

Article

Study on Driver Cross-Subject Emotion Recognition Based on Raw Multi-Channels EEG Data

Zhirong Wang, Ming Chen  and Guofu Feng *

School of Information Science, Shanghai Ocean University, Shanghai 201306, China

* Correspondence: gffeng@shou.edu.cn

Abstract: In our life, emotions often have a profound impact on human behavior, especially for drivers, as negative emotions can increase the risk of traffic accidents. As such, it is imperative to accurately discern the emotional states of drivers in order to preemptively address and mitigate any negative emotions that may otherwise manifest and compromise driving behavior. In contrast to many current studies that rely on complex and deep neural network models to achieve high accuracy, this research aims to explore the potential of achieving high recognition accuracy using shallow neural networks through restructuring the structure and dimensions of the data. In this study, we propose an end-to-end convolutional neural network (CNN) model called simply ameliorated CNN (SACNN) to address the issue of low accuracy in cross-subject emotion recognition. We extracted features and converted dimensions of EEG signals from the SEED dataset from the BCMI Laboratory to construct 62-dimensional data, and obtained the optimal model configuration through ablation experiments. To further improve recognition accuracy, we selected the top 10 channels with the highest accuracy by separately training the EEG data of each of the 62 channels. The results showed that the SACNN model achieved an accuracy of 88.16% based on raw cross-subject data, and an accuracy of 91.85% based on EEG channel data from the top 10 channels. In addition, we explored the impact of the position of the BN and dropout layers on the model through experiments, and found that a targeted shallow CNN model performed better than deeper and larger perceptual field CNN models. Furthermore, we discuss herein the future issues and challenges of driver emotion recognition in promising smart city applications.



Citation: Wang, Z.; Chen, M.; Feng, G. Study on Driver Cross-Subject Emotion Recognition Based on Raw Multi-Channels EEG Data. *Electronics* **2023**, *12*, 2359. <https://doi.org/10.3390/electronics12112359>

Academic Editor: Jichai Jeong

Received: 23 April 2023

Revised: 11 May 2023

Accepted: 18 May 2023

Published: 23 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: emotion recognition; multi-channels EEG; cross-subject; CNN

1. Introduction

As we continue to experience the era of urbanization and information technology, the concept of smart cities has become a key stage in urban development. The main goal of smart cities is to achieve human-centered development through the integration of information technology. In this context, intelligent transportation has become an important area of research, playing a key role in ensuring the efficient and sustainable movement of people and goods within and between cities. Therefore, the development of intelligent transportation systems has become an indispensable factor in achieving the ultimate goal of smart cities [1]. With the continuous improvement of transportation facilities and vehicles, the threshold for drivers is becoming lower. However, the increased number of cars has led to a rise in traffic accidents, with human factors playing a dominant role in road traffic safety [2–5]. Negative emotions such as grief, anger, and rage can significantly affect a driver's cognition, judgment, and driving behavior, leading to misjudgments and serious traffic accidents. It is therefore crucial to develop effective and reliable methods for identifying and regulating negative emotions in drivers to improve traffic safety. Advanced driver assistance systems (ADAS) [6] have been widely adopted in the market to identify and warn of dangerous driving behaviors using advanced sensor and information interconnection technology [7]. However, most of the popular driving assistance systems on the

market ignore the driver's emotional state during the driving process [8], focusing mainly on fatigue driving and distracted driving. As relying on the driver to self-regulate their emotional instability is not practical, developing a negative emotion recognition system is essential to effectively address the impact of emotional states on driving behavior and improve traffic safety.

In addition to eliciting activation signals in specific functional areas of the brain, changes in human emotions also result in various physiological reactions through the nervous system, leading to alterations in peripheral physiological signals (such as EEG, skin conductance, ECG, respiration, temperature, etc.) [8] and external expressions (such as speech, facial expressions, postures, actions, etc.), which may subsequently trigger certain behaviors. By investigating the relationship between brain functional activity and the aforementioned physiological signals and external behaviors, and analyzing subsequent behavioral performance in situational perception environments, it is possible to automatically acquire emotional information and compute and recognize emotional states in practical applications. This can also enable the prediction and regulation of subsequent behavior, thereby endowing machines with a degree of emotional intelligence. Electroencephalography (EEG) [9] signals are commonly employed in emotional recognition. The main components of EEG signals are brain rhythmic signals from various brain regions, which reflect the activity of corresponding areas [10]. The electrical activity of the cerebral cortex is transmitted to the scalp via the skull-brain structure, and thus, the obtained EEG signals are mixed signals of brain signals from different brain regions and eye movement signals, which contain a considerable amount of redundant information and have a low signal-to-noise ratio [11]. Consequently, the extraction of time-related features from EEG signals has become a pivotal factor in emotional recognition from EEG signals [12].

The recent advancements in deep learning techniques have led to a surge of interest in applying such techniques to the field of emotion-based brain-computer interfaces. This has resulted in the emergence of an increasing number of deep learning methods being utilized in this area, as evidenced by several studies [13–15]. These studies have demonstrated a strong correlation between changes in EEG signals and other physiological signals with changes in human emotions [13,16,17]. Deep learning, with its ability to effectively learn deep feature representations of samples, has been shown to be a valuable tool for extracting emotional state information contained in physiological signals. For instance, Kranti S. Kamble et al. [18] proposed a machine learning approach that achieved an AUC of 95.81% for single-subject emotions in the SEED dataset from the BCMI Laboratory. Similarly, Smith K. Khare et al. [19] proposed an eigenvector centrality method (EVCN) that exhibited the highest accuracy, at 97.24%. These findings underscore the potential of deep learning techniques for emotion recognition, and further research in this domain is warranted.

With the rapid advancement of deep learning techniques, various deep learning models have been successfully applied in the field of emotion-based brain-computer interfaces (BCIs) [20–25]. Studies have shown that changes in EEG signals and other physiological signals are strongly associated with changes in human emotions, and deep learning methods are highly effective in extracting emotional state information contained in physiological signals. For instance, Keelawat [20] trained a multilayer convolutional neural network (CNN) model with three to seven layers on the DEAP dataset, achieving accuracies of 81.54% and 86.87% for arousal and valence, respectively. This clearly indicates that CNNs are capable of extracting relevant emotional features from EEG signals and accurately classifying them based on these features.

Yang et al. [21] and Wang et al. [22] have also conducted related experiments on the SEED dataset from the BCMI Laboratory. However, due to differences in experimental scenarios, it is difficult to draw a straightforward conclusion as to which neural network architecture is more suitable for EEG-based emotion recognition. It is worth noting that most spatial and temporal neural network models employ convolutional and recurrent neural networks, such as C-RNN [23], graph convolution [24], and attention mechanism models [25], among others.

In the context of EEG signal recognition, such as P300 and SSVEP paradigms, the features are not strongly correlated with time, but only sensitive to specific brain potentials induced at a particular instant. Therefore, it is speculated that although there is an induction time, the overall induction state of the EEG should be instantaneous, with weak temporal correlations. This observation implies that an improved CNN model is sufficient to extract feature information, and recurrent neural networks and their derivative models that require more computation and training time than CNNs may not be as effective as CNNs.

Furthermore, in the realm of cross-modality, transfer learning methods have gained increasing attention from researchers. For instance, Lin YP et al. [26] proposed a novel conditional transfer learning (cTL) framework, which facilitated positive transfer by determining the similarity between the source and target domains. Their approach achieved promising results in improving the overall classification performance of 26 individuals by approximately 15% for valence classification and about 12% for arousal classification. Similarly, Zheng W L et al. [27] utilized transfer learning techniques to construct an emotion model based on personalized EEG data in the absence of labeled target data. Specifically, they explored two types of transfer methods: one that shared the structure of the source and target domains and another that trained multiple individual classifiers on the source subject, then transferred relevant classifier parameters to the target subject. They demonstrated the effectiveness of their approach by constructing emotional models for positive, neutral, and negative emotions, achieving an average accuracy of 76.31%, compared to 56.73% for traditional general classifiers. In a recent study, Luo et al. [28] proposed a novel Wasserstein generative adversarial network domain adaptation (WGANDA) framework for building emotion recognition models based on interdisciplinary electroencephalogram (EEG) data. Their proposed framework consists of GAN-like components and a two-step training procedure with pre-training and adversarial training. Evaluated on two widely used public datasets, SEED and DEAP, the framework achieved an average accuracy of 87.07% for cross-subject emotion recognition on SEED datasets from the BCMI Laboratory and a highest average accuracy of 67.99% on DEAP, outperforming traditional general classifiers. Thus, a further exploration of constructing end-to-end models for effectively extracting emotional features using deep learning techniques remains a highly compelling area of research for many scholars in the field.

This paper presents an innovative end-to-end convolutional model, SACNN, for cross-subject EEG emotion recognition. The proposed model uses large convolution kernels and wide pooling attention, enabling faster convergence and improved generalization capacity. Furthermore, this study conducts ablation experiments to identify the optimal positions of batch normalization (BN) and dropout layers. The result shows that a targeted shallow convolutional neural network outperforms deeper convolutional neural network models. These findings offer valuable insights into optimizing the architecture of CNN models for EEG emotion recognition, which can potentially facilitate the development of more effective emotion recognition systems.

The main contributions of this paper include:

1. A novel end-to-end cross-subject EEG model for emotion recognition that automatically extracts temporal features from raw EEG data through large kernel sizes and attention pooling, and outperforms other EEG-based models using the proposed features.
2. The paper explores the impact of selecting 10 channels from the frontal and temporal lobes on the training results of 62-channel EEG data.
3. The paper conducts extensive ablation experiments, mainly focusing on the impact positions of the BN and dropout layers in CNN and the influence of multilayer convolution on model training results. These experiments demonstrate the reliability of the proposed model.

2. Methodology

In this section, we provide a detailed description of the proposed method as illustrated in Figure 1. First, we describe the EEG data preprocessing approach. Subsequently,

we introduce our channel selection methodology. Finally, we provide a comprehensive overview of the SACNN architecture, including its specific algorithm for multi-channel EEG classification.

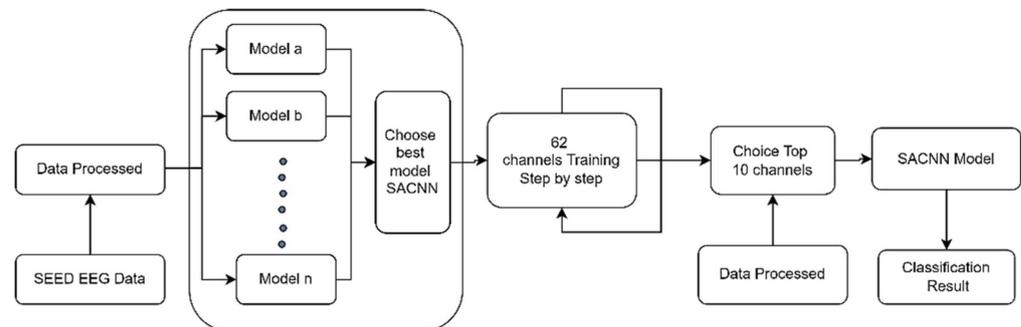


Figure 1. Flowchart of the proposed method.

2.1. Preprocessing

The dimensions and meanings of EEG data are crucial for the final results and interpretability of experiments. Different dimensions and input forms represent different specific meanings of EEG data. In this study, each experiment in the SEED dataset was from the BCMI Laboratory. Data were collected when fifteen Chinese subjects (7 males and 8 female) were watching film clips. The film clips were carefully selected to induce different types of emotion, which were positive, negative, and neutral. There was a total of 15 trials for each experiment. There was a 5 s hint before each clip, 45 s for self-assessment and 15 s to rest after each clip in one session. The order of presentation was arranged in such a way that two film clips that targetted the same emotion are not shown consecutively. For feedback, the participants were told to report their emotional reactions to each film clip by completing the questionnaire immediately after watching each clip [17]. Data contained 15 sessions (emotional experiments) with each session having a data length range of 37,001–47,601. Therefore, we have proposed a scheme to construct 62-dimensional data by intercepting all data into the minimum length of 37,000. Due to the extreme differences in data volume between individual features in the SEED dataset, it is not feasible to construct a data matrix using the direct truncation or zero-padding preprocessing method employed in the previous chapter. Therefore, this chapter proposes a dynamic downsampling approach for the SEED dataset, which exhibits significant differences in data volume. The specific downsampling algorithm is shown in the table below, which reduces the dimensionality of the data while preserving the data characteristics, normalizes the overall data, and constructs a data matrix for easy integration into the model for computation. The feature session data of each dimension are stacked together to form the data dimension of a single subject. The data are then transposed, allowing the model to train on each session in the time dimension rather than the spatial dimension. Finally, all the data are concatenated to construct a cross-subject structured emotional data belonging to the current modality. The specific downsampling process is shown in the Algorithm 1.

Based on the constructed data structure of time-feature dimensions, we recursively traverse down to the minimum feature dataset and perform downsampling with the core calculation shown in Equation (1):

$$newSampleData = \sum_k^{k+xL/minLen+[xL/minLen+flag+[xL/minLen]]} X \quad (1)$$

Here, X represents the current data, xL represents the length of the current time data, $minLen$ represents the minimum data length of all data, k represents the current loop data, and $newSampleData$ represents the new feature value under the downsampled new dimension. Subsequently, multiple $newSampleData$ form a column of feature sequences, representing the feature sequences of a certain experimental session and a subject's data collection dimension.

Algorithm 1 Preprocessing downsampling algorithm

```

input: SEED Dataset,
input: Data Data array constructed by reading in layer by layer,
input: n Data size
output: The downsampled cross-subject matrix data
1: function DownSample(Data)
2:   result  $\leftarrow$  []
3:   minLen  $\leftarrow$  The minimum data length for a single time series under a single feature
4:   resIndex  $\leftarrow$  0
5:   for <all Data> do
6:     diff = Data length of the current time series / minLen
7:     if diff == 1 then
8:       eye  $\leftarrow$  Data of the current time series
9:       result[resIndex ++]  $\leftarrow$  eye.T
10:      continue
11:    end if
12:    eye  $\leftarrow$  []
13:    eyeIndex  $\leftarrow$  0
14:    for <Time series data under current features> do
15:      temp  $\leftarrow$  0
16:      flag  $\leftarrow$  0
17:      tempTimeList  $\leftarrow$  []
18:      tempTimeListIndex  $\leftarrow$  0
19:      while temp < Length of time series under current features do
20:        if diff + flag - int(diff) >= 1 then
21:          x = diff + int(diff + flag - int(diff))
22:          flag  $\leftarrow$  flag - 1
23:        else
24:          x  $\leftarrow$  diff
25:        end if
26:        tempTimeList[tempTimeListIndex ++]  $\leftarrow$  sum(The time series k to k+int(x) under the
current feature)
27:        k  $\leftarrow$  k + int(x)
28:        if k + 1 + int(diff) >= Length of time series under current features then
29:          flag  $\leftarrow$  1
30:        end if
31:        flag  $\leftarrow$  flag + (diff - int(diff))
32:      end while
33:      eye[eyeIndex ++]  $\leftarrow$  tempTimeList
34:    end for
35:    result[resIndex ++]  $\leftarrow$  eye.T
36:  end for
37:  return result
38: end function

```

The raw EEG data structure is difficult to apply to cross-subject training. However, the features extracted through transfer learning and end-to-end learning from a single subject's single trial data are ultimately different. Therefore, the raw data were reconstructed to accommodate all subject information across all experiments, enabling the model to be trained end-to-end and enhancing its generalization performance and practical applicability. The cross-subject data reconstruction method is as follows.

First, stack the 62 dimensions of a single subject's trial session data to represent a subject's single trial session. Next, transpose the data (channel, session) to (session, channel) so that the model trains each session on the time dimension rather than the spatial dimension. Then, concatenate all experimental subject data to obtain 675 cross-subject experimental trial data in the form of Figure 2, resulting in a reconstructed SEED matrix data structure dimension of (675; 37,000; 62). In Figure 2, the different colored lines

represent different EEG channels. Due to space limitations, only six representative channels are depicted here for illustration purposes. Finally, the reconstructed cross-subject matrix data can be trained in the model as a whole, replacing segmented training one by one. The model integrates more cross-subject and cross-experiment information, enhancing its generalization performance.

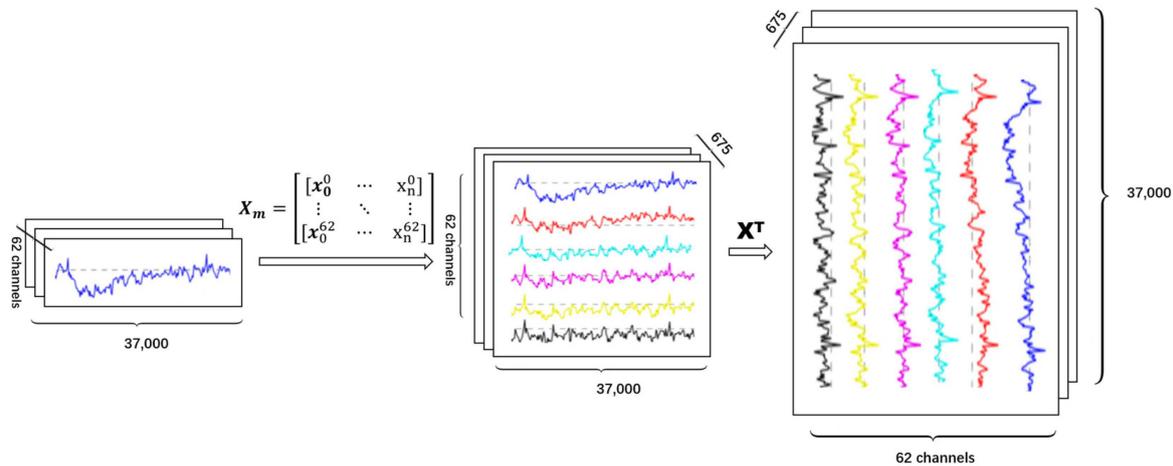


Figure 2. EEG data proposing method.

To further refine the emotional features present in the EEG data, we applied filtering to the transposed and reconstructed data. Human brainwave frequencies typically fall within the range of 0–35 Hz; some high-frequency feature data may influence the results. Therefore, we reduced the data frequency by 15 Hz and applied a low-pass filter to filter the data from 0–50 Hz.

2.2. Channel Choice

To improve computational efficiency, it is not necessary to use all available channels of EEG data for the model. Therefore, in this study, the SACNN was used to train each of the 62 channels of data separately, and the results were validated. To ensure the reliability of the experimental results and reduce errors, the experiments were repeated five times with different initial random weights for each network model. The channel importance was determined using the following formula:

$$R = \text{sort}\left(\frac{1}{N} \sum_{i=0}^n Acc_i\right) \tag{2}$$

Here, R represents the sorted channel importance result, N and n represent the number of repeated trials (5 in this case), Acc represents the mean value after 10-fold cross-validation, and the sort function sorts the values in descending order. By using this method, the most important channels for the classification task can be identified and used for the final model, thus improving the efficiency of the computation.

2.3. SACNN Model

To describe the proposed framework for EEG emotion recognition based on CNN, there are seven key components, namely ConvRelu-10, BN, AveragePooling, ConvRelu-5, MaxPooling, dropout, and fully connected neural network (FCN). Figure 3 provides a detailed illustration of these components.

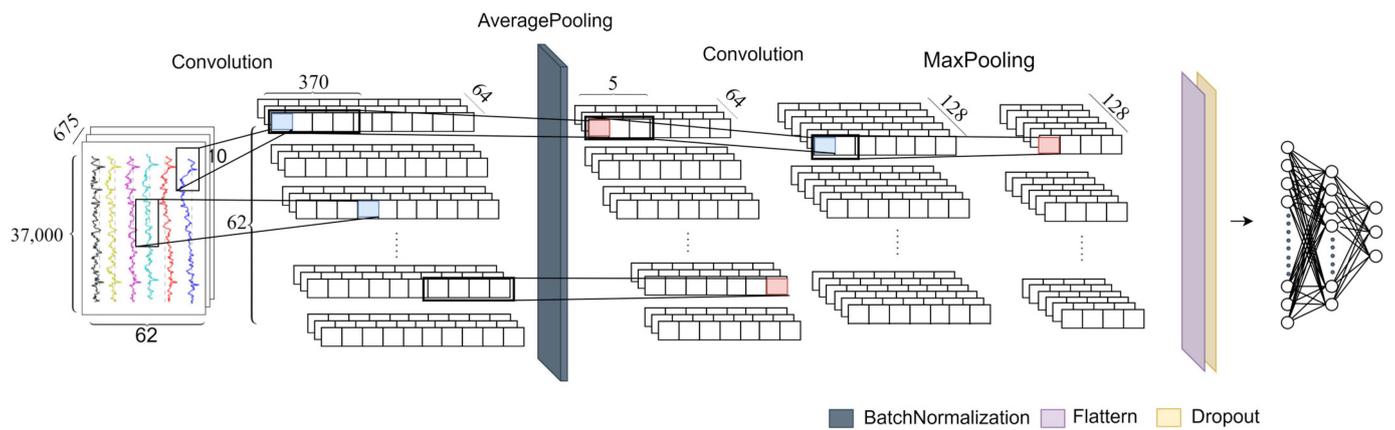


Figure 3. The framework of SACNN.

3. Experiments and Result

3.1. Dataset

The SEED dataset contains EEG recordings of three emotional categories, namely positive, negative, and neutral. The experiment involved 15 subjects who performed three experiments, with an interval of approximately one week between each session. The subjects’ emotions were stimulated by watching a film, and each experiment comprised 15 trials. After each trial, subjects were asked to provide feedback. The EEG signals were recorded using 62 channels at a sampling rate of 1000 Hz, which were then downsampled to 200 Hz. The EEG data underwent noise removal and artifact removal, followed by bandpass filtering from 0–75 Hz.

The negative emotions in the SEED dataset can be extrapolated to negative emotions experienced by drivers during driving.

3.2. Experimental Setup

We train our model on NVIDIA RTX 3090 GPU and 64 GB of memory. For each model, we set three batch sizes and train 100 epochs. The learning rate is set to 0.001. For each experiment, we randomly shuffle the samples. The ratio of the training set to test set is 8:2.

3.3. Ablation Studies

To demonstrate the effectiveness of BN, dropout, and multi-Conv blocks at different layers of the model, ablation experiments were conducted on the SEED dataset. Pre-experiments showed no significant difference between placing the BN layer in the front, middle or rear of the conv and pooling layers, so it was placed in the middle in this study. Table 1 outlines the specific structures of the multi-Conv blocks model, which consists of nine models with different combinations of convolutional layers and BN layers.

Table 1. The structure of multi-Conv blocks model.

Model a	Model b	Model c	Model d	Model e	Model f	Model g	Model h	Model i
CNN-1 (MP + BN + DP)	CNN-2 (AP + BN + DP)	CNN-2 (AP_MP + BN + DP)	CNN-2 (MP_AP + BN + DP)	CNN-5 (MP + 1_BN + DP)	CNN-5 (MP + Multi_BN + DP)	CNN-10 (MP + 1_BN + DP)	CNN-10 (MP + Multi_BN + DP)	CNN-10 (MP_AP + BN + DP + Res)
Input (62 Channels Raw EEG Data)								
Conv1D 10-64 BN	Conv1D 10-64 BN	Conv1D 10-64 BN	Conv1D 10-64 BN	Conv1D 10-64 BN	Conv1D 10-64 BN	Conv1D 10-64 BN	Conv1D 10-64 BN	Conv1D 10-64 BN
Maxpooling1D 370	Averagepooling1D 370				Maxpooling1D 370			
DP 0.2	Conv1D 5-128	Conv1D 5-128	Conv1D 5-128	Conv1D 5-128	Conv1D 5-128 BN	Conv1D 5-128	Conv1D 5-128 BN	Conv1D 5-128 BN
	Averagepooling1D 2	Maxpooling1D 2	Averagepooling1D 2			Maxpooling1D 2		
		DP 0.2		Conv1D 5-256	Conv1D 5-128 BN	Conv1D 5-256 Conv1D 5-256	Conv1D 5-256 Conv1D 5-256 BN	Conv1D 5-256 Conv1D 5-256

Table 3 clearly demonstrates that adding a BN layer as seen in model c8 greatly improves the accuracy of the model. When comparing models c1–c3, the benefits of adding BN to the Conv block are greater than the FCN region. Similarly, when comparing models c1–c7, the benefits of a multilayer BN are no better than a single-layer BN. This indicates that for EEG emotion recognition, an excessive normalization of the data is not beneficial to model learning.

(C) Dropout Layers Ablation Studies

Table 4 presents the dropout layers ablation experiment results on each model. Different blocks mean the position of dropout layers set. In this table, the checkmarks indicate the addition of the corresponding DP (Dropout) layer in that particular layer. The absence of a checkmark indicates that no Dropout layer was added in that layer. Here, we choose nine models, meaning that 20% of dropout data is featured respectively in the first layer, second layer, third layer, both in first and second layers, both in first and third layers, both in second and third layers, in all layers, no dropout layer in model c9, and 50% dropout data features in third layer in model c8.

Table 4. Dropout layers ablation studies results.

	Model c1	Model c2	Model c3	Model c4	Model c5	Model c6	Model c7	Model c8	Model c9
Block1-DP 0.2	✓			✓	✓		✓		
Block2-DP 0.2		✓		✓		✓	✓		
Block3-DP 0.2			✓		✓	✓	✓		
Block3-DP 0.5								✓	
Loss	1.12 ± 0.0083	1.164 ± 0.02	1.081 ± 0.023	1.11 ± 0.017	1.1 ± 0.01	1.107 ± 0.006	1.114 ± 0.02	1.10 ± 0.02	1.107 ± 0.006
Accuracy	0.326 ± 0.005	0.33 ± 0.016	0.437 ± 0.015	0.34 ± 0.016	0.33 ± 0.005	0.333 ± 0.001	0.348 ± 0.019	0.36 ± 0.015	0.33 ± 0.007

Table 4 shows that adding a dropout layer to model c9 and others can improve the model’s accuracy by 10%. Comparing models c1–c3, adding dropout to the FCN block is more effective than adding it to the Conv block. However, comparing models c1–c7, the gain of multilayer dropout is not as good as that of single-layer dropout. This suggests that dropping out some data are beneficial for improving the model’s generalization, but dropping out too much data hinders the model’s ability to learn data features. In addition, dropping 0.2 data is more stable than dropping 0.5 data in terms of loss, indicating that dropping too much data is not good for feature discrimination.

(D) Multi-Channels Results

After single-channel training, Figure 4a indicates that the blue region exhibits a higher accuracy, primarily in the prefrontal and bilateral temporal regions.

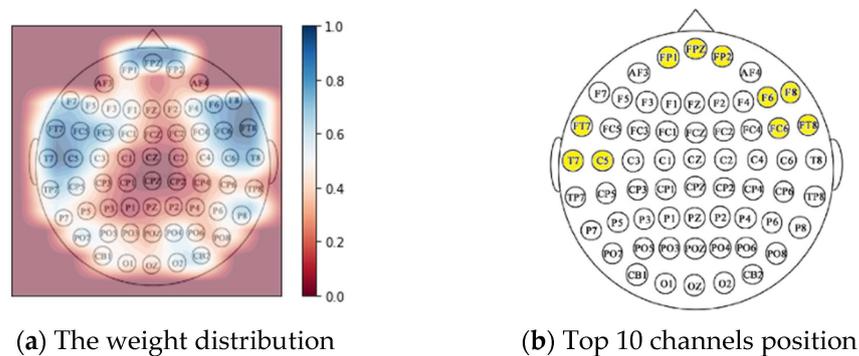


Figure 4. 62-channels training result.

Table 5 shows the results for the top 10 channels (F8, T7, FT7, FC6, FPZ, FT8, FP2, F6, C5, and FP1), with accuracies ranging from 59.9% to 72.4% and a stable mean standard deviation of 0.02. Figure 4b illustrates that the top 10 potentials are primarily concentrated in the front half of the brain, with FT7, T7, C5, and F6 showing oblique symmetry rather than exact left–right symmetry, and F8, FC6, and FT8 also exhibiting oblique symmetry.

Table 5. Top 10 channels result.

	Channel	Mean-acc	Channel	Mean-acc	Channel	Mean-acc		
1	F8	0.724 ± 0.015	5	FPZ	0.644 ± 0.022	8	F6	0.620 ± 0.022
2	T7	0.691 ± 0.012	6	FT8	0.641 ± 0.016	9	C5	0.610 ± 0.017
3	FT7	0.689 ± 0.021	7	FP2	0.624 ± 0.025	10	FP1	0.599 ± 0.023
4	FC6	0.652 ± 0.012						

(E) Compared Models Results

Here, we choose four models, SVM, DBN, LSTM, and CNN-LSTM, as the comparison models.

Table 6 displays the accuracy (acc/std) of the five models using single-channel, multi-channels, and all channels. The single channel represents the highest accuracy, F8, in the top 10 chosen as the input to the model, while multi means that the 10 channels in Table 5 are selected as the model input. It can be observed from Table 6 that training the models with raw EEG data using multi-channels improved the accuracy of all models to varying degrees, indicating that filtering effective channel data positively impacts model learning. The SACNN achieves an accuracy of 88.16% with all channels trained, and 91.85% with multi-channels, with a maximum accuracy of 94.81%. These results demonstrate that a simple CNN network can achieve high accuracy after model improvement and valid data channel selection.

Table 6. The comparison models results on SEED dataset for different channels.

Model	Single	Multi	All
SVM	0.748 ± 0.001	0.763 ± 0.001	0.333 ± 0.001
DBN	0.726 ± 0.013	0.826 ± 0.017	0.392 ± 0.027
LSTM	0.541 ± 0.021	0.444 ± 0.031	0.348 ± 0.058
CNN-LSTM	0.6593 ± 0.018	0.844 ± 0.014	0.444 ± 0.043
Ours	0.724 ± 0.015	0.918 ± 0.012	0.881 ± 0.022

Finally, this section compared the SACNN model with other existing cross-modal models, including the transfer learning [27], Wasserstein generative adversarial network domain adaptation (WGANDA) [28], the deep subdomain associate adaptation network (DSAAN) [29], the multi-modal domain adaptive variational autoencoder (MMDA-VAE) [30], and the dynamic domain adaptation (DDA) [31]. Although these models were not trained end-to-end, they are all cross-subject models based on the SEED dataset and therefore have certain comparability. As shown in Table 7, the SACNN model achieved the best accuracy of 91.82% on the SEED dataset, which is higher than the other compared models.

Table 7. The comparison models result on SEED dataset.

Model	Transfer Learning	WGANDA	DSAAN	MMDA-VAE	DDA	Ours
Res	76.31%	87.07%	88.25%	89.64%	91.08%	91.82%

4. Conclusions

In the field of cognitive neuroscience, the analysis of brain activity has been recognized as a valuable tool for investigating the mechanisms underlying emotional and behavioral responses in humans. Specifically, electroencephalography (EEG) has been used to study the brain's typical patterns of activation and their corresponding levels under various contextual stimuli during driving. Studies have revealed that negative emotional changes in humans are often associated with specific activation signals in distinct brain regions, thereby enabling a deeper understanding of the underlying mechanisms of human emotional responses. Consequently, machine learning models that are trained using relevant EEG

channels have shown to effectively improve the accuracy of model recognition for emotional state identification. These findings highlight the potential utility of EEG and cognitive neuroscience techniques in enabling an automatic acquisition of emotional information, which can facilitate the prediction and regulation of subsequent behavior and enhance the machine's emotional intelligence.

In this paper, we propose a novel method called the SACNN for recognizing emotions from raw SEED EEG data, which is capable of recognizing EEG emotion cross different subjects. Through ablation experiments, we compared the effectiveness of BN, dropout layers, and multilayer CNNs for EEG emotion recognition with CNN networks. Our results indicate that the BN layer is more suitable in Conv blocks, while the dropout layer is more suitable in FCN blocks, and both are better to add in single locations rather than multiple locations. We also found that too many dropouts will hinder the model's ability to learn corresponding features. For multilayer CNNs, fewer layers with large perceptual fields and large pooling is better than deeper layers for EEG emotion recognition. In many practical situations, deep neural networks perform better than shallow neural networks. However, in certain cases, shallow neural networks can lead to better results. In the field of EEG emotion recognition, it has been found that shallow neural networks perform better than deep neural networks, with better interpretability, possibly due to the closed form under linear combinations in limited data [32]. Similarly, experiments conducted in Kim D E's work [33] have demonstrated that shallow neural networks can successfully identify malicious network traffic more effectively than complex deep neural networks, with the former achieving an average detection rate of 98.50% compared to the latter's average high level of 48.30%. Furthermore, David Anderson presented "A Two-Stage Deep Learning Approach to Chest X-Ray Analysis" at the Denver Medical Imaging Informatics Conference in 2019, demonstrating that using two "shallow" neural networks in building diagnostic models can achieve faster, more accurate, and more interpretable AI in radiology and other imaging technologies. This indicates that under specific conditions, shallow neural networks can outperform deep neural networks.

We used raw SEED EEG data-trained SACNN, with the top 10 channels for accuracy being F8, T7, FT7, FC6, FPZ, FT8, FP2, F6, C5, and FP1. Among them, FT7, T7, C5, and F6, as well as F8, FC6, and FT8 exhibit oblique symmetry, which allows for the further exploration of functional connectivity in the brain during emotionally evoked states.

We compared SACNN with three other competitive models, namely DBN, SVM, and LSTM, which were trained using the same inputs as SACNN. Our results demonstrate that SACNN outperformed the other models, achieving a higher accuracy rate of 88.16% and 91.85% in all channel data and multichannel data, respectively. In contrast, the temporal feature dimension extraction of DBN and SVM models was not as high as that of SACNN, and the CNN-LSTM model's reconstruction of data to fit the characteristics of its spatiotemporal network model yielded lower results than SACNN.

There is still a lot to explore regarding CNN feature extraction in relation to EEG emotion temporal aspects, and more efficient applications of neural network learning based on the theoretical meaning of itself. As the same model may yield different results after data reconstruction and processing, it is worth exploring further.

There are still many areas for improvement in the experiment. For example, the current model construction is based solely on the SEED dataset, and in future work, we will validate the conclusions on the generality of the BN layer and dropout layer positions for the model on more datasets. Additionally, the current data can only distinguish whether negative emotions are present, without differentiating their degree. Therefore, in future work, we will further explore the differentiation of the degree of negative emotions and their correlation with the level of human behavioral danger, with the aim of finding a benchmark between normal and dangerous states.

By exploring the intricate relationship between functional brain activity and emotion, it is possible to combine this knowledge with a subsequent behavior analysis within context-aware environments to automatically acquire emotional information and calculate

and recognize emotional states in practical application settings. This approach can enable the prediction and regulation of subsequent behavior, thereby endowing machines with a certain degree of emotional intelligence. Going forward, there is a promising avenue for further development by utilizing multi-channel and multi-modal signal-fusion recognition methods, which can enable the achievement of a higher accuracy in emotion recognition tasks. Thus, this area of research has significant potential for advancing our understanding of the interaction between the brain and emotions, as well as for developing novel applications in various domains, including human–machine interactions and mental health.

Author Contributions: Conceptualization, M.C. and G.F.; methodology, Z.W.; software, Z.W.; validation, Z.W.; formal analysis, Z.W.; investigation, Z.W.; resources, M.C.; writing—original draft preparation, Z.W.; writing—review and editing, G.F.; supervision, M.C. and G.F.; project administration, M.C. and G.F.; funding acquisition, M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Shanghai Science and Technology Innovation Action Planning grant number No. 20dz1203800.

Institutional Review Board Statement: This study did not require ethical approval.

Informed Consent Statement: Not applicable.

Data Availability Statement: Dataset: <https://bcmi.sjtu.edu.cn/~seed/seed.html> (accessed on 1 May 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Matthias, G.; Hu, Y.H.; Zhou, X.; Ning, B.W.; Mi, J.Y.; Liu, J.; Wang, G.Y.; Wang, J.; Dong, C.; Zhang, L.D. Road traffic and transport safety development report. *China Emerg. Manag.* **2017**, *2018*, 48–58.
2. Fajola, A.; Oduneye, F.; Ogbimi, R.; Mosuro, O.O.; Oyo-Ita, A.A.; Ovwigho, U. Taking Goal Zero Outside the Fence: Lifestyle and Health Influences on Tanker and Commercial Drivers' Performance and Road Safety. In Proceedings of the SPE International Conference and Exhibition on Health, Safety, Environment, and Sustainability, Bogotá, Colombia, 27–31 July 2020; pp. 110–120.
3. Niu, Z.; Lin, M.; Chen, Q.; Bai, L. Correlation analysis between risky driving behaviors and characteristics of commercial vehicle drivers. In Proceedings of the 2015 International Conference on Information Technology and Intelligent Transportation Systems ITITS 2015, Xi'an, China, 12–13 December 2015; pp. 677–685.
4. Kaplan, S.; Guvensan, M.A.; Yavuz, A.G.; Karalurt, Y. Driver Behavior Analysis for Safe Driving: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 3017–3032. [[CrossRef](#)]
5. Zhang, L.; Liu, T.; Pan, F.; Guo, T.; Liu, R. Analysis of the influence of driver factors on road traffic accident indicators. *China J. Saf. Sci.* **2014**, *24*, 79–84.
6. Gietelink, O.; Ploeg, J.; Schutter, B.D.; Verhaegen, M. Development of advanced driver assistance systems with vehicle hardware-in-the-loop simulations. *Veh. Syst. Dyn.* **2006**, *44*, 569–590. [[CrossRef](#)]
7. Jo, J.; Lee, S.J.; Jung, H.G.; Park, K.R.; Kim, J. Vision-based method for detecting driver drowsiness and distraction in driver monitoring system. *Opt. Eng.* **2011**, *50*, 7202–7208. [[CrossRef](#)]
8. Peng, L.; Wu, C.; Huang, Z.; Zhong, M. Novel vehicle motion model considering driver behavior for trajectory prediction and driving risk detection. *Transp. Res. Rec.* **2014**, *2434*, 123–124. [[CrossRef](#)]
9. Stemmler, G. Somatovisceral Activation during Anger. In *International Handbook of Anger*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 103–121.
10. Da Silva, F.L. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2005.
11. Korats, G.; Le Cam, S.; Ranta, R.; Hamid, M. Applying ICA in EEG: Choice of the Window Length and of the Decorrelation Method. In Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies, Valetta, Malta, 1–4 February 2012; Springer: Berlin/Heidelberg, Germany, 2012.
12. Yang, H.; Han, J.; Min, K. A Multi-Column CNN Model for Emotion Recognition from EEG Signals. *Sensors* **2019**, *19*, 4736. [[CrossRef](#)]
13. Ekman, P.; Davidson, R.J. Voluntary smiling changes regional brain activity. *Psychol. Sci.* **1993**, *4*, 342–345. [[CrossRef](#)]
14. Wei, C.; Chen, L.; Song, Z.; Lou, X.G.; Li, D.D. EEG-based emotion recognition using simple recurrent units network and ensemble learning. *Biomed. Signal Process. Control* **2020**, *58*, 101756. [[CrossRef](#)]
15. Kamble, K.; Sengupta, J. A comprehensive survey on emotion recognition based on electroencephalograph (EEG) signals. *Multimed. Tools Appl.* **2023**, 1–36. [[CrossRef](#)]

16. Zheng, W.L.; Zhu, J.Y.; Lu, B.L. Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* **2017**, *10*, 417–429. [[CrossRef](#)]
17. Duan, R.N.; Zhu, J.Y.; Lu, B.L. Differential entropy feature for EEG-based emotion classification. In Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), San Diego, CA, USA, 6–8 November 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 81–84.
18. Kamble, K.S.; Sengupta, J. Ensemble machine learning-based affective computing for emotion recognition using dual-decomposed EEG signals. *IEEE Sens. J.* **2021**, *22*, 2496–2507. [[CrossRef](#)]
19. Khare, S.K.; Bajaj, V. An evolutionary optimized variational mode decomposition for emotion recognition. *IEEE Sens. J.* **2020**, *21*, 2035–2042. [[CrossRef](#)]
20. Keelawat, P.; Thammasan, N.; Kijirikul, B.; Numao, M. Subject-Independent Emotion Recognition During Music Listening Based on EEG Using Deep Convolutional Neural Networks. In Proceedings of the 2019 IEEE 15th International Colloquium on Signal Processing Its Applications (CSPA), Pulau Pinang, Malaysia, 8–9 March 2019; pp. 21–26.
21. Yang, Y.; Wu, Q.M.J.; Zheng, W.L.; Lu, B.L. EEG-based emotion recognition using hierarchical network with subnetwork nodes. *IEEE Trans. Cogn. Dev. Syst.* **2017**, *10*, 408–419. [[CrossRef](#)]
22. Wang, Y.; Qiu, S.; Li, J.; Ma, X.; Liang, Z.; Li, H.; He, H. EEG-based emotion recognition with similarity learning network. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1209–1212.
23. Li, X.; Song, D.; Zhang, P.; Yu, G.; Hou, Y.; Hu, B. Emotion recognition from multi-channel EEG data through Convolutional Recurrent Neural Network. In Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Shenzhen, China, 15–18 December 2016; pp. 352–359.
24. Song, T.; Zheng, W.; Song, P.; Cui, Z. EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks. *IEEE Trans. Affect. Comput.* **2020**, *11*, 532–541. [[CrossRef](#)]
25. Liu, J.; Zhang, L.; Wu, H.; Zhao, H. Transformers for EEG emotion recognition. *arXiv* **2021**, arXiv:2110.06553.
26. Lin, Y.P.; Jung, T.P. Improving EEG-based emotion classification using conditional transfer learning. *Front. Hum. Neurosci.* **2017**, *11*, 334. [[CrossRef](#)]
27. 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.
28. Luo, Y.; Zhang, S.Y.; Zheng, W.L.; Lu, B.L. WGAN domain adaptation for EEG-based emotion recognition. In Proceedings of the International Conference on Neural Information Processing, Siem Reap, Cambodia, 13–16 December 2018; Springer: Cham, Switzerland, 2018; pp. 275–286.
29. Meng, M.; Hu, J.; Gao, Y.; Kong, W.; Luo, Z. A deep subdomain associate adaptation network for cross-session and cross-subject EEG emotion recognition. *Biomed. Signal Process. Control* **2022**, *78*, 103873. [[CrossRef](#)]
30. Wang, Y.; Qiu, S.; Li, D.; Du, C.; Lu, B.L.; He, H. Multi-modal domain adaptation variational autoencoder for eeg-based emotion recognition. *IEEE/CAA J. Autom. Sin.* **2022**, *9*, 1612–1626. [[CrossRef](#)]
31. Li, Z.; Zhu, E.; Jin, M.; Fan, C.; He, H.; Cai, T.; Li, J. Dynamic Domain Adaptation for Class-Aware Cross-Subject and Cross-Session EEG Emotion Recognition. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 5964–5973. [[CrossRef](#)]
32. Herrera, S.R.; Ceberio, M.; Kreinovich, V. When Is Deep Learning Better and When Is Shallow Learning Better: Qualitative Analysis. *Int. J. Parallel Emerg. Distrib. Syst.* **2022**, *37*, 589–595. [[CrossRef](#)]
33. Kim, D.E. Comparison of Shallow and Deep Neural Networks in Network Intrusion Detection. Master’s Thesis, California State University, Fullerton, CA, USA, 2017.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.