T$ sampling points. Then, the DE features of all channel data at each timestamp are calculated, and the DE feature data of each channel at each timestamp are converted into a 32×32 2D format in the same way as in the frequency domain. Since there are a total of T timestamps, 3D data representation is finally obtained.

3.3. Attention mechanism

The activation degree of different emotions on different network channels, timestamps, frequency bands, and brain regions varies (Asif et al., 2023). Therefore, we designed a multi-attention mechanism in the spatial domain of the channel G time/frequency G (as shown in Fig. 5) to adaptively focus on specific patterns that recognize different emotions.



Download: [Download high-res image \(58KB\)](#)

[Download: Download full-size image](#)

Fig. 5. Channel G time/frequency G space domain triple attention mechanism.

To be more precise, the EEG signals that are fed into the network may be depicted as either $\in \mathbf{R}^{c \times t \times h \times w}$, or $\mathbf{F} \in \mathbf{R}^{c \times s \times h \times w}$ with the former indicating the EEG signals that are fed into the flow of the network in the frequency domain and the latter into the flow of the network in the time domain. Metrics including accuracy, precision, recall, F1-score, and mean squared error are frequently used to quantify accuracy. Efficiency is the amount of time, memory, energy, and hardware that an artificial neural network (ANN) consumes to produce its outputs. Through the use of attention mechanisms, the network is able to adjust to contextual cues, preferentially focus on important information, and strike a more advantageous balance between accuracy and processing efficiency. In this case, c is the channel count, s is the frequency band count (with an initial value of 5), t is the timestamp count (with an initial value of 8), and h and w are the 2D EEG signal width and length, respectively, with initial values of 32.

3.3.1. Channel attention

The channel attention module draws on the ideas of SENet ([Wang et al., 2023](#)), which automatically learns the importance of each feature channel and then enhances useful channels while suppressing channels that are less useful for the current task. Specifically, it first averages the w and h dimensions of the input feature map \mathbf{F} to aggregate spatial information, generating $\mathbf{F} \in \mathbf{R}^{c \times \frac{s}{t} \times 1 \times 1}$. Then, it averages the s/t dimension to aggregate information across frequency bands or timestamps, resulting in $\mathbf{F} \in \mathbf{R}^{c \times 1 \times 1 \times 1}$.

Finally, the channel attention matrix is realized through a fully connected layer with SoftMax activation.

$$\left\{ \begin{array}{l} \&F_{c,s/t,h} = GA P_w(F_{c,s/t,h,w}) \\ \&F_{c,s/t} = GA P_h(F_{c,s/t,h}) \\ \&F_c = GA P_{s/t}(F_{c,s/t}) \\ \&M_{elect} = \text{softmax}(F_c) \end{array} \right. \quad (4)$$

Here, $F_{c,s/t,h}$, $F_{c,s/t}$, F_c represent the sequentially aggregated information, and $GA P_w$, $GA P_h$, $GA P_{s/t}$ denote the averaging operations over the respective dimensions. M_{elect} represents the channel attention map. Finally, the generated attention map M_{elect} is multiplied with the input feature map F to obtain the final feature map F' with channel attention mechanism, as shown in Eq. (5).

$$F' = M_{elect} \odot F \quad (5)$$

3.3.2. Frequency attention

The intensity of emotional expression varies across different timestamps, and the importance of different frequency bands in recognizing different emotions also varies. There have been many dimensional models of emotion established, but just a handful are still widely acknowledged as the dominant models at this time. The vector model, the circumplex model, and the Positive Activation – Negative Activation (PANA) model are the three most well-known two-dimensional models. Main emotions, such as happiness, sorrow, anger, fear, contempt, and surprise, and secondary emotions, which arouse a mental image associated with a memory or main emotion, are the two categories into which emotions can be classified.

To achieve more accurate classification, it is necessary to adaptively learn the importance of timestamps and frequency bands. Therefore, we designed a time/frequency attention mechanism. Similarly, it first averages the w and h dimensions of the input feature map F to aggregate spatial information, generating $F \in R^{c \times s/t \times 1 \times 1}$. Then, it averages the c dimension to aggregate channel information, resulting in $F \in R^{1 \times s/t \times 1 \times 1}$. Finally, the time/frequency attention matrix is realized through a fully connected layer with SoftMax activation, as shown in [Eq. \(6\)](#).

$$\left\{ \begin{array}{l} & \& F_{c,s/t,h} = GAP_w(F_{c,s/t,h,w}) \\ & \& F_{c,s/t} = GAP_h(F_{c,s/t,h}) \\ & \& F_{s/t} = GAP_c(F_{c,s/t}) \\ & \& M_{s/t} = \text{softmax}(F_{s/t}) \end{array} \right. \quad (6)$$

Here, $F_{c,s/t,h}$ and $F_{c,s/t}, F_{s/t}$ represent the sequentially aggregated information, and GAP_w, GAP_h, GAP_c denote the averaging operations over the respective dimensions. $M_{s/t}$ represents the time/frequency attention map. Finally, the generated time/frequency attention map $M_{s/t}$ is multiplied with the input feature map F' to obtain the final feature map F'' with time/frequency attention mechanism, as shown in [Eq. \(7\)](#).

$$F'' = M_{s/t} \odot F \quad (7)$$

3.3.3. Spatial attention

The degree to which various brain regions are activated changes in response to various emotional states. In order to tackle this, we present a spatial attention mechanism that will enhance classification accuracy even further. In the same way as earlier attention approaches, it aggregates information from the temporal and frequency domains by averaging the s/t dimensions of the input feature map F , producing $F \in R^{1 \times 1 \times h \times w}$. Afterwards, $F \in R^{c \times 1 \times h \times w}$ is obtained by averaging the c dimension in order to aggregate channel information. Eq. (8) shows that a completely connected layer with SoftMax activation is also used to realize the final spatial attention matrix.

$$\left\{ \begin{array}{l} & F_{c,h,w} = GA P_{s/t}(F_{c,s/t,h,w}) \\ & F_{h,w} = GA P_c(F_{c,h,w}) \\ & M_{h,w} = \text{softmax}(F_{h,w}) \end{array} \right. \quad (8)$$

In this case, the aggregated data is denoted by $F_{c,h,w}$ and $F_{h,w}$, while the averaging processes across the corresponding dimensions are denoted by $GA P_{s/t}$ and $GA P_c$. $M_{h,w}$ is the spatial attention map that was constructed. As seen in Eq. (9), the output feature map F''' with spatial attention mechanism is obtained by multiplying the produced spatial attention map with the input feature map F'' .

$$F''' = M_{h,w} \odot F'' \quad (9)$$

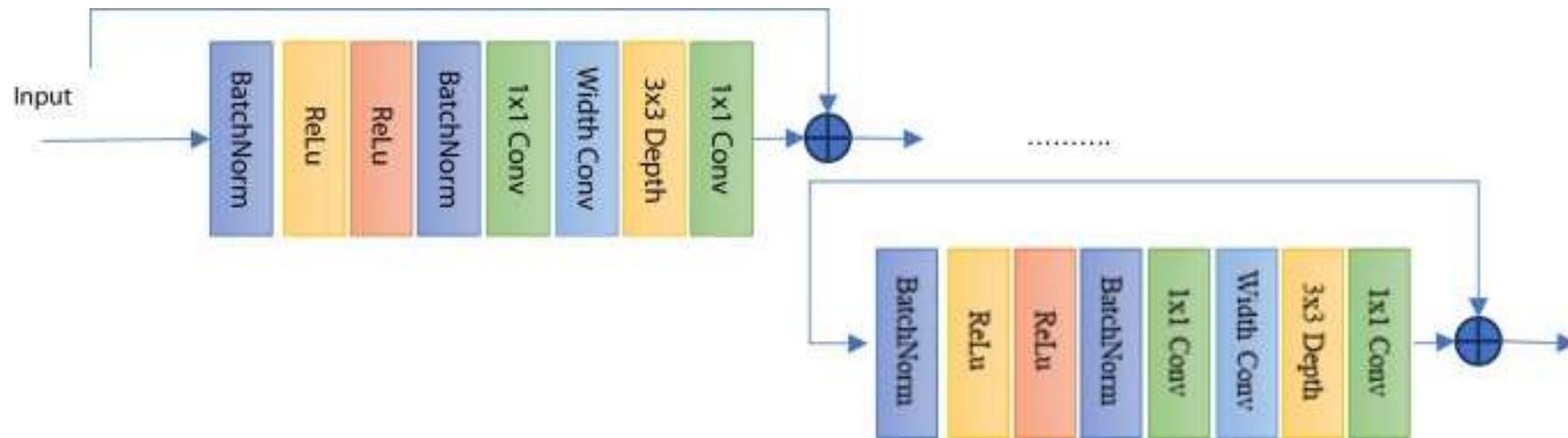
3.3.4. Post-attention mechanism

The attention mechanism helps the network focus on more effective information but also increases the training and inference time. Therefore, adding attention modules at all stages of the network is not the

best choice. To address this, we designed experiments to explore the effectiveness of attention modules in different relative positions within the network. We observed that in the four residual learning modules, models with attention mechanisms in the last two modules performed better than those with attention mechanisms in all modules, while also having fewer parameters.

3.4. Residual learning module

The Residual Learning Module is an important component of the dual-stream neural network. Each branch network consists of one convolution layer and four stacked residual learning modules. Based on the post-attention mechanism, only the last two residual learning modules include multiple attention modules. Other than the attention modules, the structure of these residual learning modules is the same as those without attention mechanisms. Specifically, a residual learning module is primarily composed of a MobileBlock (as shown in Fig. 6), convolution, pooling, and possibly attention modules. The MobileBlock consists of multiple identical residual structures stacked together, with the number of residual structures referred to as the depth scaling factor, adjustable in size, and set to 2 here.



[Download: Download high-res image \(133KB\)](#)

[Download: Download full-size image](#)

Fig. 6. A scalable MobileBlock based on inverted residual and depth-separable convolution.

Taking the first residual learning module of the frequency domain branch in the network as an example, the detailed information such as the resolution and feature channels of the input image is listed in [Table 1](#). Stage 1 is the 3D convolution module before the residual learning module, Stages 2 and 3 are MobileBlock modules containing two stacked residual structures, and Stages 4 and 5 are 3D convolution and average pooling modules, respectively. The stacked residual structures of the MobileBlock make full use of the feature information while addressing the difficulty of training deep neural networks [\[53\]](#). The basic

principle is shown in Eq. (10).

$$\begin{cases} \&y_l = h(x_l) + F(x_l, W_l) \\ \&x_{l+1} = f(y_l) \end{cases} \quad (10)$$

Table 1. Internal structure of the residual learning module.

Stage	Layer	Operation	Resolution	Channels
1	-	Conv3D	$5 \times 32 \times 32$	4
2	1	BatchNormReLU	$5 \times 16 \times 16$	4
	2	Conv3D	$5 \times 16 \times 16$	32
	3	BatchNormReLU	$5 \times 16 \times 16$	32
	4	DW Conv3D	$5 \times 16 \times 16$	32
	5	Conv3D	$5 \times 16 \times 16$	4
	6	BatchNormReLU	$5 \times 16 \times 16$	8
3	1	Conv3D	$5 \times 16 \times 16$	64
	2	BatchNormReLU	$5 \times 16 \times 16$	64
	3	DW Conv3D	$5 \times 16 \times 16$	64
	4	Conv3D	$5 \times 16 \times 16$	64

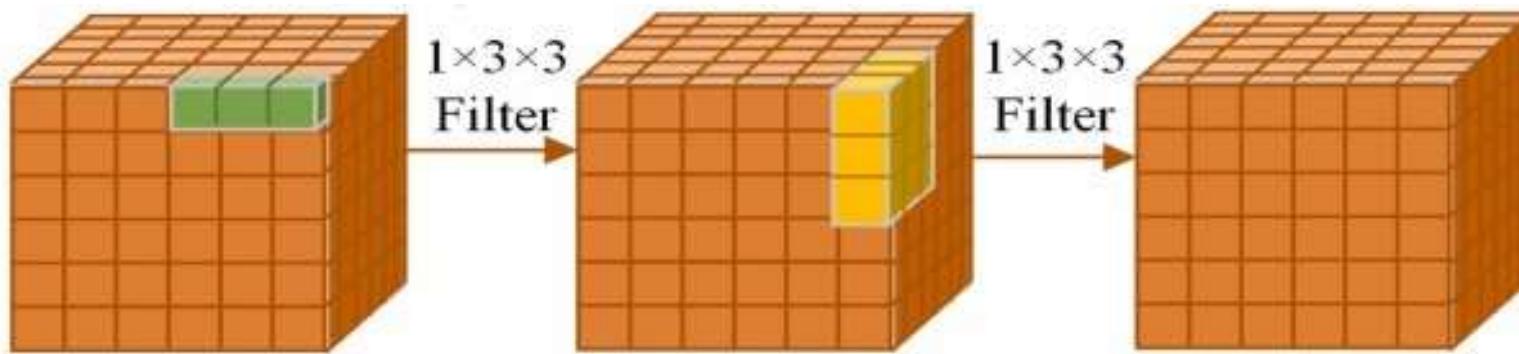
Stage	Layer	Operation	Resolution	Channels
	5	Conv3D	$5 \times 16 \times 16$	8
4	-	Conv3D	$5 \times 16 \times 16$	8
5	-	AvgPooling	$5 \times 8 \times 8$	8

Among them, x_l and x_{l+1} are the input and output of the l-th residual unit; $h(x_l)$ is the identity function; F is the residual function, representing the learned residual; f is the ReLU activation function.

Taking a page out of MobileNet's playbook, we use depthwise separable convolution for the residual block's internal structure rather than regular 3D convolution to keep the network lightweight without sacrificing classification performance. The term "grouped convolution" refers to a collection of convolutions for the layer that have several kernel groups and matching various output channel groups. Due to a shortage of graphics card memory, the model is split between two GPUs. Depth wise separable convolution layers form the foundation of mobile networks. There are two types of convolutions in each depth wise separable convolution layer: pointwise and depth wise.

Ordinary 3D convolution, Pseudo 3D convolution, and 3D depthwise separable convolution are the three main methods for implementing 3D convolution [54]. The use of neural architecture search (NAS) to build efficient, light-weight networks has grown in popularity. NAS can outperform manually built networks by carefully setting convolution kernel types and sizes, network width, and depth. This study scales the

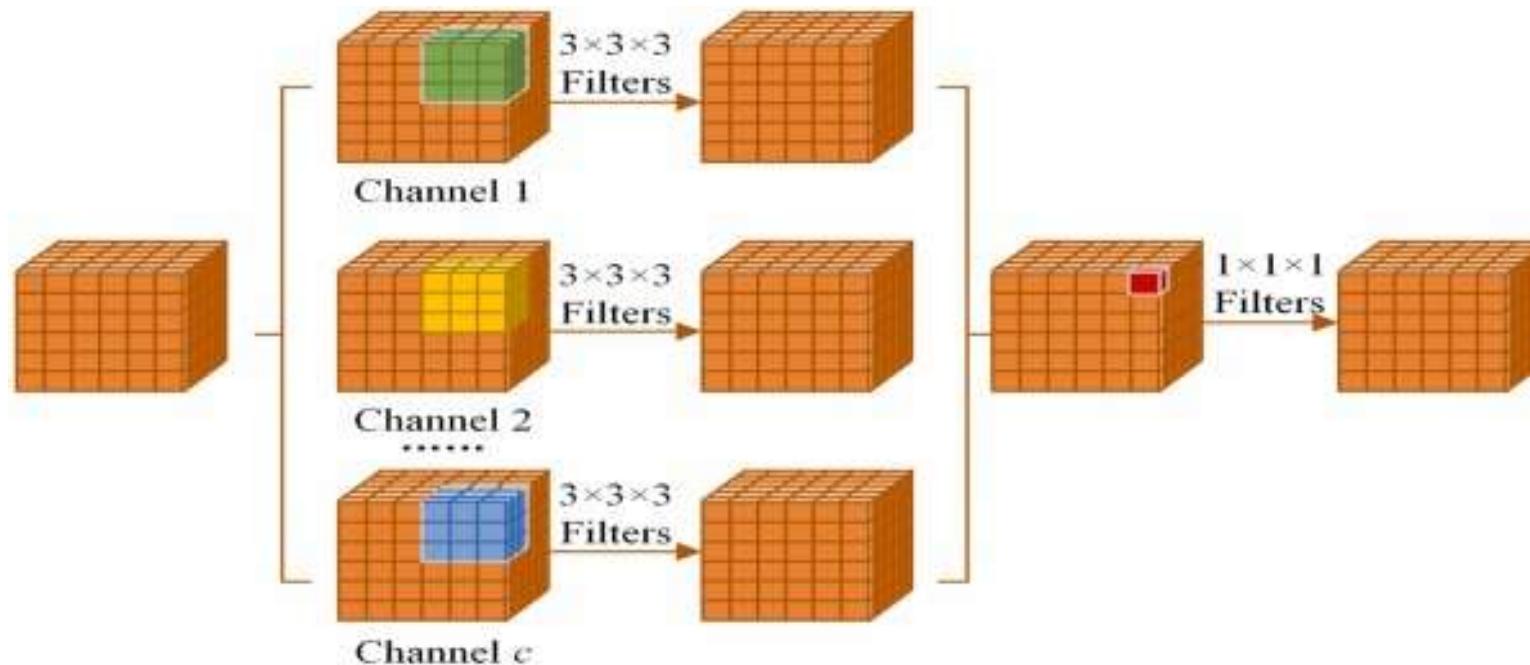
network parameters, drawing inspiration from EfficientNet, and applies MobileNet's lightweight network design principles to the backbone network to achieve a balance between accuracy and parameter volume. Using a lightweight network for EEG emotion categorization not only expedites training and inference but also simplifies model installation on-device, hence enhancing real-time performance. Fig. 7 illustrates the fundamental concept of Pseudo 3D convolution, which entails simulating the commonly used $3 \times 3 \times 3$ kernel 3D convolution using a 2D spatial convolution and a 1D temporal convolution with kernel sizes of $1 \times 3 \times 3$ and $3 \times 1 \times 1$, respectively. By separating standard three-dimensional convolution into its component parts—depthwise and pointwise—we get depthwise separable convolution. Fig. 8 illustrates that pointwise convolution utilises a 1×1 kernel, whereas depthwise convolution assigns a distinct convolutional kernel to every input channel.



Download: [Download high-res image \(113KB\)](#)

Download: [Download full-size image](#)

Fig. 7. Pseudo 3D convolution.



[Download: Download high-res image \(163KB\)](#)

[Download: Download full-size image](#)

Fig. 8. 3D depth separable convolution.

Comparing their parameter sizes, the relationship is often: ordinary 3D convolution > Pseudo 3D convolution > 3D depthwise separable convolution. For example, for 3D input data with shape l, w, h , processed by a 3D convolution with kernel size k , the parameter size for ordinary 3D convolution is

$$P_{\text{Conv}} = k \times k \times k \times c_b \times c_a \times l_F \times w_F \times h_F;$$

for Pseudo 3D convolution, it is
 $P_{\text{P3DConv}} = k \times k \times c_b \times c_b \times l_F \times w_F \times h_F + k \times c_b \times c_a \times l_F \times w_F \times h_F$; for 3D depthwise separable convolution, it is
 $P_{\text{DWConv}} = k \times k \times k \times c_b \times l_F \times w_F \times h_F + c_b \times c_a \times l_F \times w_F \times h_F$. The comparison

among them is shown in Eq. (11):

$$\begin{aligned} \mathcal{G} \frac{P_{\text{DWConv}}}{P_{\text{Conv}}} &= \frac{1}{c_a} + \frac{1}{k^3} \\ \mathcal{G} \frac{P_{\text{DWConv}}}{P_{\text{P3DConv}}} &= \frac{k^3 + c_a}{k^2 \times c_b + k \times c_a} \approx \frac{k}{c_b} \end{aligned} \quad (11)$$

The input data F's length, width, and height are represented by lF, wF, and hF, respectively, while the number of channels before and after processing is denoted by cb. It is evident that depthwise separable convolution uses a lot fewer parameters than regular convolution because all of these variables are either 1 or more. Deepwise separable convolution uses fewer parameters than Pseudo 3D convolution because, in most cases, the kernel size k is smaller than the number of output channels. In the section on the ablation experiment, we will compare and contrast the three types of convolutions in terms of their parameters and classification accuracy.

Additionally, considering that the number of feature maps extracted by depthwise separable convolution is closely related to the number of input channels, an inverted residual strategy (Li et al., 2022) is adopted to make the extracted information as rich as possible. This involves first using a 1×1 convolution to expand the number of input channels to eight times the original, processing it with a 3×3 convolution, and then reducing the number of channels through another 1×1 convolution. The channel number changes in the ordinary residual structure are for reducing computational load, but with the lightweight depthwise separable convolution, saving parameters is less meaningful and might lose useful information. Moreover, a convolutional layer and pooling layer are connected after the MobileBlock to further aggregate feature information.

3.5. Model scaling

To better balance the number of parameters and accuracy, the idea of EfficientGNet[46] is applied to scale some parameters of the network. EfficientGNet's scaling parameters include the input image resolution, network width (number of channels), and network depth (number of sub-module stacks). Given the limited resolution of the 2D EEG input data, we explore width and depth, which will be detailed in the ablation experiment section.

3.6. Feature fusion and classification

The dual-stream structure of the network extracts spatio-temporal and spatio-frequency features of EEG respectively. To achieve higher classification accuracy, we fuse these features for classification. Before feature fusion, adaptive pooling is applied to convert the size of the 3D feature maps to $1 \times 1 \times 1$. Then, the downsampled spatio-temporal and spatio-frequency features are concatenated along the channel dimension and sent to the classifier for processing. Given the large input feature dimension, the classifier consists of two fully connected layers with a dropout operation added between them to prevent overfitting. Crossentropy is used as the loss function to quantify the difference between the predicted and true distributions, as shown in Eq. (12).

$$\text{Loss} = - \sum_{c=1}^M p_c \log (q_c) \quad (12)$$

Where Loss represents the loss, M represents the total number of categories, p_c represents the true distribution, and q_c represents the predicted distribution.

4. Experiments

4.1. Dataset

To ensure the proposed strategy worked, experiments were carried out using the SEED dataset ([Asif et al., 2023](#), [Tian et al., 2021](#)). All three of these emotions—positive, neutral, and negative—are present in the SEED dataset. Fifteen video clips, one for each of the three emotions, each lasting four minutes to avoid monotony and featuring instantly identifiable content that evokes the desired emotion, serve to elicit these feelings. The data collection involved fifteen subjects, each of whom saw fifteen video segments over the course of three trials. Four minutes of video, 45 seconds of self-assessment, a 5-second start prompt, and a 15-second respite make up each clip. You won't see clips evoking different emotions one after the other. In order to document their emotional reaction, subjects fill out a questionnaire right after each clip.

Preprocessing is done on the raw EEG signals. A 62-channel EEG equipment based on the worldwide 10–20 system with a 1000 Hz sample rate is used in the tests. A bandpass filter ranging from 0 to 75 Hz is applied to the raw EEG signals before downsampling them to 200 Hz. As a last step before feeding the model, the EEG signals are divided into 200 ms samples, and DE features are extracted by several processes, including the Fourier transform.

4.2. Experimental configuration

The Adam optimizer was used to train and evaluate the suggested approach on an NVIDIA RTX 8000 GPU in order to optimise the loss function across 121 epochs. To balance training efficiency and stability, 0.7

decay factors are applied at epochs 30, 60, and 90 to the learning rate of 5×10^{-4} . Every subject's data is used for experiments, and each sample is divided into a training set and a test set in a 6:4 ratio at random. Both the output channels of the first convolution in the network stream and the number of MobileBlocks in a residual learning module are programmable.

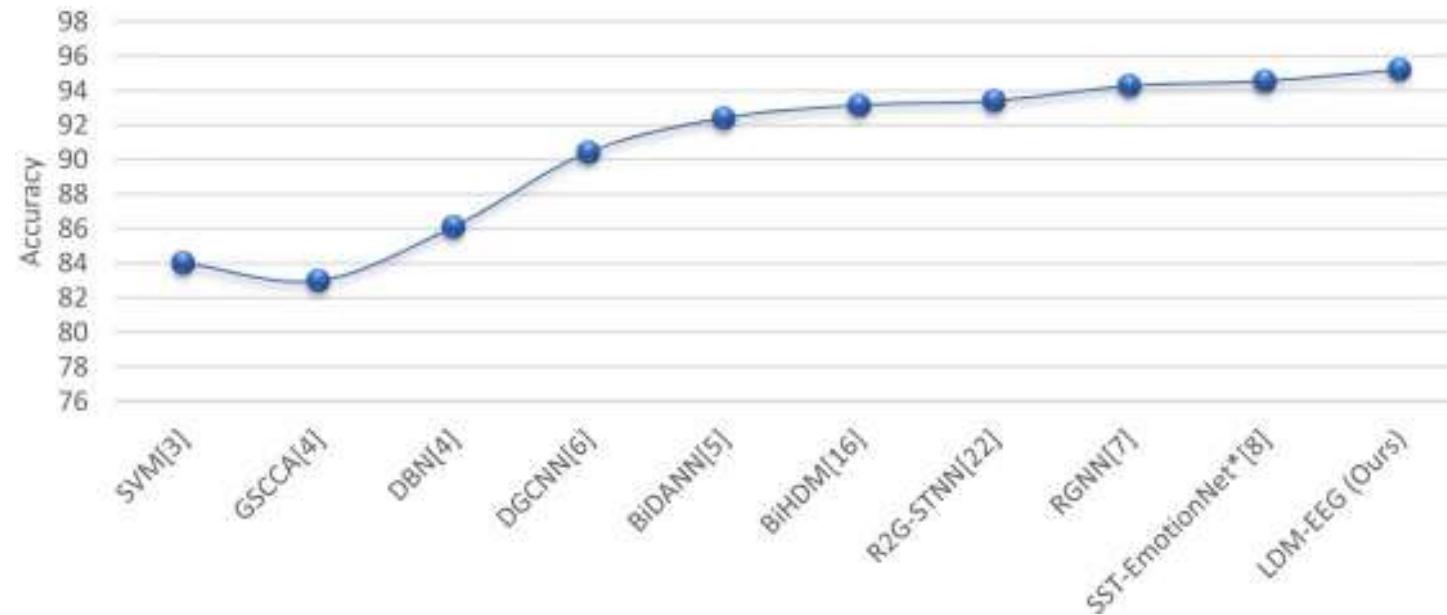
4.3. Result analysis and comparison

The performance of the suggested strategy was compared to the most recent research using the SEED dataset as of 2021. Among the techniques used were the group sparse canonical correlation analysis (GSCCA) and the least squares support vector machine (SVM) (Asif et al., 2023). The deep belief network (DBN) (Gao et al., 2022); the dynamic graph convolutional network (DGCNN) (Feng et al., 2022); the bi-hemisphere difference model (BiHDM) (Xu et al., 2023); the regularized graph neural network (RGNN) (Xu et al., 2023) the spatiotemporal attention-based 3D dense neural network (SST-EmotionNet) (Zhang, 2020). A Hierarchical Spatio-Temporal Neural Network (R2G-STNN) to learn about global and regional spatiotemporal information for EEG emotion recognition. In order to fully extract the information from the three features of time, frequency, and space, the aforementioned features and models frequently only take into account one or two combinations of these domains. In order to create a 3D representation input into a spatiotemporal-frequency dual-stream network, Jia et al.'s work SST-EmotionNet used DE as an EEG feature and made full use of information from the time, frequency, and spatial domains, yielding good emotion classification results. Because our work incorporates several concepts from the top model, SST-EmotionNet, a Pytorch version of it was created for comparison, using consistent optimizer, learning rate, and training epoch parameters. First, as shown in [Table 2](#), the model's classification accuracy was

compared. Since there are two MobileBlocks in each residual learning module, the number of MobileBlocks is set to two in this instance. With this configuration, the model's classification accuracy increased, as shown in Fig. 9, Fig. 10.

Table 2. Performance comparison of the latest models on the SEED data set.

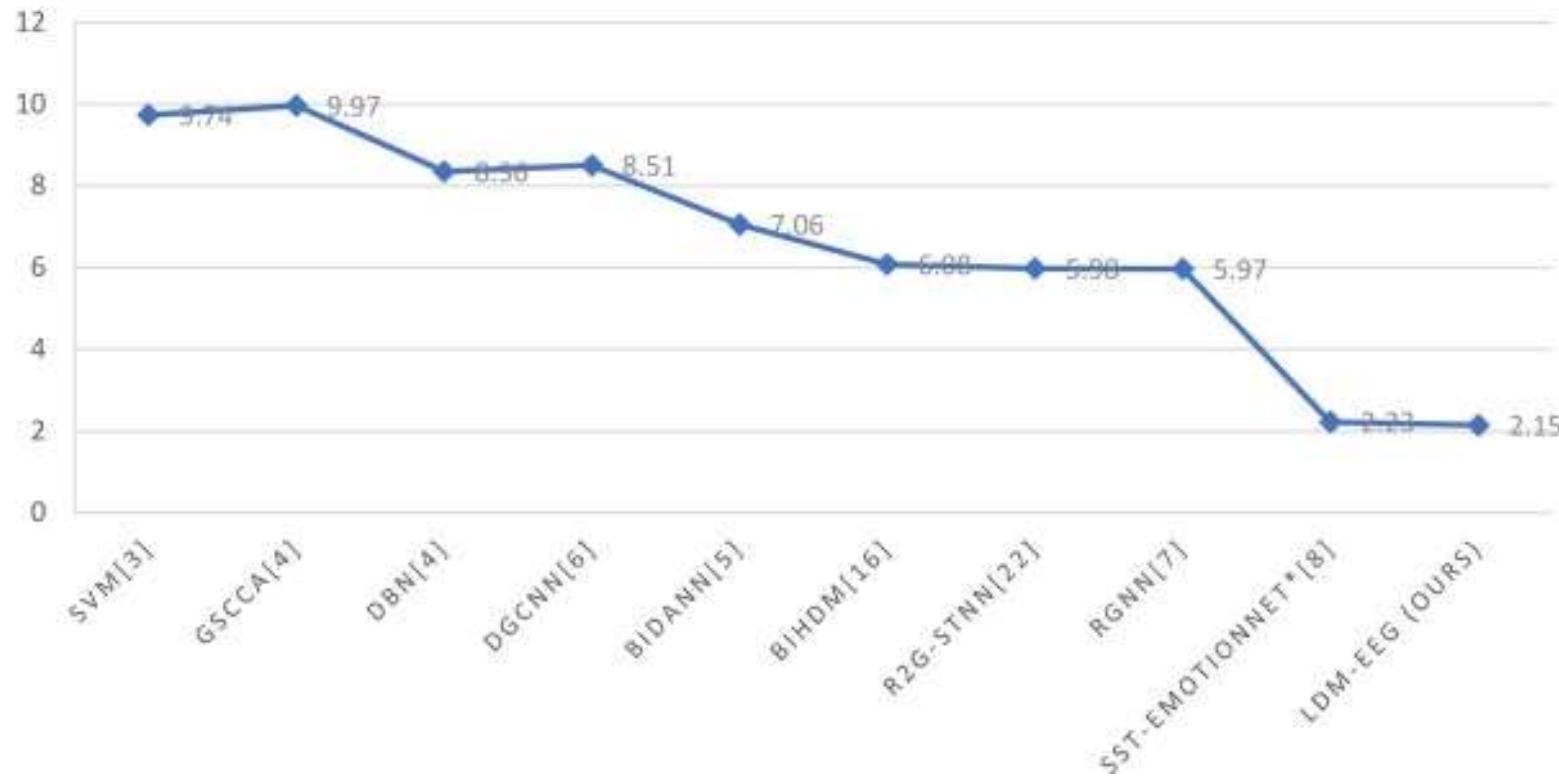
Model	ACC	STD
SVM[53]	84.01	9.74
GSCCA[54]	82.98	9.97
DBN[4]	86.1	8.36
DGCNN[6]	90.42	8.51
BiDANN[55]	92.4	7.06
BiHDM[56]	93.14	6.08
R2G-STNN[22]	93.4	5.98
RGNN[7]	94.26	5.97
SST-EmotionNet*[8]	94.54	2.23
LDM-EEG (Ours)	95.2	2.15



[Download: Download high-res image \(121KB\)](#)

[Download: Download full-size image](#)

Fig. 9. Accuracy of different Model.



Download: [Download high-res image \(121KB\)](#)

Download: [Download full-size image](#)

Fig. 10. STD of different Model.

5. Conclusion

To tackle the issue of EEG emotion identification, this work presents a lightweight architecture that uses multiple attention mechanisms and dual-stream structure scaling. The time-frequency dual-stream

structure, attention mechanism, and depthwise separable 3D convolution are crucial components of this architecture, which was built on top of MobileNet. They accomplish two goals at once: lowering the model's parameter count and increasing the model's classification accuracy. To make sure that important ideas including depthwise separable 3D convolutions, post-attention processes, and multiple attention mechanisms worked, a battery of ablation tests were run. A few caveats, nonetheless, are present in this work. Manual selection is still king, but we scaled the model parameters by depth and breadth to strike a compromise between accuracy and model parameters. Thus, moving forward, we will incorporate concepts like EfficientNet's neural architecture search into our work to improve the automatic balancing of classification accuracy and parameter count. Further complicating matters when comparing parameter counts among models is the paucity of publicly available EEG publications that furnish comprehensive source code and precise parameter setups. In order to overcome this limitation and make continuous improvements, we plan to undertake additional trials in future work.

Funding Statement

There is no funding received by this research.

CRediT authorship contribution statement

Shaleen Bhatnagar: Writing – review & editing, Supervision, Project administration, Methodology, Investigation. **Mohammed I. Khalaf:** Writing – original draft, Data curation, Conceptualization. **Naresh Kumar Gunda:** Writing – original draft, Data curation, Conceptualization. **Vaibhav**

Bhatnagar: Writing – review & editing, Visualization, Formal analysis. **Shtwai Alsubai:** Writing – review & editing, Visualization, Validation, Formal analysis. **Leeladhar Gudala:** Writing – review & editing, Methodology, Formal analysis. **Aadam Quraishi:** Writing – original draft, Methodology, Investigation. **Faisal Yousef Alghayadh:** Writing – original draft, Validation, Methodology. **Ashok Kumar Pamidi Venkata:** Writing – original draft, Validation, Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for supporting this research work through the project number ([PSAU/2024/01/29802](#))

[Special issue articles](#) [Recommended articles](#)

Data availability

I have shared the link of the dataset used in this research in the main manuscript.

References

[Apicella et al., 2022](#) A. Apicella, P. Arpaia, F. Isgrò, G. Mastrati, N. Moccaldi
A survey on EEG-based solutions for emotion recognition with a low number of channels

IEEE Access, vol. 10 (2022), pp. 117411-117428, [10.1109/ACCESS.2022.3219844 ↗](https://doi.org/10.1109/ACCESS.2022.3219844)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Asif et al., 2023](#) M. Asif, S. Mishra, M.T. Vinodbhai, U.S. Tiwary

Emotion recognition using temporally localized emotional events in EEG with naturalistic context: DENS# Dataset

IEEE Access, vol. 11 (2023), pp. 39913-39925, [10.1109/ACCESS.2023.3266804 ↗](https://doi.org/10.1109/ACCESS.2023.3266804)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Chai et al., 2016](#) X. Chai, Q. Wang, Y. Zhao, X. Liu, O. Bai, Y. Li

Unsupervised domain adaptation techniques based on auto-encoder for non-stationary EEG-based emotion recognition

Comput. Biol. Med., 79 (2016), pp. 205-214

 [View PDF](#) [View article](#) [View in Scopus ↗](#) [Google Scholar ↗](#)

[Chai et al., 2017](#) X. Chai, Q. Wang, Y. Zhao, Y. Li, Q. Wang, X. Liu, O. Bai

A Fast, efficient Domain adaptation technique for cross-domain electroencephalography (EEG)-based emotion recognition

Sensors, 17 (2017), p. 1014

[Crossref ↗](#) [View in Scopus ↗](#) [Google Scholar ↗](#)

[Cheng et al., Feb. 2021](#) J. Cheng, et al.

Emotion recognition from multi-channel EEG via deep forest

IEEE J. Biomed. Health Inform., vol. 25 (2) (Feb. 2021), pp. 453-464, [10.1109/JBHI.2020.2995767 ↗](#)

↗

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Du et al., 2023](#) X. Du, et al.

MMPosE: movie-induced multi-label positive emotion classification through EEG signals

1 Oct.-Dec

IEEE Trans. Affect. Comput., vol. 14 (4) (2023), pp. 2925-2938, [10.1109/TAFFC.2022.3221554 ↗](#)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Feng et al., Nov. 2022](#) L. Feng, C. Cheng, M. Zhao, H. Deng, Y. Zhang

EEG-based emotion recognition using spatial-temporal graph convolutional LSTM with attention mechanism

IEEE J. Biomed. Health Inform., vol. 26 (11) (Nov. 2022), pp. 5406-5417,
[10.1109/JBHI.2022.3198688 ↗](#)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Gao et al., 2022](#) Y. Gao, X. Fu, T. Ouyang, Y. Wang

EEG-GCN: spatio-temporal and self-adaptive graph convolutional networks for single and multi-view EEG-based emotion recognition

IEEE Signal Process. Lett., vol. 29 (2022), pp. 1574-1578, [10.1109/LSP.2022.3179946 ↗](#)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Gilakjani and Osman, 2024](#) S.S. Gilakjani, H.A. Osman

A graph neural network for EEG-based emotion recognition with contrastive learning and generative adversarial neural network data augmentation

IEEE Access, vol. 12 (2024), pp. 113-130, [10.1109/ACCESS.2023.3344476 ↗](#)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Gu et al., May 2023](#) Y. Gu, X. Zhong, C. Qu, C. Liu, B. Chen

A domain generative graph network for EEG-based emotion recognition

IEEE J. Biomed. Health Inform., vol. 27 (5) (May 2023), pp. 2377-2386,
[10.1109/JBHI.2023.3242090 ↗](#)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Hou et al., 2023](#) F. Hou, *et al.*

EEG-based emotion recognition for hearing impaired and normal individuals with residual feature pyramids network based on time–frequency–spatial features

Art no. 2505011

IEEE Trans. Instrum. Meas., vol. 72 (2023), pp. 1-11, [10.1109/TIM.2023.3240230 ↗](https://doi.org/10.1109/TIM.2023.3240230)

[Google Scholar ↗](#)

Kim et al., Jan. 2022 S.-H. Kim, H.-J. Yang, N.A.T. Nguyen, S.K. Prabhakar, S.-W. Lee

WeDea: A new EEG-based framework for emotion recognition

IEEE J. Biomed. Health Inform., vol. 26 (1) (Jan. 2022), pp. 264-275, [10.1109/JBHI.2021.3091187 ↗](https://doi.org/10.1109/JBHI.2021.3091187)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Li et al., June 2022 J.W. Li, *et al.*

Single-channel selection for EEG-based emotion recognition using brain rhythm sequencing

IEEE J. Biomed. Health Inform., vol. 26 (6) (June 2022), pp. 2493-2503,
[10.1109/JBHI.2022.3148109 ↗](https://doi.org/10.1109/JBHI.2022.3148109)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Li et al., 2023 W. Li, Y. Tian, J. Dong, C. Fang

A hierarchical three-dimensional mlp-based model for EEG emotion recognition

Art no. 7005004

IEEE Sens. Lett., vol. 7 (10) (2023), pp. 1-4, [10.1109/LSENS.2023.3307111 ↗](https://doi.org/10.1109/LSENS.2023.3307111)

[Google Scholar ↗](#)

[Li et al., 2024](#) W. Li, C. Fang, Z. Zhu, C. Chen, A. Song

Fractal spiking neural network scheme for EEG-based emotion recognition

IEEE J. Transl. Eng. Health Med., vol. 12 (2024), pp. 106-118, [10.1109/JTEHM.2023.3320132 ↗](https://doi.org/10.1109/JTEHM.2023.3320132)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Li et al., July 2022](#) X. Li, F. Shen, Y. Peng, W. Kong, B.-L. Lu

Efficient sample and feature importance mining in semi-supervised EEG emotion recognition

IEEE Trans. Circuits Syst. II: Express Briefs, vol. 69 (7) (July 2022), pp. 3349-3353,
[10.1109/TCSII.2022.3163141 ↗](https://doi.org/10.1109/TCSII.2022.3163141)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Li et al., 2021](#) Y. Li, W. Zheng, Y. Zong, Z. Cui, T. Zhang, X. Zhou

A Bi-hemisphere domain adversarial neural network model for EEG emotion recognition

1 April-June

IEEE Trans. Affect. Comput., vol. 12 (2) (2021), pp. 494-504, [10.1109/TAFFC.2018.2885474 ↗](https://doi.org/10.1109/TAFFC.2018.2885474)

[Google Scholar ↗](#)

[Li et al., 2022](#) Y. Li, W. Zheng, L. Wang, Y. Zong, Z. Cui

From regional to global brain: a novel hierarchical spatial-temporal neural network model for EEG emotion recognition

1 April-June

IEEE Trans. Affect. Comput., vol. 13 (2) (2022), pp. 568-578, [10.1109/TAFFC.2019.2922912 ↗](https://doi.org/10.1109/TAFFC.2019.2922912)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Lin, May 2020](#) Y.-P. Lin

Constructing a personalized cross-day EEG-based emotion-classification model using transfer learning

IEEE J. Biomed. Health Inform., vol. 24 (5) (May 2020), pp. 1255-1264,
[10.1109/JBHI.2019.2934172 ↗](https://doi.org/10.1109/JBHI.2019.2934172)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Liu et al., Nov. 2022](#) S. Liu, *et al.*

3DCANN: a spatio-temporal convolution attention neural network for EEG emotion recognition

IEEE J. Biomed. Health Inform., vol. 26 (11) (Nov. 2022), pp. 5321-5331,
[10.1109/JBHI.2021.3083525 ↗](https://doi.org/10.1109/JBHI.2021.3083525)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Liu et al., 2018](#) Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, Y. Shi

Real-time movie-induced discrete emotion recognition from EEG signals

1 Oct.-Dec

IEEE Trans. Affect. Comput., vol. 9 (4) (2018), pp. 550-562, [10.1109/TAFFC.2017.2660485 ↗](https://doi.org/10.1109/TAFFC.2017.2660485)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Lu et al., 2023](#) W. Lu, T.-P. Tan, H. Ma

Bi-branch vision transformer network for EEG emotion recognition

IEEE Access, vol. 11 (2023), pp. 36233-36243, [10.1109/ACCESS.2023.3266117 ↗](#)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Pan et al., 2023](#) D. Pan, *et al.*

MSFR-GCN: a multi-scale feature reconstruction graph convolutional network for EEG emotion and cognition recognition

IEEE Trans. Neural Syst. Rehabil. Eng., vol. 31 (2023), pp. 3245-3254,
[10.1109/TNSRE.2023.3304660 ↗](#)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Peng et al., 2021](#) Y. Peng, *et al.*

Self-weighted semi-supervised classification for joint EEG-based emotion recognition and affective activation patterns mining

Art no. 2517111

IEEE Trans. Instrum. Meas., vol. 70 (2021), pp. 1-11, [10.1109/TIM.2021.3124056 ↗](#)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Peng et al., 2022](#) Y. Peng, F. Jin, W. Kong, F. Nie, B.-L. Lu, A. Cichocki

OGSSL: a semi-supervised classification model coupled with optimal graph learning for EEG emotion recognition

IEEE Trans. Neural Syst. Rehabil. Eng., vol. 30 (2022), pp. 1288-1297,
[10.1109/TNSRE.2022.3175464 ↗](https://doi.org/10.1109/TNSRE.2022.3175464)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Peng et al., 2023 Y. Peng, H. Liu, J. Li, J. Huang, B.-L. Lu, W. Kong

Cross-session emotion recognition by joint label-common and label-specific EEG features exploration

IEEE Trans. Neural Syst. Rehabil. Eng., vol. 31 (2023), pp. 759-768,
[10.1109/TNSRE.2022.3233109 ↗](https://doi.org/10.1109/TNSRE.2022.3233109)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Peng et al., Sept. 2022 Y. Peng, F. Qin, W. Kong, Y. Ge, F. Nie, A. Cichocki

GFIL: a unified framework for the importance analysis of features, frequency bands, and channels in EEG-based emotion recognition

IEEE Trans. Cogn. Dev. Syst., vol. 14 (3) (Sept. 2022), pp. 935-947, [10.1109/TCDS.2021.3082803 ↗](https://doi.org/10.1109/TCDS.2021.3082803)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Pereira et al., 2021 E.T. Pereira, H.M. Gomes, L.R. Veloso, M.R.A. Mota

Empirical evidence relating EEG signal duration to emotion classification performance
1 Jan.-March

IEEE Trans. Affect. Comput., vol. 12 (1) (2021), pp. 154-164, [10.1109/TAFFC.2018.2854168 ↗](https://doi.org/10.1109/TAFFC.2018.2854168)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Sarma and Barma, 2022 P. Sarma, S. Barma

Emotion recognition by discriminating EEG segments with high affective content from automatically selected relevant channels

Art no. 4000812

IEEE Trans. Instrum. Meas., vol. 71 (2022), pp. 1-12, [10.1109/TIM.2022.3147876 ↗](https://doi.org/10.1109/TIM.2022.3147876)

[Google Scholar ↗](#)

Shen et al., 2023 X. Shen, X. Liu, X. Hu, D. Zhang, S. Song

Contrastive learning of subject-invariant EEG representations for cross-subject emotion recognition

1 July-Sept

IEEE Trans. Affect. Comput., vol. 14 (3) (2023), pp. 2496-2511, [10.1109/TAFFC.2022.3164516 ↗](https://doi.org/10.1109/TAFFC.2022.3164516)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Song et al., 2019 T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, Z. Cui

MPED: a multi-modal physiological emotion database for discrete emotion recognition

IEEE Access, vol. 7 (2019), pp. 12177-12191, [10.1109/ACCESS.2019.2891579 ↗](https://doi.org/10.1109/ACCESS.2019.2891579)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Song et al., 2020 T. Song, W. Zheng, P. Song, Z. Cui

EEG emotion recognition using dynamical graph convolutional neural networks

1 July-Sept

IEEE Trans. Affect. Comput., vol. 11 (3) (2020), pp. 532-541, [10.1109/TAFFC.2018.2817622 ↗](https://doi.org/10.1109/TAFFC.2018.2817622)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Tian et al., 2021 Z. Tian, D. Li, Y. Song, Q. Gao, Q. Kang, Y. Yang

EEG-based emotion recognition of deaf subjects by integrated genetic firefly algorithm

Art no. 2516911

IEEE Trans. Instrum. Meas., vol. 70 (2021), pp. 1-11, [10.1109/TIM.2021.3121473 ↗](https://doi.org/10.1109/TIM.2021.3121473)

[Google Scholar ↗](#)

Wang et al., 2024 X. Wang, J. Zhang, C. He, H. Wu, L. Cheng

A novel emotion recognition method based on the feature fusion of single-lead EEG and ECG signals

1 March1

IEEE Internet Things J., vol. 11 (5) (2024), pp. 8746-8756, [10.1109/JIOT.2023.3320269 ↗](https://doi.org/10.1109/JIOT.2023.3320269)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Wang et al., 2023 Y. Wang, *et al.*

Wearable wireless dual-channel EEG system for emotion recognition based on machine learning

15 Sept.15

IEEE Sens. J., vol. 23 (18) (2023), pp. 21767-21775, [10.1109/JSEN.2023.3303441 ↗](https://doi.org/10.1109/JSEN.2023.3303441)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Wang et al., 2022](#) Z. Wang, Y. Wang, C. Hu, Z. Yin, Y. Song

Transformers for EEG-based emotion recognition: a hierarchical spatial information learning model

1 March

IEEE Sens. J., vol. 22 (5) (2022), pp. 4359-4368, [10.1109/JSEN.2022.3144317 ↗](#)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Wu et al., 2023](#) M. Wu, W. Teng, C. Fan, S. Pei, P. Li, Z. Lv

An investigation of Olfactory-enhanced video on EEG-based emotion recognition

IEEE Trans. Neural Syst. Rehabil. Eng., vol. 31 (2023), pp. 1602-1613,

[10.1109/TNSRE.2023.3253866 ↗](#)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Xu et al., 2023](#) X. Xu, T. Jia, Q. Li, F. Wei, L. Ye, X. Wu

EEG feature selection via global redundancy minimization for emotion recognition

1 Jan.-March

IEEE Trans. Affect. Comput., vol. 14 (1) (2023), pp. 421-435, [10.1109/TAFFC.2021.3068496 ↗](#)

[View in Scopus ↗](#) [Google Scholar ↗](#)

[Zhang, 2020](#) H. Zhang

Expression-EEG based collaborative multimodal emotion recognition using deep autoencoder

IEEE Access, vol. 8 (2020), pp. 164130-164143, [10.1109/ACCESS.2020.3021994 ↗](https://doi.org/10.1109/ACCESS.2020.3021994)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Zhang et al., 2023 Y. Zhang, H. Liu, D. Zhang, X. Chen, T. Qin, Q. Zheng

EEG-based emotion recognition with emotion localization via hierarchical self-attention

1 July-Sept

IEEE Trans. Affect. Comput., vol. 14 (3) (2023), pp. 2458-2469, [10.1109/TAFFC.2022.3145623 ↗](https://doi.org/10.1109/TAFFC.2022.3145623)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Zheng et al., 2017 Zheng, H.L.; Fu, J.L.; Mei, T.; Luo, J.B. Learning Multi-Attention Convolutional Neural Network for Fine-Grained Image Recognition. In Proceedings of the 16th IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.

[Google Scholar ↗](#)

Zheng, Sept. 2017 W. Zheng

Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis

IEEE Trans. Cogn. Dev. Syst., vol. 9 (3) (Sept. 2017), pp. 281-290, [10.1109/TCDS.2016.2587290 ↗](https://doi.org/10.1109/TCDS.2016.2587290)

[View in Scopus ↗](#) [Google Scholar ↗](#)

Zheng and Lu, 2016 Zheng, W.-L.; Lu, B.-L. Personalizing EEG-based affective models with transfer learning. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016; pp. 2732–2738.

[Google Scholar ↗](#)

Zheng et al., 2015 Zheng, W.-L.; Zhang, Y.-Q.; Zhu, J.-Y.; Lu, B.-L. Transfer components between subjects for EEG-based emotion recognition. In Proceedings of the 2015 International Conference on Active Computing and Intelligent Interaction (ACII), Xi'an, China, 21–24 September 2015; pp. 917–922.

[Google Scholar ↗](#)

Cited by (1)

Pushing the boundaries of brain-computer interfacing (BCI) and neuron-electronics
2024, Journal of Neuroscience Methods

© 2024 The Author(s). Published by Elsevier B.V.



All content on this site: Copyright © 2024 Elsevier B.V., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the Creative Commons licensing terms apply.

