Article

EEG-Based Emotion Recognition Using a 2D CNN with Different Kernels

Yuqi Wang [1,2], Lijun Zhang [1,2], Pan Xia [2,3], Peng Wang [2,3], Xianxiang Chen [2,3], Lidong Du [2,3], Zhen Fang [2,3,4,*] and Mingyan Du [5,*]

1   Institute of Microelectronics of Chinese Academy of Sciences, Beijing 100029, China; wangyuqi182@mails.ucas.edu.cn (Y.W.); zhanglijun@ime.ac.cn (L.Z.)
2   University of Chinese Academy of Sciences, Beijing 100049, China; xiapan17@mails.ucas.edu.cn (P.X.); wangpeng01@aircas.ac.cn (P.W.); chenxx@aircas.ac.cn (X.C.); lddu@mail.ie.ac.cn (L.D.)
3   Aerospace Information Research Institute, Chinese Academy of Sciences (AIRCAS), Beijing 100190, China
4   Personalized Management of Chronic Respiratory Disease, Chinese Academy of Medical Sciences, Beijing 100190, China
5   China Beijing Luhe Hospital, Capital Medical University, Beijing 101199, China
*   Correspondence: zfang@mail.ie.ac.cn (Z.F.); my_du@sina.com (M.D.)

Abstract: Emotion recognition is receiving significant attention in research on health care and Human-Computer Interaction (HCI). Due to the high correlation with emotion and the capability to affect deceptive external expressions such as voices and faces, Electroencephalogram (EEG) based emotion recognition methods have been globally accepted and widely applied. Recently, great improvements have been made in the development of machine learning for EEG-based emotion detection. However, there are still some major disadvantages in previous studies. Firstly, traditional machine learning methods require extracting features manually which is time-consuming and rely heavily on human experts. Secondly, to improve the model accuracies, many researchers used user-dependent models that lack generalization and universality. Moreover, there is still room for improvement in the recognition accuracies in most studies. Therefore, to overcome these shortcomings, an EEG-based novel deep neural network is proposed for emotion classification in this article. The proposed 2D CNN uses two convolutional kernels of different sizes to extract emotion-related features along both the time direction and the spatial direction. To verify the feasibility of the proposed model, the pubic emotion dataset DEAP is used in experiments. The results show accuracies of up to 99.99% and 99.98% for arousal and valence binary classification, respectively, which are encouraging for research and applications in the emotion recognition field.

Keywords: emotion recognition; machine learning; convolutional neural network; electroencephalogram

1. Introduction

1.1. Background

Emotions are mainly the individuals' inner responses (such as attention, memorization, achieving goals, awareness of priority, knowledge motivation, communication with others, learning development, mood status, and effort motivation) [1,2] to whether or not the objective conditions meet their psychological expectations which could also reflect their attitudes and perceptions [3]. Preliminary researchers in neuroscience, psychology, and cognitive science have confirmed that emotions play a key role in rational decision making, interpersonal communication, learning, memory, physical, and even mental health [4]. Long-term negative emotions and depression can interfere with individuals' normal behaviors and health [5]. Affective computing is a branch of artificial intelligence that relates to, arises from, or influences emotions and it has emerged as a significant field of study that aims to develop systems that can automatically recognize emotions [2]. In the past decade, there have been a variety of methods to recognize emotions, including those based on

self-evaluation (such as the Self-Assessment Manikin (SAM)) [6], which reports behavioral responses (such as vocal intonations, facial expressions, and body postures), and physiological signals. However, due to the social nature of human beings, individuals usually do not desire to express their real emotional states, they habitually disguise their expressions and movements when a camera is on them [7]. Moreover, to ensure the monitored data is of high quality, users are required to maintain a stable posture and position in front of the camera, microphone, and other sensors. These limitations lead it to be challenging to apply in a practical sense. By contrast, individuals are not able to disguise or conceal their physiological responses easily, thus, physiological signals can truly reflect the emotional changes of users. Consequently, methods of recognition based on physiological signals have become mainstream [8]. Among these physiological signals used in emotion recognition, electrocardiogram (ECG), respiratory (RSP), body temperature, etc. have their disadvantages since the variability of these signals is usually subtle and the change rate is typically slower. By contrast, electroencephalography (EEG) signals have the advantage of a high real-time differential and of being non-fakeable. In addition, emotion is essentially regulated by the central nervous system. Therefore, using EEG signals to recognize emotion is usually more accurate and objective than using other peripheral physiological signals [9–13]. EEG-based emotion recognition has attracted an increasing number of scholars and has been proved to be an effective method of emotion recognition in a multitude of research [14].

1.2. Related Work

1.2.1. Emotion Model

Since people have different ways to express their emotions, judging their emotions is a challenging task. There are two main models of emotion. Some researchers hold the view that emotions are composed of several basic discrete emotions. For example, Ekman believes that emotions consist of happiness, sadness, fear, and anger. Under different cultural backgrounds and social environments, these basic emotions form more complex emotions by a combination in different forms [15]. By contrast, other researchers believe that emotional states are consecutive and indissociable. Plutchik proposed the famous emotion wheel model. As Figure 1 shows, the middle circle indicates the basic emotions. The outer circle and the inner circle respectively represent the undersaturation and the oversaturation of basic emotion [16].
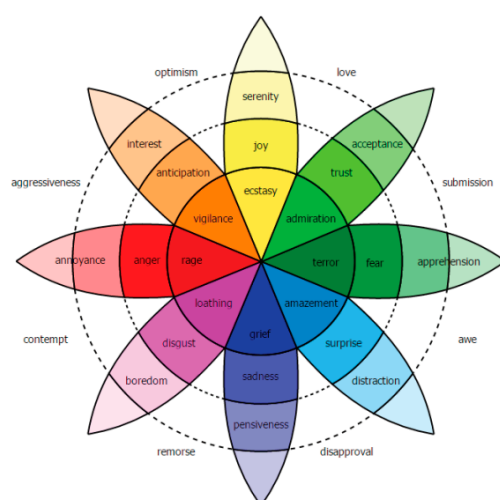

Figure 1. Plutchik emotion wheel.

The most widely accepted emotion model is the two-dimensional model 'Arousal-Valence', proposed in 1980 by Russell. Figure 2 illustrates the Arousal-Valence model, the x-axis represents the Valence Dimension and the y-axis represents the Arousal Dimension.

Different emotions can be located in this model. The emotion model used in the most popular public emotion dataset 'DEAP' is the extended version of 'Arousal-Valence'.
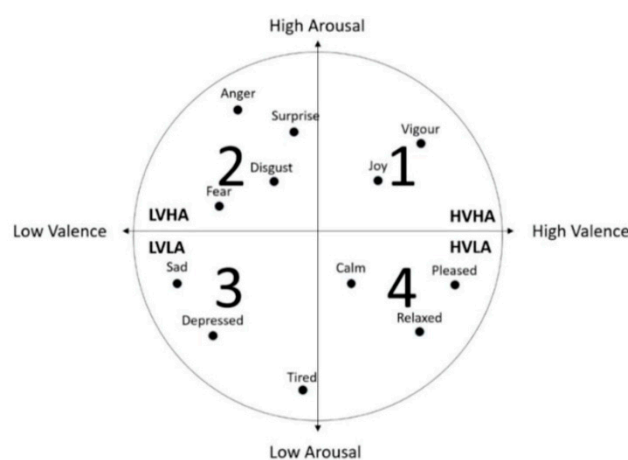

Figure 2. The Arousal-Valence model.

1.2.2. EEG-Based Emotion Recognition

Most existing approaches are based on machine learning techniques for EEG emotion recognition [17]. For the classifiers in emotion recognition, the traditional machine learning algorithms such as support vector machine (SVM) and k-nearest neighbor (KNN) are frequently used and achieve good results. Kroupi designed a linear discriminant analysis model as the classification and used the power spectral density as a feature to recognize emotion [18]. Bahari used a k-Nearest to classify the emotion and got the accuracy of 64.56%, 58.05%, and 67.42% for three classes of arousal, valence, and liking [19]. Zheng proposed selecting 12 channel electrode features in SVM, which provided 86.65% on average [20].

However, the methods that use traditional machine learning algorithms have required the extraction of the emotion-related features from the origin EEG fragment. The extraction is time-consuming and the process uncertain. Moreover, the emotion recognition accuracies of these methods could be improved. Therefore, deep learning-based methods in emotion recognition have become increasingly popular. Tripathi used a DNN as the classifier to obtain better results as the accuracy of valence and arousal is 81.4 and 73.4%, respectively [21]. Zhang used the sparse autoencoder (SAE) and logistic regression to predict the emotion status. The recognition accuracy has improved to 81.21% for valence and 81.26% for arousal [22]. Nevertheless, the accuracy of emotion recognition by using CNN or SAE is still not high. Alhagry proposed a long-short term (LSTM) model to address emotion recognition. They used the DEAP dataset to test the method and the accuracies were 85.45% and 85.65% for valence and arousal, respectively [23]. Salama recognized emotions by a 3D convolutional neural network(3D-CNN) model. They extracted multi-channel EEG signals into 3D data for the Spatio-temporal feature extraction. The recognition accuracies on the DEAP dataset were improved to 87.44% and 88.49% for valence and arousal, respectively [24]. Song designed a dynamical graph convolutional neural network (DGCNN) which used graph relation to represent EEG and then use graph convolution network (GCN) to classify emotion. They tested the method on the DREAMER database and achieved the recognition accuracies of 86.23%, 84.54%, and 85.02% for valence, arousal, and dominance, respectively [25]. Zhong presented a regularized graph neural network (RGNN) to capture both local and global inter-channel relations and the accuracy is 85.30 on the SEED dataset [26]. Yin combined the graph CNN and LSTM. They took advantage of both methods to extract the graph domain features for emotion recognition and attained the average classification accuracy of 84.81% and 85.27% for valence and arousal [27]. Yang subtracted the Base Mean outcome from raw EEG data, then the processed data were converted to 2D EEG frames. They proposed a fusion model of CNN and LSTM and

achieved high performance with a mean accuracy of 90.80% and 91.03% on valence and arousal classification tasks respectively [28]. Liu used a deep neural network and sparse autoencoder combined model to classify emotion [29]. Zhang tried many kinds of deep learning methods to classify emotions and got the best performance by using a CNN-LSTM model (accuarcy:94.17%) [30]. Donmez designed their own experiment and collected an emotion dataset, then used CNN to classify it and obtained an accuracy of 84.69% [31]. There are also some studies whose goals are not concerned with emotional recognization, but the methods they used for the classification of EEG signals using deep learning are worth learning from. For example, Abdani used Convolutional Neural Networks to test subjects to determine whether the task is new or routine [32], and Anwar used the AlexNet to classify motor imagery tasks [33].

1.3. The Contributions of This Study

There are still three main limitations, however, that need to be addressed in the studies in the field of emotion recognition to improve performance. The first one is the feature extraction problem. The shallow traditional machine learning models such as SVM and KNN used in emotion recognition require researchers to extract the emotional-related features manually as the input of their models. Some studies that used the deep learning model also extracted features manually to improve the classification performance. Manually extracting these features is time-consuming and the quality of these extracted features is also unstable given the involvement of human subjective consciousness and experience. Another main issue is that researchers used the user-dependent model to improve their model's performance. The training and testing data are chosen from the same subject in the User-dependent model. Thus, the User-dependent emotion recognition model typically shows high accuracy. Nevertheless, the User-dependent model lacks generalization and universality. For every different subject, the User-dependent model requires training data of the specific individual to perform the tuning process, which means the model needs to be retrained every time the subject changes. The accuracies of most emotion recognition models are not high enough, thus there is still room for improvement. To address the above issues, we, therefore, propose a novel 2D-CNN model with two different sizes of convolution kernels that convolve along with the time and space directions, respectively. The proposed model does not need to manually extract the emotional-related features. The emotion recognition classification result could be directly derived from the raw EEG data by the proposed model, which realizes the end-to-end functionality. Furthermore, the proposed model is a User-independent model that is more applicable to new users because there is no need to create a new model for every single subject. The model just needs to be trained once for all the subjects and it could effectively monitor the emotion of a new user. Finally, the effectiveness of our model is examined on the DEAP dataset. The proposed model achieves the state-of-the-art accuracy of 99.97% on the valence and 99.93% on the arousal.

The layout of the paper is as follows: In Section 1, important background on emotion reignition is described and also previous works are reviewed. Section 2 begins by introducing the domains of EEG signals and how they will be used in emotion recognition. It then introduces the DEAP dataset and the process of the dataset. The experiment setup and the entire process of our model are presented in detail at the end of Section 2. Section 3 explains the results achieved by the proposed method to demonstrate its effectiveness. Finally, the conclusion of this work and future work follows in Section 4.

2. Materials and Methods

2.1. EEG on Emotion

As the reflection of the central nervous system (CNS), asynchronous activity occurs in different locations of the brain during emotions [34]. Therefore, EEG can reveal significant information on emotions [35]. EEG is a waveform recording system that reads scalp electrical activity generated by the human brain over a period of time. It measures voltage

fluctuations resulting from the ionic current flowing through the neurons of the brain. The EEG signal has a low amplitude which ranges from 10 uV to 100 uV [36]. The frequency range of EEG signals is typically 0.5–100 Hz [37]. According to various mental states and conditions, researchers divide EEG signals into five frequency sub-bands that are named the delta (1–4 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (>30 Hz), respectively [38].

2.2. The Dataset and Process

The majority of the past studies about EEG—based emotion recognition in the last 10 years used the public open dataset to compare with other researchers and to demonstrate the advantages of their methods. The most used public open datasets are the MAHNOB-HCI, the DEAP, and the SEED. Among the articles conducted by public open dataset, most of them used the DEAP dataset [39]. To evaluate our proposed model, we adopted the most popular emotion dataset, DEAP, to conduct the experiments and verify the effectiveness. The DEAP (Database for Emotion Analysis using Physiological Signals) is a public open emotion dataset collected by researchers from the Queen Mary's University of London, the University of Twente in the Netherlands, the University of Geneva in Switzerland, and the Swiss Federal Institute of Technology in Lausanne. They recorded the EEG and peripheral physiological signals of 32 healthy participants aged between 19 and 32 with an equal number of males and females. All 32 subjects were stimulated to produce certain emotions. Subjects were asked to relax for the first two minutes of the experiment. The first twominute recording is regarded as the baseline. Then subjects were asked to watch 40 music video excerpts of oneminute duration and these served as the sources of emotion elicitation. Figure 3 shows the process of the emotion elicitation experiment.
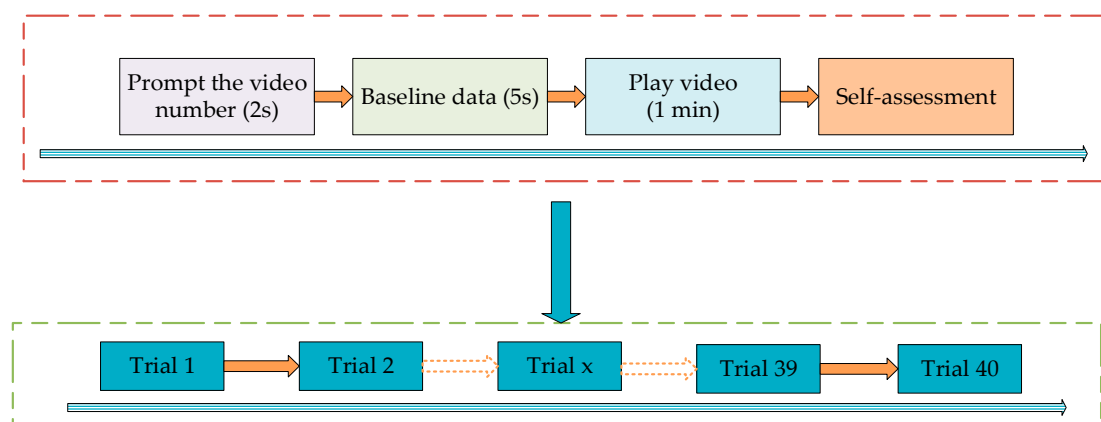

Figure 3. Experiment process.

During this process, the EEG signals were recorded, at a sampling rate of 512 Hz, from 32 electrodes placed on the scalp according to the international 10–20 system [40]. At the same time, the peripheral physiological signals including skin temperature, blood volume pressure, electromyogram, and galvanic skin response were recorded from the other 8 channels. Every subject completed 40 trials corresponding to the 40 videos. As a result, there are 1280 (32 subjects × 40 trials) signal sequences in the DEAP dataset. For each trial, the first 3 s are used as a baseline because participants did not watch the videos during this time. Then they watched the one minute video. At the end of each trial, subjects were asked to perform a self–assessment to evaluate their emotional levels of arousal, valence, liking, and dominance in the range of 1 to 9. The self–assessment scales are a manikin designed by Morris, as shown in Figure 4 [41].
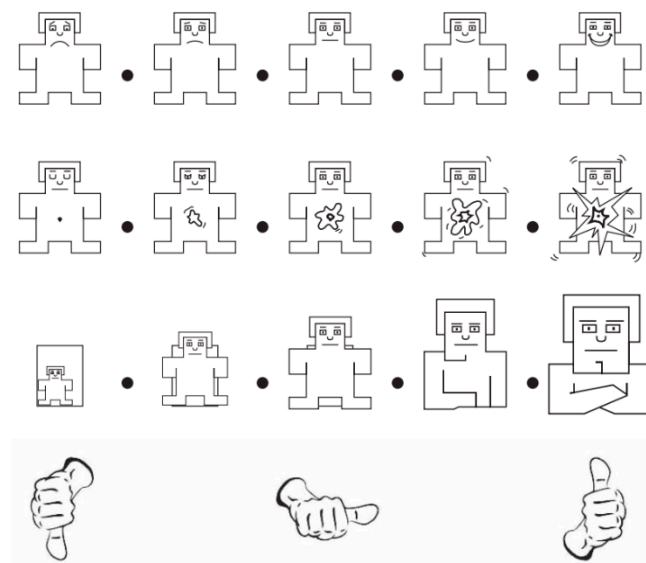
Figure 4. The self–assessment scales.

This scale represents from top to bottom, the levels of Valence, Arousal, Dominance, and Liking, respectively. Furthermore, DEAP provides a preprocessed version of the recorded EEG signals. In this study, we used the preprocessed version that the EEG signal is downsampled into 128 Hz. To reduce the noises and cut the EOG artifacts, the EEG signal was filtered by a band-pass filter with a frequency from 4 Hz to 45 Hz. The size of the physiological signal matrix for each subject is $40 \times 40 \times 8064$, which corresponds to 40 trials $\times$ 40 channels $\times$ 8064 (128 Hz $\times$ 63 s) sampling points [40].

2.3. Experiment Setting

Valence and Arousal are the most dimensional expression of basic emotions that researchers usually focus on. Therefore, in this study, we chose Valence and Arousal as the two scales for emotion recognition. Among the 40 channels of the DEAP dataset, we chose the first 32 channels that contain the EEG signal as the basis of emotion recognition. The labels of each data depend on the self-assessment rating values of Arousal and Valence states. We divide the rating values of 1-9 into two binary classification problems with a threshold of 5: If the self-assessment is more than 5, the label of this data is 1 (represents high valence/arousal), otherwise, the label of this data is 0 (represents to low valence/arousal). The process of recognizing emotions using EEG signals is shown in Figure 5.
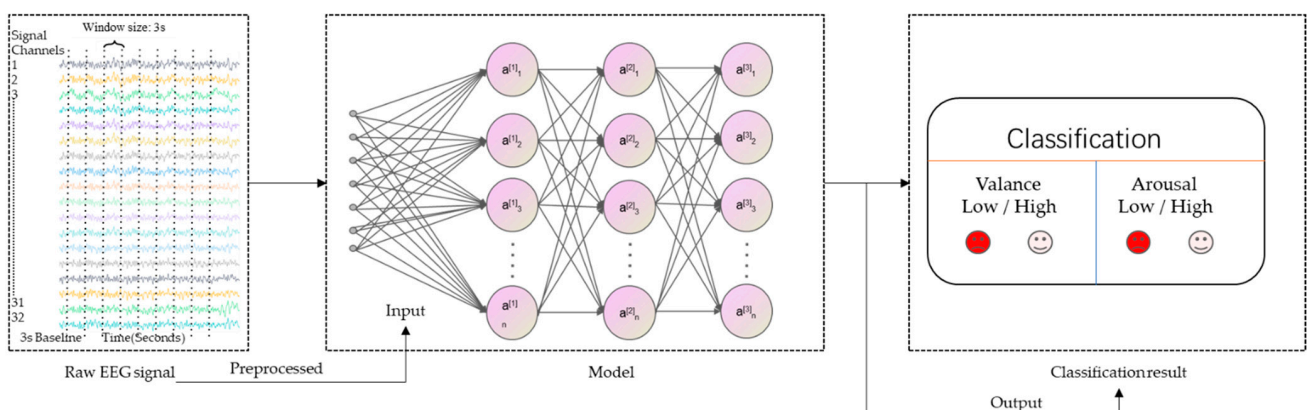

Figure 5. Process of emotion recognition using EEG.

The raw EEG signals from the DEAP original version were preprocessed to reduce the noise and cut the EOG artifacts. After this, the preprocessed version DEAP dataset was produced. Then the preprocessed raw EEG signals are extracted for the deep learning model. Finally, the emotion recognition classification results are obtained after the training and testing of the model.

2.4. Proposed Method

2.4.1. Deep Learning Framework

To have the full benefit of the preprocessed raw EEG structure, we propose a two-dimensional (2D) Convolutional Neural Network (CNN) model in this study for emotion recognition. CNN is a class of deep neural networks widely used in a number of fields [42].

Compared with traditional machine learning methods, 2D CNN models have a better ability to detect shape-related features and complex edges. In a typical CNN network, there could be components named convolutional layers, pooling layers, dropout layers, and fully connected layers. Features from original EEG signals are concatenated into images and then sent to convolutional layers. After being convolved in convolutional layers, the data is further subsampled to smaller size images in pooling layers. During the process, network weights are learned iteratively through the backpropagation algorithm.

The input vector of the CNN structure is the two-dimensional feature, shown below:

$$a = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{1}$$

The shape of the input vector a is $m \times n$. Then the input vector a is convolved with Wk in the convolution layer, Wk is given by:

$$c_k = \begin{pmatrix} c_{11} \\ c_{21} \\ \cdots \\ c_{i1} \end{pmatrix} \tag{2}$$

In Formula (2), the length of b_k is i which must be less than m in Formula (1). The feature map is finally obtained after the convolution, which is calculated as:

$$f(\alpha) = f(c_k \times a + b_k) \tag{3}$$

After the convolutional layer, BatchNorm2d is used to normalize the data. This would keep the data size from being too large and prevent gradient explosion before the LeakyReLU layer. BatcgNorm2d is calculated as:

$$y = \frac{x - \mathbf{mean}(x)}{\sqrt{\mathbf{Var}(x) + \mathbf{eps}}} * \mathbf{gamma} + \mathbf{beta} \tag{4}$$

where x is the value of the input vector, $\mathbf{mean}(x)$ is the mean value and the $\sqrt{\mathbf{Var}(x)}$ denotes the variance value. Eps is a small floating-point constant to ensure that the denominator is not zero. Gamma and beta are trainable vector parameters.

In Formula (3), f is the activation function, in this study, we use the leaky rectified linear unit (LeakyReLU).

LeakyReLU is defined in Formula (5):

$$y_\alpha = \begin{cases} x_\alpha & \text{if } x_\alpha \geq 0 \\ \frac{x_\alpha}{a_\alpha} & \text{if } x_\alpha < 0 \end{cases} \tag{5}$$

where α is defined in Formula (3) and a_i is a fixed parameter in the range of 1 to $+\infty$.

ReLU has more advantages in avoiding gradient disappearance than traditional neural network activation functions, such as *sigmoid* and *tanh*. LeakyReLU has the same advantage in avoiding gradient disappearance as ReLU. Moreover, during backpropagation, the gradient can also be calculated for the part whose input is less than zero by LeakyReLU (instead of having a value of zero as in ReLU). Thus, LeakyReLU can avoid the gradient direction sawtooth problem. The b_k is the bias value, k is the filter, and the $c_k \in Ri \times 1$ is the weight matrix. The total number of filters in the convolutional layer is denoted by *n*.

A dropout of 0.25 is used in the Leaky Relu layer for reducing the network complexity and reducing over-fitting. Thus, enhancing the generalization of the model. As Figure 6 shows, some neural network units are temporarily discarded with a certain probability, in our model, the probability is 0.25.
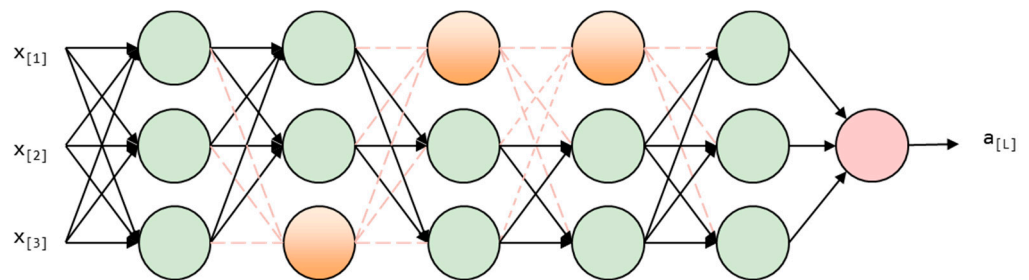

Figure 6. Dropout layer diagram.

The feature map is then downsampled through the max-pooling layer. Figure 7 shows the principle of Max pooling.
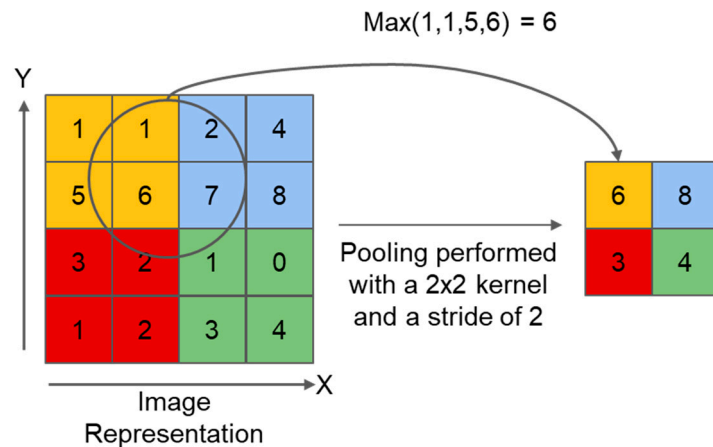

Figure 7. The process of Max pooling.

The maximum value from the given feature map is extracted by this function. The fully connected layer is flattened after the last polling layer. Finally, since the task is a binary classification task, SoftMax is used in the output layer. Adam is used as the optimizer and binary CrossEntropy is applied to calculate the model loss because the labels of the arousal or valence classification are high value and low value. The Cross-Entropy in the binary-classification task is calculated as below:

$$L = -[y^* \log \hat{y} + (1 - y) * \log(1 - \hat{y})] \tag{6}$$

For the data with *N* samples, the calculation process is:

$$L = -\sum_{n=1}^{N} \hat{y}_i \log y_i + (1 - \hat{y}_i) \log(1 - \hat{y}_i) \tag{7}$$

where N is the number of samples, y is the one-hot value, and $\hat{y}$ is the output.

2.4.2. CNN Model

To better obtain the details of EEG signals associated with emotions, the 2D CNN is employed to feature extraction and classification in our study. Table 1 shows the major hyper-parameters and the related information, such as the value or the type of them of the proposed trained CNN model.

Table 1. The value or types of the proposed model's hyper-parameters.

Hyper-Parameter of the Proposed Model	Value/Type
Batch size	128
Learning rate	0.0001
Momentum	0.9
Dropout	0.25
Number of epochs	200
Pooling layer	Max pooling
Activation function	LeakyReLU
Window size	3 S
Optimizer	Adam
Loss Function	Cross Entropy

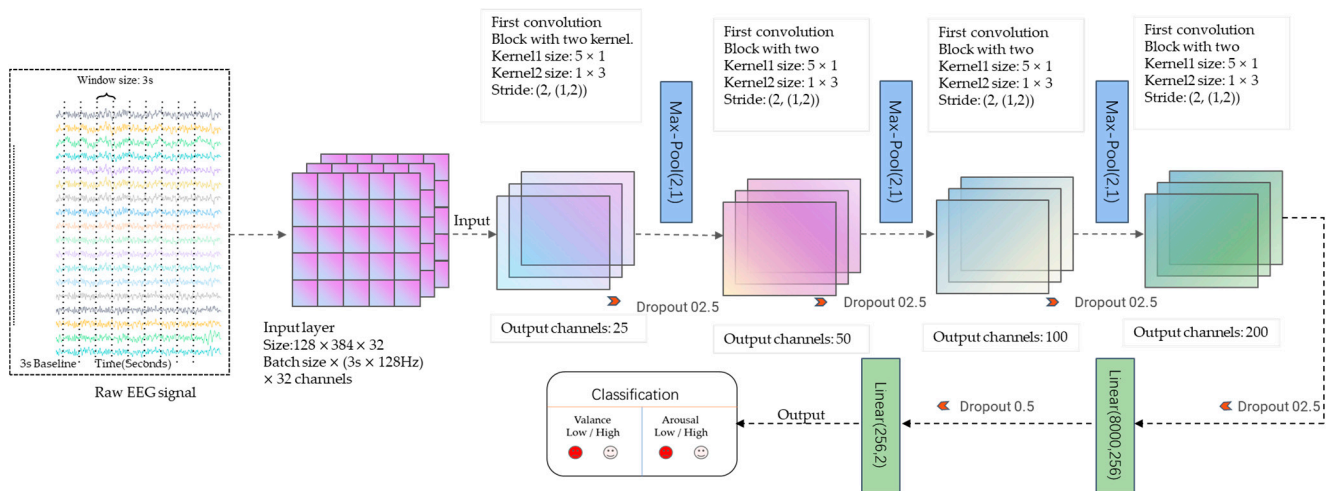Our proposed CNN architecture is illustrated in Figure 8.

Figure 8. The architecture of the proposed network.

The model is designed using Python3.7. As Figure 8 and Table 1 show, the input size is width × height, where width is 32 (the number of electrode channels), and height equals $3 \times 128 = 384$ ((window size: 3 s) × (sampling rate 128 Hz)). The batch size of the model is 128. So, the shape of the input data is (12,838,432). In the proposed model, Conv2 represents the dimensional (2D) convolutional layer, Pooling2D denotes 2D max pooling, BatchNorm2d is the 2D batch normalization, and Liner denotes the fully connected layer. Every convolution layer is followed by an activation layer, in this model we use the Leaky Relu as the activation function with the alpha = 0.3. The proposed model contains eight convolution layers, four batch normalizations, four drop-out layers with the probability of 0.25, three max-pooling layers, and two fully connected layers. All above layers collectively form the four convolution blocks. One convolution block has two convolution layers with two different convolution kernels to extract the emotion-relevant features. The two convolution kernels have distinct kernel sizes, in particular 5 × 1 and 1 × 3. The convolution kernel with size 5 × 1 convolves the data along the time direction

and the other convolution kernel with size 1 × 3 convolves along the spatial direction. Every single convolution layer uses LeakyReLU as the activation function. Each convolution block has a normalization layer and a dropout layer (0.25). The first three convolution blocks are connected with Max pooling layers at the end of them. The last convolution block is followed by a fully connected layer and it is connected to the last output fully connected layer after a dropout layer with the probability of 0.5. The final output of this model is the classification results of the emotion recognition. Table 2 shows the shapes of the proposed model.

Table 2. The shapes of the proposed model.

Number of Layers	Layer Type	Numbers of Input Channels/Output Channels
1	Input (shape:1, 384, 32)	
2	conv_1 (Conv2d)	1/25 (kernel size: 5 × 1)
3	droputout1 (Dropout=0.25)	1/25
4	conv_2 (Conv2d)	25/25 (kernel size: 1 × 3, stride = (1,2))
5	bn1 (BatchNorm2d)	25
6	pool1 (MaxPool2d (2,1))	25/25
7	conv_3 (Conv2d)	25/50 (kernel size: 5 × 1)
8	droputout2 (Dropout = 0.25)	25/50
9	conv_4 (Conv2d)	50/50 (kernel size: 1 × 3, stride = (1,2))
10	bn2 (BatchNorm2d)pool2	50
11	(MaxPool2d (2,1))	50
12	conv_5 (Conv2d)	50/100 (kernel size: 5 × 1)
13	droputout3 (Dropout = 0.25)	50/100
14	conv_6 (Conv2d)	100/100 (kernel size: 1 × 3, stride = (1,2))
15	bn3 (BatchNorm2d)	100
16	pool3 (MaxPool2d (2,1))	100
17	conv_7 (Conv2d)	100/200 (kernel size: 5 × 1)
18	droputout4 (Dropout = 0.25)	100/200
19	conv_8 (Conv2d)	200/200 (kernel size: 1 × 3)
20	bn4 (BatchNorm2d)	200
21	flatten (Flatten layer)	Shape: 128 × 8000
22	Linear1 (Linear)	8000/256
23	Droputout5 (Dropout = 0.5)	256/2 (binary classification task, number
24	Linear2 (Linear)	of classes = 2)

3. Results

We use the most popular public emotion dataset DEAP of the EEG signals to evaluate the proposed model. Seventy percent of the data of the DEAP dataset is randomly divided into the training set and the other thirty percent is the test set. The classification metric is the accuracy (ACC) which is the most commonly used evaluated guideline and represents the proportion of the sample that is classified correctly, given by:

$$\mathbf{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \times \mathbf{100}\% \tag{8}$$

where *TP*, *TN*, *FP*, and *FN* were denoted as the number of true positives, true negatives, false positives, and false negatives, respectively.

Our model achieves the accuracies of 99.99% and 99.98% on Arousal and Valence binary classification, respectively. The accuracy and loss classification of Valence and Arousal are shown in Figures 9 and 10.
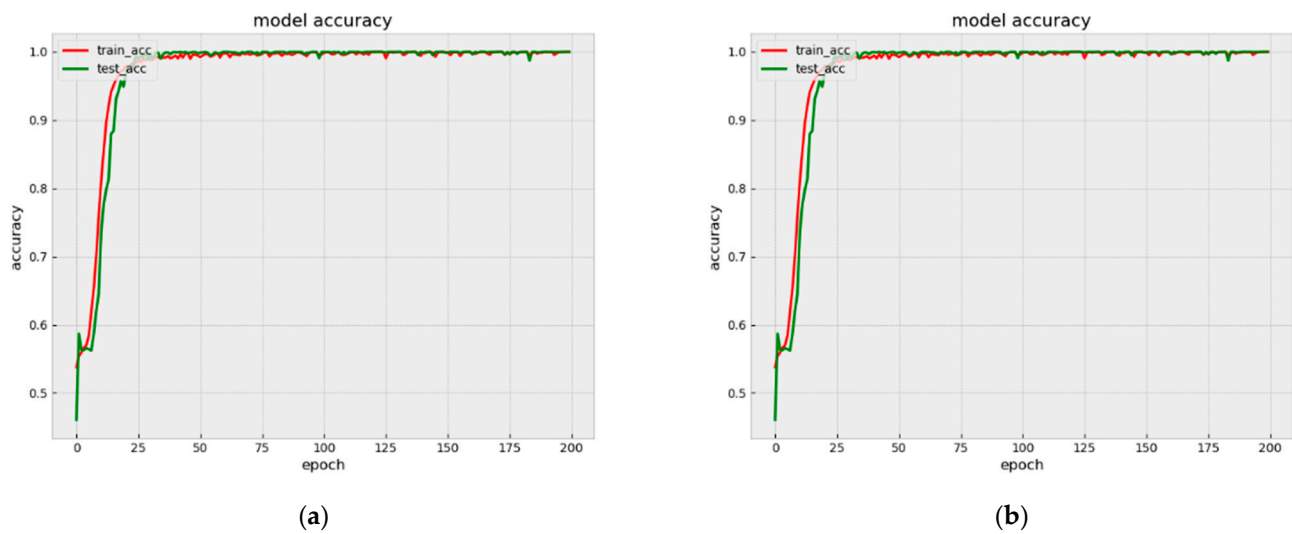
(**a**)

(**b**)

Figure 9. Model accuracy and loss curves (**a**) accuracy in Valence, (**b**) loss in Valence.
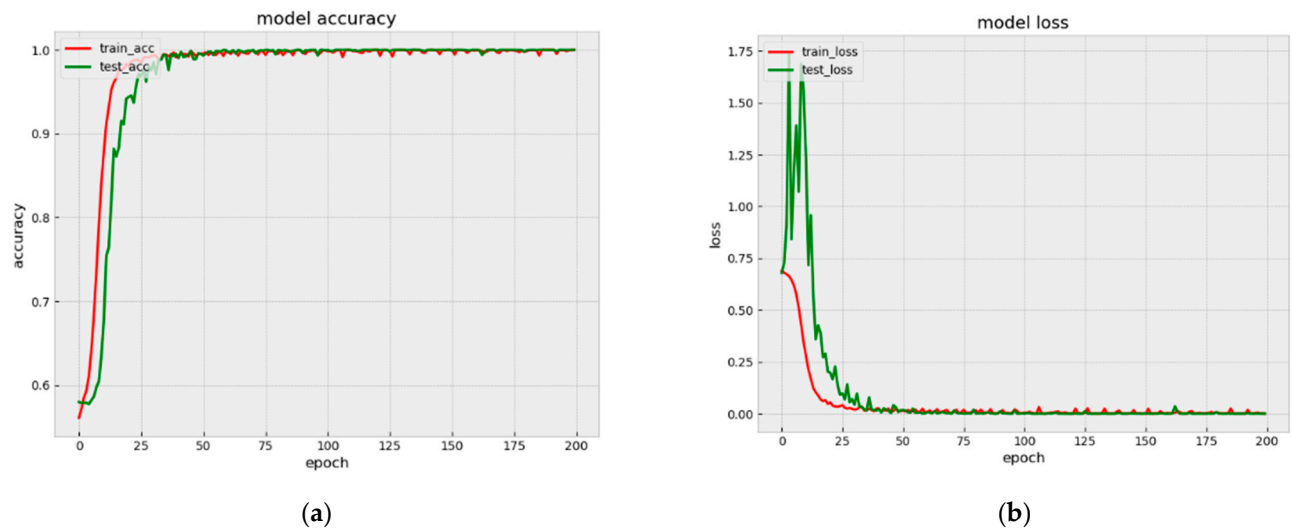

(**a**)

(**b**)

Figure 10. Model accuracy and loss curves (**a**) accuracy in Arousal, (**b**) loss in Arousal.

As can be seen from Figures 9 and 10, for both Arousal and Valence classification, the proposed model converged quickly from epoch 0 to epoch 25. Figure 11 shows the Model accuracy and loss curves during the first 25 epochs. The model accuracies have already achieved 98.36% and 98.02% on Arousal and Valence after 25 epochs, respectively. Then from epoch 25 to epoch 50, the model accuracies improved slowly. Until the 50 epochs, the model accuracies are 99.56% and 99.68% on Arousal and Valence, respectively. Then the model is stabilizing during 50 to 200 epochs and finally achieves the high accuracies of 99.99% and 99.98% on Arousal and Valence, respectively.

In Table 3, we compare the Arousal and Valence binary classification accuracies to other studies which used the DEAP dataset. From Table 3 and Figure 12, it can be concluded that among all the above methods of emotion recognition, our method has the best performance.
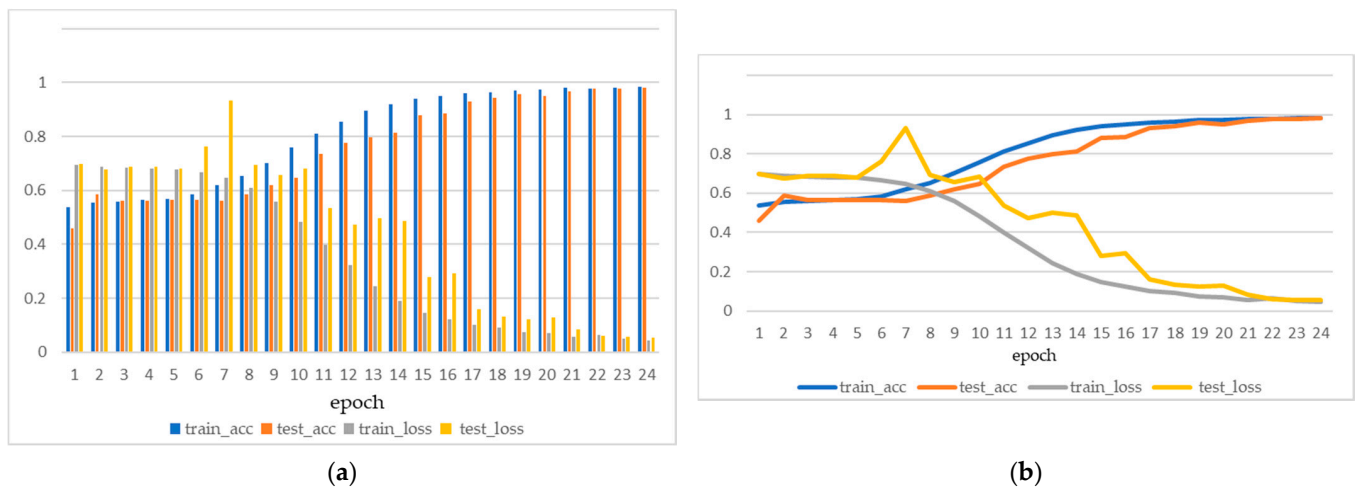
(**a**)

(**b**)

Figure 11. Model accuracy and loss curves during the first 25 epochs. (**a**) histogram, (**b**) line chart.

**Table 3.** Comparison with other studies that used the DEAP dataset.

Author	Accuracies (%)
Chao et al. [43]	Val:68.28, Aro:66.73 (deep learning)
Pandey and Seeja [44]	Val:62.5, Aro:61.25 (deep learning)
Islam and Ahmad [45]	Val:81.51, Aro:79.42 (deep learning)
Alazrai et al. [46]	Val:75.1, Aro:73.8 (traditional machine learn)
Yang et al. [28]	Val:90.80, Aro:91.03 (deep learning)
Alhagry et al. [23]	Val:85.45, Aro:85.65 (deep learning)
Liu et al. [47]	Val:85.2, Aro:80.5 (deep learning)
Cui et al. [48]	Val:96.65, Aro:97.11 (deep learning)
Vijiayakumar et al. [49]	Val:70.41, Aro:73.75 (traditional machine learn)
Li et al. [50]	Val:95.70, Aro:95.69 (traditional machine learn)
Luo [51]	Val:78.17, Aro:73.79 (deep learning)
Zhong and Jianhua [52]	Val:78.00, Aro:78.00 (deep learning)
Menezes et al. [53]	Val:88.00, Aro:69.00 (deep learning)
Zhang et al. [54]	Val:94.98, Aro:93.20 (deep learning)
Kumar et al. [55]	Val:61.17, Aro:64.84 (Non—machine learning)
Atkinson and Campos [56]	Val:73.06, Aro:73.14 (deep learning)
Lan et al. [57]	Val:73.10, Aro:71.75 (deep learning)
Mohammadi et al. [58]	Val:86.75, Aro:84.05 (deep learning)
Koelstra et al. [39]	Val:57.60, Aro:62.00 (traditional machine learn)
Tripathi et al. [21]	Val:81.40, Aro:73.40 (deep learning)
Cheng et al. [59]	Val:97.69, Aro:97.53 (deep learning)
Wen et al. [60]	Val:77.98, Aro:72.98 (deep learning)
Wang et al. [61]	Val:72.10, Aro:73.10 (deep learning)
Gupta et al. [62]	Val:79.99, Aro:79.95 (traditional machine learn)
Yin et al. [27]	Val:85.27, Aro:84.81 (deep learning)
Salama et al. [24]	Val:88.49, Aro:87.44 (deep learning)
Our method	Val:99.99, Aro:99.98
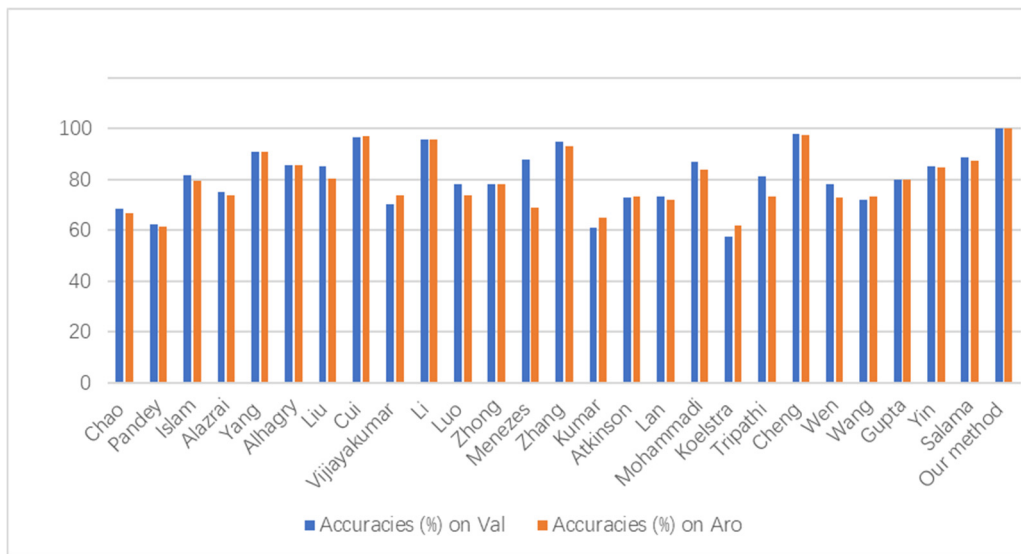
(**a**) The rank of Accuracies (%) of the above 27 works
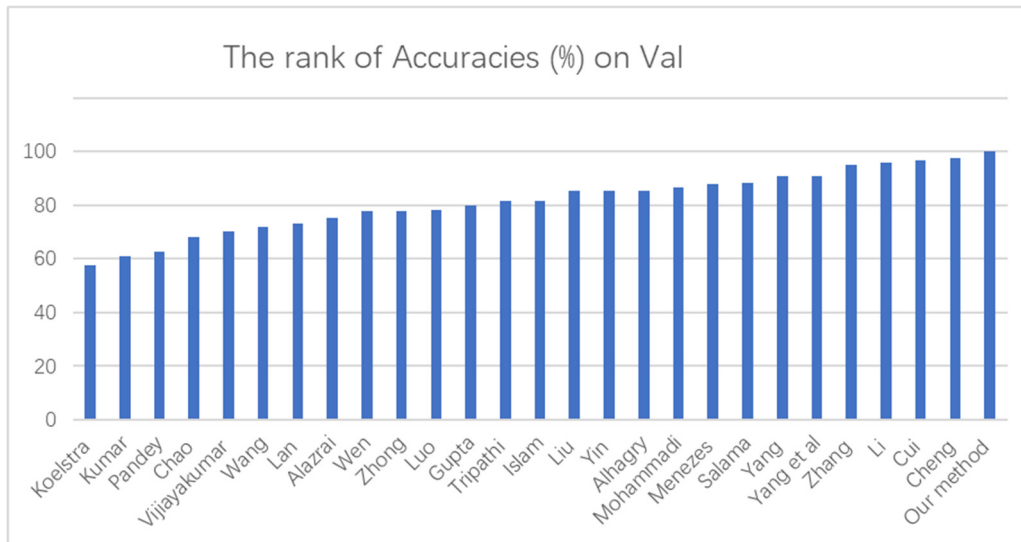
(**b**) The rank of Accuracies (%) on Val
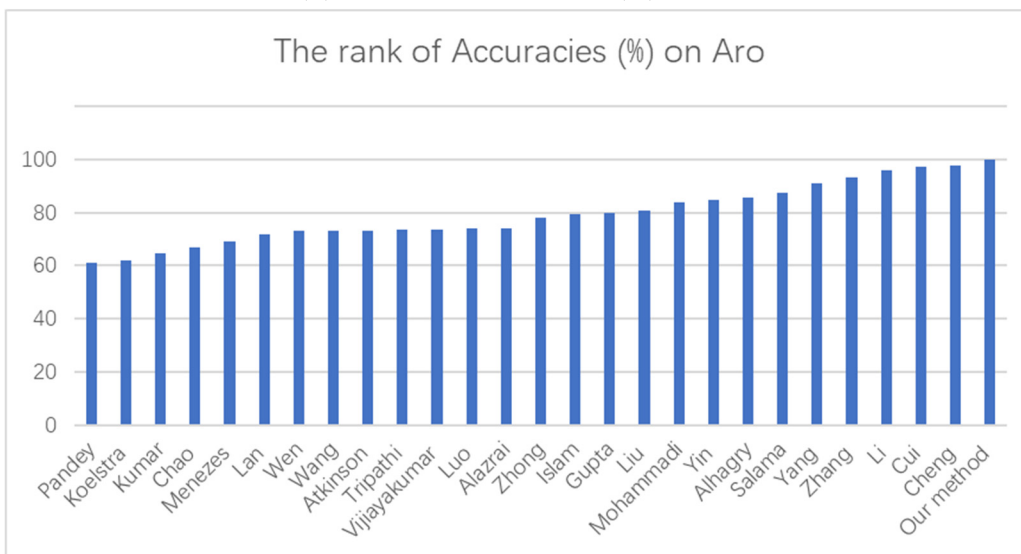
(**c**) The rank of Accuracies (%) on Aro

Figure 12. (**a**–**c**) Comparison in the form of histograms.

4. Conclusions

In this paper, we proposed the Convolutional Neural Network model with two different sizes of convolution kernels to recognize emotion from EEG signals. This model is user-independent and has high generalization and universality. It is more applicable to new users because there is no need to create a new model for every single subject. It also is an end-to-end model which is time-saving and stable. The effectiveness of our method is ascertained on the DEAP public dataset and the performance has been proved to be at the top of the area. Wearable devices have made great progress in people's daily lives, and they have been widely used in the application of healthcare [63]. Measuring EEG signals with miniaturized wearable devices becomes possible. Future works will consider designing our own emotion experiments and transferring the model to other public emotion datasets such as the SEED dataset and evaluating the performance.

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

Data Availability Statement: Data we use in this article is a public data set names"DEAP". It is available for download via this link: https://www.eecs.qmul.ac.uk/mmv/datasets/deap/ (accessed on 1 April 2022).

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

References

1. Northoff, G. *Minding the Brain: A Guide to Philosophy and Neuroscience*; Macmillan International Higher Education: London, UK, 2014.
2. Wyczesany, M.; Ligeza, T. Towards a constructionist approach to emotions: Verification of the three-dimensional model of affect with EEG-independent component analysis. *Exp. Brain Res.* **2015**, *233*, 723–733. [CrossRef] [PubMed]
3. Barrett, L.F.; Lewis, M.; Haviland-Jones, J.M. *Handbook of Emotions*; Guilford Publications: New York, NY, USA, 2016.
4. Cao, Q.; Zhang, W.; Zhu, Y. Deep learning-based classification of the polar emotions of "moe"-style cartoon pictures. *Tsinghua Sci. Technol.* **2021**, *26*, 275–286. [CrossRef]
5. Berger, M.; Wagner, T.H.; Baker, L.C. Internet use and stigmatized illness. *Soc. Sci. Med.* **2005**, *61*, 1821–1827. [CrossRef] [PubMed]
6. Bradley, M.M.; Lang, P.J. Measuring emotion: The self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* **1994**, *25*, 49–59. [CrossRef]
7. Porter, S.; Ten Brinke, L. Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychol. Sci.* **2008**, *19*, 508–514. [CrossRef]
8. Wioleta, S. Using physiological signals for emotion recognition. In Proceedings of the 2013 6th International Conference on Human System Interaction (HSI), Sopot, Poland, 6–8 June 2013; pp. 556–561.
9. Wagner, J.; Kim, J.; Andre, E. From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification. In Proceedings of the 2005 IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands, 6 July 2005; pp. 940–943. [CrossRef]
10. Kim, K.H.; Bang, S.W.; Kim, S.R. Emotion recognition system using short-term monitoring of physiological signals. *Med. Biol. Eng. Comput.* **2004**, *42*, 419–427. [CrossRef]
11. Brosschot, J.F.; Thayer, J.F. Heart rate response is longer after negative emotions than after positive emotions. *Int. J. Psychophysiol.* **2003**, *50*, 181–187. [CrossRef]
12. Coan, J.; Allen, J. Frontal EEG asymmetry as a moderator and mediator of emotion. *Biol. Psychol.* **2004**, *67*, 7–50. [CrossRef]
13. Petrantonakis, P.C.; Hadjileontiadis, L.J. A Novel Emotion Elicitation Index Using Frontal Brain Asymmetry for Enhanced EEG-Based Emotion Recognition. *IEEE Trans. Inf. Technol. Biomed.* **2011**, *15*, 737–746. [CrossRef]
14. Li, X.; Hu, B.; Zhu, T.; Yan, J.; Zheng, F. Towards affective learning with an EEG feedback approach. In Proceedings of the 1st ACM International Workshop on Multimedia Technologies for Distance Learning, Beijing, China, 23 October 2009; pp. 33–38.
15. Ekman, P.E.; Davidson, R.J. *The Nature of Emotion: Fundamental Questions*; Oxford University Press: New York, NY, USA, 1994.
16. Plutchik, R. A psycho evolutionary theory of emotion. *Soc. Sci. Inf.* **1980**, *21*, 529–553. [CrossRef]

17. Alarcao, S.M.; Fonseca, M.J. Emotions Recognition Using EEG Signals: A Survey. *IEEE Trans. Affect. Comput.* **2017**, *10*, 374–393. [CrossRef]

18. Kroupi, E.; Vesin, J.-M.; Ebrahimi, T. Subject-Independent Odor Pleasantness Classification Using Brain and Peripheral Signals. *IEEE Trans. Affect. Comput.* **2015**, *7*, 422–434. [CrossRef]

19. Bahari, F.; Janghorbani, A. EEG-based emotion recognition using Recurrence Plot analysis and K nearest neighbor classifier. In Proceedings of the 2013 20th Iranian Conference on Biomedical Engineering (ICBME), Tehran, Iran, 18–20 December 2013; pp. 228–233. [CrossRef]

20. Zheng, W.-L.; Lu, B.-L. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Trans. Auton. Ment. Dev.* **2015**, *7*, 162–175. [CrossRef]

21. Tripathi, S.; Acharya, S.; Sharma, R.D.; Mittal, S.; Bhattacharya, S. Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4746–4752.

22. Zhang, D.; Yao, L.; Zhang, X.; Wang, S.; Chen, W.; Boots, R. Cascade and parallel convolutional recurrent neural networks on EEG-based intention recognition for brain-computer interface. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence, AAAI 2018, New Orleans, LA, USA, 2–7 February 2018.

23. Alhagry, S.; Fahmy, A.A.; El-Khoribi, R.A. Emotion recognition based on EEG using lstm recurrent neural network. *Emotion* **2017**, *8*, 355–358. [CrossRef]

24. Salama, E.S.; El-Khoribi, R.A.; Shoman, M.E.; Wahby, M.A. EEG-Based Emotion Recognition using 3D Convolutional Neural Networks. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 329–337. [CrossRef]

25. Song, T.; Zheng, W.; Song, P.; Cui, Z. EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks. *IEEE Trans. Affect. Comput.* **2018**, *11*, 532–541. [CrossRef]

26. Zhong, P.; Wang, D.; Miao, C. EEG-based emotion recognition using regularized graph neural networks. *IEEE Trans. Affect. Comput.* **2020**. [CrossRef]

27. Yin, Y.; Zheng, X.; Hu, B.; Zhang, Y.; Cui, X. EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM. *Appl. Soft Comput.* **2020**, *100*, 106954. [CrossRef]

28. Yang, Y.; Wu, Q.; Qiu, M.; Wang, Y.; Xiaowei, C. Emotion Recognition from Multi-Channel EEG through Parallel Convolutional Recurrent Neural Network. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN) 2018, Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–7.

29. Liu, J.; Wu, G.; Luo, Y.; Qiu, S.; Yang, S.; Li, W.; Bi, Y. EEG-Based Emotion Classification Using a Deep Neural Network and Sparse Autoencoder. *Front. Syst. Neurosci.* **2020**, *14*, 43. [CrossRef]

30. Zhang, Y.; Chen, J.; Tan, J.H.; Chen, Y.; Chen, Y.; Li, D.; Yang, L.; Su, J.; Huang, X.; Che, W. An investigation of deep learning models for EEG-based emotion recognition. *Front. Neurosci.* **2020**, *14*, 622759. [CrossRef]

31. Donmez, H.; Ozkurt, N. Emotion Classification from EEG Signals in Convolutional Neural Networks. In Proceedings of the 2019 Innovations in Intelligent Systems and Applications Conference (ASYU), Izmir, Turkey, 31 October–2 November 2019; pp. 1–6. [CrossRef]

32. Zulkifley, M.; Abdani, S.R. EEG signals classification by using convolutional neural networks. In Proceedings of the IEEE Symposium on Acoustics, Speech and Signal Processing (SASSP 2019), Kuala Lumpur, Malaysia, 20 March 2019.

33. Anwar, A.M.; Eldeib, A.M. EEG signal classification using convolutional neural networks on combined spatial and temporal dimensions for BCI systems. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 434–437.

34. Tatum, W.O. *Handbook of EEG Interpretation*; Demos Medical Publishing: New York, NY, USA, 2008.

35. Koelsch, S. Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* **2014**, *15*, 170–180. [CrossRef] [PubMed]

36. Teplan, M. Fundamentals of EEG measurement. *Meas. Sci. Technol.* **2002**, *2*, 1–11.

37. Niedermeyer, E.; da Silva, F. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2005.

38. Bos, D.O. EEG-Based Emotion Recognition: The Influence of Visual and Auditory Stimuli. Capita Selecta (MSc course). 2006.

39. Islam, M.R.; Moni, M.A.; Islam, M.M.; Rashed-Al-Mahfuz, M.; Islam, M.S.; Hasan, M.K.; Hossain, M.S.; Ahmad, M.; Uddin, S.; Azad, A.; et al. Emotion recognition from EEG signal focusing on deep learning and shallow learning techniques. *IEEE Access* **2021**, *9*, 94601–94624. [CrossRef]

40. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.-S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. DEAP: A Database for Emotion Analysis; Using Physiological Signals. *IEEE Trans. Affect. Comput.* **2011**, *3*, 18–31. [CrossRef]

41. Morris, J.D. SAM: The Self-Assessment Manikin and Efficient Cross-Cultural Measurement of Emotional Response. *J. Advert. Res.* **1995**, *35*, 63–68.

42. Zhang, J.; Wu, H.; Chen, W.; Wei, S.; Chen, H. Design and tool flow of a reconfigurable asynchronous neural network accelerator. *Tsinghua Sci. Technol.* **2021**, *26*, 565–573. [CrossRef]

43. Chao, H.; Dong, L.; Liu, Y.; Lu, B. Emotion Recognition from Multiband EEG Signals Using CapsNet. *Sensors* **2019**, *19*, 2212. [CrossRef]

44. Pandey, P.; Seeja, K. Subject independent emotion recognition from EEG using VMD and deep learning. *J. King Saud Univ. Comput. Inf. Sci.* **2019**, *34*, 1730–1738. [CrossRef]

45. Islam, R.; Ahmad, M. Virtual Image from EEG to Recognize Appropriate Emotion using Convolutional Neural Network. In Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 3–5 May 2019; pp. 1–4. [CrossRef]

46. Alazrai, R.; Homoud, R.; Alwanni, H.; Daoud, M.I. EEG-Based Emotion Recognition Using Quadratic Time-Frequency Distribution. *Sensors* **2018**, *18*, 2739. [CrossRef]

47. Liu, W.; Zheng, W.-L.; Lu, B. Emotion recognition using multimodal deep learning. In *Proceedings of The 23rd International Conference on Neural Information Processing (ICONIP 2016), Kyoto, Japan, 16–21 October 2016*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; Volume 9948, pp. 521–529. [CrossRef]

48. Cui, H.; Liu, A.; Zhang, X.; Chen, X.; Wang, K.; Chen, X. EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. *Knowl.-Based Syst.* **2020**, *205*, 106243. [CrossRef]

49. Vijayakumar, S.; Flynn, R.; Murray, N. A Comparative Study of Machine Learning Techniques for Emotion Recognition from Peripheral Physiological Signals. In Proceedings of the ISSC 2020. 31st Irish Signals and System Conference, Letterkenny, Ireland, 11–12 June 2020; pp. 1–21.

50. Li, M.; Xu, H.; Liu, X.; Lu, S. Emotion recognition from multichannel EEG signals using K-nearest neighbor classification. *Technol. Health Care* **2018**, *26*, 509–519. [CrossRef] [PubMed]

51. Luo, Y.; Lu, B.-L. EEG Data Augmentation for Emotion Recognition Using a Conditional Wasserstein GAN. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 2535–2538. [CrossRef]

52. Zhong, Y.; Jianhua, Z. Subject-generic EEG feature selection for emotion classification via transfer recursive feature elimination. In Proceedings of the 2017 36th Chinese Control Conference (CCC), Dalian, China, 26–28 July 2017; pp. 11005–11010. [CrossRef]

53. Menezes, M.L.R.; Samara, A.; Galway, L.; Sant'Anna, A.; Verikas, A.; Alonso-Fernandez, F.; Wang, H.; Bond, R. Towards emotion recognition for virtual environments: An evaluation of eeg features on benchmark dataset. *Pers. Ubiquitous Comput.* **2017**, *21*, 1003–1013. [CrossRef]

54. Zhang, Y.; Ji, X.; Zhang, S. An approach to EEG-based emotion recognition using combined feature extraction method. *Neurosci. Lett.* **2016**, *633*, 152–157. [CrossRef] [PubMed]

55. Kumar, N.; Khaund, K.; Hazarika, S.M. Bispectral Analysis of EEG for Emotion Recognition. *Procedia Comput. Sci.* **2016**, *84*, 31–35. [CrossRef]

56. Atkinson, J.; Campos, D. Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Syst. Appl.* **2016**, *47*, 35–41. [CrossRef]

57. Lan, Z.; Sourina, O.; Wang, L.; Liu, Y. Real-time EEG-based emotion monitoring using stable features. *Vis. Comput.* **2016**, *32*, 347–358. [CrossRef]

58. Mohammadi, Z.; Frounchi, J.; Amiri, M. Wavelet-based emotion recognition system using EEG signal. *Neural Comput. Appl.* **2016**, *28*, 1985–1990. [CrossRef]

59. Cheng, J.; Chen, M.; Li, C.; Liu, Y.; Song, R.; Liu, A.; Chen, X. Emotion Recognition From Multi-Channel EEG via Deep Forest. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 453–464. [CrossRef]

60. Wen, Z.; Xu, R.; Du, J. A novel convolutional neural network for emotion recognition based on EEG signal. In *Proceedings of the 2017 International Conference on Security, Pattern Analysis and Cybernetics (SPAC), Shenzhen, China, 15–17 December 2017*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2017; pp. 672–677.

61. Wang, Y.; Huang, Z.; McCane, B.; Neo, P. EmotioNet: A 3-D Convolutional Neural Network for EEG-based Emotion Recognition. In *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2018.

62. Gupta, V.; Chopda, M.D.; Pachori, R.B. Cross-Subject Emotion Recognition Using Flexible Analytic Wavelet Transform From EEG Signals. *IEEE Sens. J.* **2019**, *19*, 2266–2274. [CrossRef]

63. Zhang, Z.; Cong, X.; Feng, W.; Zhang, H.; Fu, G.; Chen, J. WAEAS: An optimization scheme of EAS scheduler for wearable applications. *Tsinghua Sci. Technol.* **2021**, *26*, 72–84. [CrossRef]