

Article

Human Emotion Recognition with Electroencephalographic Multidimensional Features by Hybrid Deep Neural Networks

Youjun Li ¹ , Jiajin Huang ¹, Haiyan Zhou ¹ and Ning Zhong ^{1,2,3,*}

¹ Institute of International WIC, Beijing University of Technology, Beijing 100124, China; lyj@ncut.edu.cn (Y.L.); hjj@emails.bjut.edu.cn (J.H.); zhouhaiyan@bjut.edu.cn (H.Z.)

² Beijing Advanced Innovation Center for Future Internet Technology, Beijing University of Technology, Beijing 100124, China

³ Knowledge Information Systems Lab, Department of Life Science and Informatics, Maebashi Institute of Technology, Maebashi 371-0816, Japan

* Correspondence: zhong@maebashi-it.ac.jp; Tel.: +81-27-265-7366

Received: 11 September 2017; Accepted: 11 October 2017; Published: 13 October 2017

Featured Application: The method presented in this study can be applied in many fields, such as mental health care, entertainment consumption behavior, society safety, and so on. For example, in the mental health care field, an automatic emotion analysis system can be constructed with our method to monitor the emotional variation of the subjects. With accurate and objective emotion analysis results from EEG signals, our method can provide useful treatment effect information to the medical staff.

Abstract: The aim of this study is to recognize human emotions by electroencephalographic (EEG) signals. The innovation of our research methods involves two aspects: First, we integrate the spatial characteristics, frequency domain, and temporal characteristics of the EEG signals, and map them to a two-dimensional image. With these images, we build a series of EEG Multidimensional Feature Image (EEG MFI) sequences to represent the emotion variation with EEG signals. Second, we construct a hybrid deep neural network to deal with the EEG MFI sequences to recognize human emotional states where the hybrid deep neural network combined the Convolution Neural Networks (CNN) and Long Short-Term-Memory (LSTM) Recurrent Neural Networks (RNN). Empirical research is carried out with the open-source dataset DEAP (a Dataset for Emotion Analysis using EEG, Physiological, and video signals) using our method, and the results demonstrate the significant improvements over current state-of-the-art approaches in this field. The average emotion classification accuracy of each subject with CLRNN (the hybrid neural networks that we proposed in this study) is 75.21%.

Keywords: emotion recognition; EEG signal; multidimensional features; hybrid neural networks; CNN; LSTM RNN

1. Introduction

Emotion is an important symbol of human intelligence; as such, an important intelligence symbol of artificial intelligence is that the machine can understand human emotions. As early as the 1980s, Minsky, one of the founders of artificial intelligence, proposed that a machine without emotions is not intelligent. Recently, research on human emotion recognition has been applied in many fields such as entertainment [1], safe driving [2,3], health care [4], social security [5], etc. Picard et al. [6] believed that the emotional changes of the human were embodied in speech [7], facial expressions [8], body posture [9], the central nervous system, autonomic nerve physiological activities [10], etc. Thus,

the study of human emotions through behavioral, facial, or physiological features has gradually become a focus of much attention. However, voice and facial expressions can be deliberately hidden by people on some social occasions. For this reason, researchers have tended to study human emotion through physiological signals such as electroencephalograms (EEG), electrooculography (EOG), temperature (TEM), blood volume pressure (BVP), electromyograms (EMG), and many other methods. Of all of these physiological signals, the EEG signal is of more interest to researchers as it comes directly from the human brain. Therefore, changes in EEG signals can directly reflect changes in human emotional states.

In this study, we recognize human emotional states with EEG signals. Two important aspects must be ensured during the emotion recognition process: (1) EEG feature extraction and expression and (2) emotion classifiers construction. For the first aspect, most of the previously used methods have only focused on the time and frequency dimension, and rarely combine the spatial dimension. Therefore, how to integrate and present the spatial features of the EEG signal with the time and frequency features is one key problem. For the second aspect, the key problem lies in how to construct a classifier to automatically learn the changes from the EEG multidimensional features over time and classify the changes into different emotion states. Corresponding to these two aspects, we mainly undertook the following work in this study:

- A new method is proposed to integrate the different EEG domain features. With the integration of multidimensional features, a sequence of two-dimensional images is constructed to express the variation in emotion.
- A hybrid deep learning neural network named CLRNN (Convolutional Networks (CNN) and Long-Short-Term-Memory Recurrent Neural Networks (RNN)) is built to undertake the recognition of human emotion from the EEG multidimensional feature image sequences.
- Empirical research is conducted with the open-source database DEAP [11] using our method, and the results demonstrate significant improvements over current state-of-the-art approaches in this field.

The rest of the paper is organized as follows: Related work is presented in Section 2. As data preparation, the methods of building EEG MFI and emotion labels are presented in Section 3. Next, we introduce the construction of CLRNN in Section 4. Section 5 describes the procedure of the experiment and reports the results. Finally, the conclusions and their discussion are detailed in Section 6.

2. Related Work

In this section, we review the related work on EEG features extraction and emotion classifying, respectively.

2.1. EEG Feature Extraction

We extend the study in [12] and review a wide range of EEG feature extraction methods proposed in the past 10 years. As seen in Table 1, most previous EEG feature extraction methods only focused on the time and frequency dimensions, and rarely combined them with spatial dimension information.

The time domain features study EEG signal through the variation of signal time series. The features include Hjorth features (Activity, Mobility, and Complexity [13,14]), statistics features (Power, Mean, Standard Deviation, etc. [15]), High Order Crossing features (HOC [16,17]), and so on. Time-domain features are not predominant. However, there are still many studies that have researched human emotion through time domain characteristics.

The frequency domain features study EEG signal by transforming the raw time domain EEG signal into frequency domain EEG signal with Fourier Transform method usually. The most popular features in the frequency domain are power features of different sub-frequency bands known as alpha, beta, theta, and delta. The most widely used algorithm is the Fast Fourier Transform (FFT),

which is applied in [18–24], and alternatives include Short-Time Fourier Transform (STFT) [25–28]. Another frequency feature is Power Spectra Density (PSD), which is usually estimated by Welch’s method [29].

Table 1. A summary of feature extraction for emotion recognition from EEG ¹.

Author and Study	Year	EEG Features	Extraction Method	Dimension
Ansari et al. [13]	2007	Activity, Mobility, and Complexity	Sevcik’s method	Time
Chanel et al. [25]	2007	9 sub-bands of the EEG (4–20 Hz)	STFT	Frequency
Horlings [14]	2008	Activity, Mobility, and Complexity	Welch’s Method	Time
Khalili and Moradi [18]	2008	Sub-band: θ , α , β , γ	FFT	Frequency
Li and Lu [19]	2009	EEG γ band (30–100 Hz)	FFT	Frequency
Petrantonakis and Hadjileontiadis [16,17]	2010	Higher Order Crossing	DWT	Time
Murugappan et al. [15,30]	2010	Power	DWT	Time
Nie et al. [26]	2011	Sub-band: δ , θ , α (8–13 Hz), β (1–30 Hz), γ (36–40 Hz)	STFT	Frequency
Kroupi et al. [31]	2011	Sub-band: θ , α , β , γ , NLD, NSI	Welch’s Method	Frequency
Liu and Sourina [20]	2012	β/α , Sub-band: β	FFT	Frequency
Hadjidimitriou et al. [32]	2012	HHS-based Feature Vectors	HHS	Time and Frequency
Reuderink et al. [33]	2013	The change and asymmetry in Sub-band of α	Welch’s Method	Frequency and Spatial
Rozgic et al. [21]	2013	Spectral Power and Spectral Power Differences	FFT	Frequency and Spatial
Lee and Hsieh [22]	2014	Correlation, Coherence, and Phase Synchronization	FFT	Frequency
Zheng et al. [27]	2014	PSD, DE, DASM and RASM	STFT	Frequency
Lahane and Sangaiah [34]	2015	Density Estimate	Kernel Density Estimation	Frequency
Paul et al. [35]	2015	Sub-band: α , β , θ	MF DFA	Frequency
Bashivan et al. [23]	2015	Sum of squared absolute values of the Sub-band: α , β , θ	FFT	Frequency Spatial
Thammasan et al. [29]	2016	Fractal Dimension (FD) and Power Spectral Density (PSD)	Welch’s Method	Frequency
Zheng et al. [28]	2016	PSD, DE, DASM, RASM, ASM, and DCAU	STFT	Frequency
Li et al. [36]	2017	Multi-scale entropy	HHT	Time and Frequency
Yin et al. [24]	2017	Frequency Features and Time-Frequency Features	FFT	Time and Frequency

¹ EEG, electroencephalographic; DE, density estimate; DWT, discrete wavelet transform; FFT, Fast Fourier transform; STFT, Short-time Fourier transform; HHS, Hilbert-Huang spectrum; PSD, power spectra density; ASM, asymmetry; DASM, differential asymmetry; RASM, rational asymmetry; DCAU, differential causality. MF DFA, multifractal detrended fluctuation analysis; NLD, normalized length density; NSI, non-stationarity index.

Since EEG signals are non-stationary, people proposed new methods combining time and frequency domain features to access additional information. The Hilbert–Huang Transform (HHT) is one method of studying EEG signals from both time and frequency domain. It decomposes the signal into Intrinsic Mode Functions (IMF) along with a trend, and obtains instantaneous frequency data. Hadjidimitriou et al. extracted HHS-based energy as the EEG features to study the music liking of the subjects [32]. They found that time–frequency features were more resistant to noise than the STFT-based features, which only extracted frequency features. Li et al. used HHT to improve the extraction of

multi-scale entropy as the EEG emotional features [36], and their results demonstrated that the time–frequency combined feature obtained better results than the traditional single-domain features.

EEG signals are obtained by measuring the electrical voltage signals of the multiple electrodes affixed to different positions on the scalp. From the obtaining method of the EEG signal, we can see that the information is highly correlated with the spatial, time, and frequency dimensions. However, seldom have previous studies paid attention to the spatial domain. The spatial information studies were limited to the asymmetry between the electrode pairs. The methods mostly calculate the differences in the power bands of the corresponding electrodes pairs on the left/right hemisphere of the scalp [21,37]. Recently, Bashivan et al. transformed EEG activities into a sequence of topology-preserving multi-spectral images to study human cognitive function [23], but few studies have analyzed human emotions with the spatial information of the EEG signals.

The method to integrate EEG multidimensional features is based on the spatial distribution of EEG electrodes (according to the 10–20 system [38]), and map the frequency domain characteristics to a two-dimensional image. With this method, we obtain a sequence of images from consecutive time windows from the EEG signal. The details of the construction method are presented in Section 3.

2.2. Emotion Classification Methods

In order to provide a comparison to our method, we chose studies that classified human emotions with scales of Valence and Arousal in Table 2. It also lists the classification accuracy and the number of subjects. As seen in Table 2, the most commonly used emotional classification methods include k-Nearest Neighbor (k-NN, used in [15,39]), Support Vector Machine (SVM, used in [14,40–42]), Random Decision Forest (RDF), Bayes Neural Networks (used in [43]) and Neural Networks (used in [44,45]). These methods are all used as baseline methods for comparison with our method, with details given in Section 5.2.

Table 2. Survey of the studies on emotion classification methods with EEG signal ¹.

Author and Study	Emotion Classification Basis	Subjects	Accuracy	Classification Method
Horlings [14]	Valence and Arousal (2 classes)	10	81%	SVM
Schaaff [41]	Valence and Arousal (3 classes)	30	66.7%	SVM
Frantzidis [40]	Valence and Arousal (2 classes, respectively)	28	81.3%	SVM
Murugappan [15]	Valence(2 classes)	12	71.3%	k-NN
Brown [39]	Valence (2 classes)	9	82%	SVM, k-NN
Hosseini [42]	Valence and Arousal (2 classes)	15	82%	SVM
Chung [43]	Valence and Arousal (2/3 classes respectively)	32	66.6%, 66.4% (2) 53.4%, 51.0% (3)	Bayes neural network
Li [44]	Valence and Arousal (2 classes, respectively)	32	74.12%	C-RNN
Our Method	Valence and Arousal (4 classes)	32	75.21%	CLRNN

¹ SVM, Support Vector Machine; CNN, Convolution Neural Networks; RNN, Recurrent Neural Networks; C-RNN, CNN+RNN; LSTM, Long Short-Term-Memory; CLRNN, CNN + LSTM RNN.

It is noteworthy that most of the methods listed in Table 2 classify emotions statically, except for the method used in [44] where the LSTM RNN was adopted to learn from the EEG features incrementally and dynamically. Another point worth noting is that only CNN is suitable for automatically extracting features from the image out of these methods. These two points are the reason for selecting CNN and LSTM RNN as parts of our classification method. The second column of Table 2 shows the classification basis and the number of the difference classes in the previous studies. As we can see, previous studies have basically divided emotions into categories two to three. In this study, we divided the emotion state into four classes. All the studies in Table 2 classify the emotion by Valence and Arousal. The third column of Table 2 shows the number of subjects included in the evaluated dataset.

3. Materials and Methods

The data preparation phase mainly included two aspects: the construction of EEG MFI sequences and the building of the emotion classification labels.

3.1. The Construction of EEG MFI Sequences

The International 10–20 System is an internationally recognized method of describing and applying the location of scalp electrodes in the context of an EEG test. The system is based on the relationship between the location of an electrode and the underlying area of the cerebral cortex. The “10” and “20” refer to the fact that the actual distances between the adjacent electrodes are either 10% or 20% of the total front–back or right–left distance of the skull [46].

Figure 1 shows a plan view of the International 10–20 System and a generalized square matrix from it. We can see that the left of Figure 1 is the International 10–20 System, where the EEG electrodes circled in red are the test points used in the DEAP dataset. In this study, we generalized the International 10–20 System with test electrodes used in the DEAP dataset to form a square matrix ($N \times N$), where N is the maximum point number between the horizontal or vertical test points. With the DEAP dataset, N equals 9. The square matrix without filling the EEG frequency features is represented at the right of Figure 1. The gray triangle above the center of the square matrix represents the nasion, while the red points are the electrodes corresponding to the red circles in the International 10–20 System. The gray points are added to form a fully matrix. The value of the red point corresponded to the frequency feature (PSD) of the EEG electrode. The value of the gray point is the interpolation of the red points surrounding it.

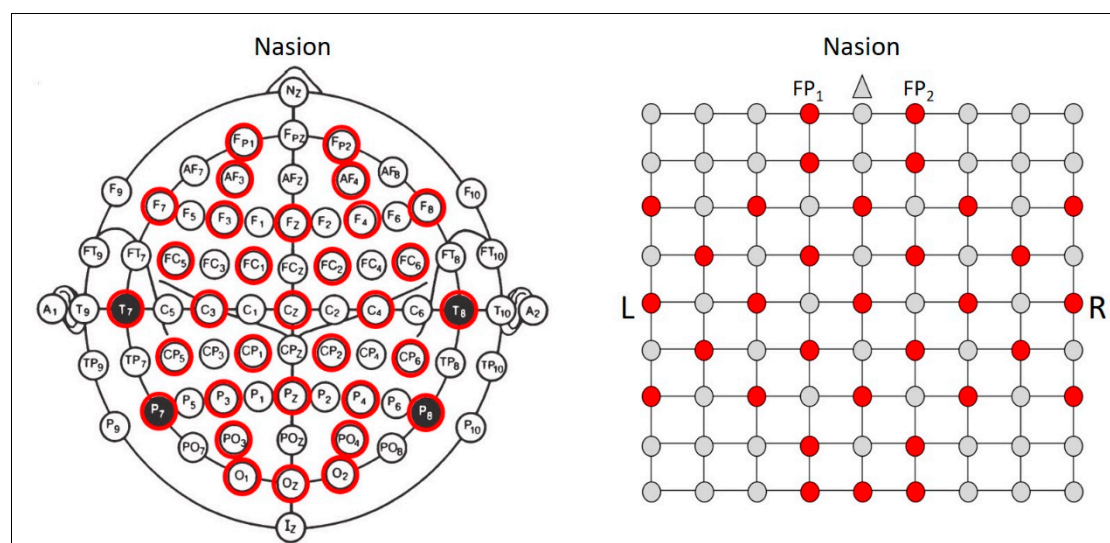


Figure 1. The International 10–20 System and the corresponding square EEG feature matrix (9×9) with tested EEG electrodes (the red points are tested in the trial and the gray points are not tested).

Figure 1 presents the method of mapping the International 10-20 System to a generalized EEG feature matrix. With this method, a single frame EEG MFI can be built from the EEG signal within a time window. With the time window moving forward, an EEG MFI sequence is constructed from the EEG signal. The process is presented in Figure 2. The definition of the red points and gray points is as same as it is defined in Figure 1. The different colors in EEG images represent the value of the EEG feature. The higher the feature is, the closer it is to the dark red. The lower the feature is, the closer it is to the dark blue. And the range of the EEG feature value is from 0 to 1.

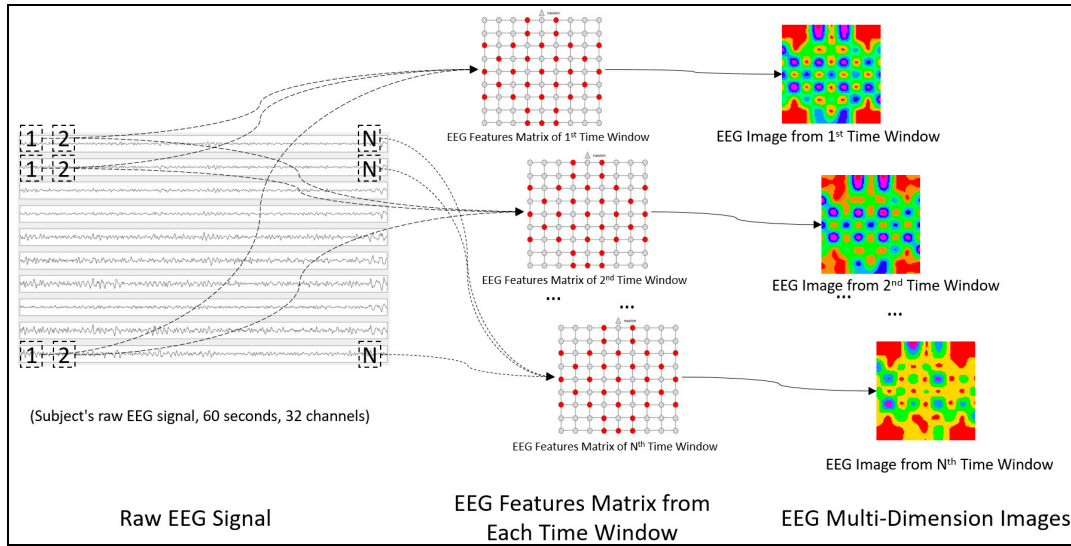


Figure 2. The construction process diagram of the electroencephalographic (EEG) Multidimensional Feature Image (MFI) sequence.

The EEG MFI sequence construction process consists of three steps. First, the raw EEG signals are extracted from DEAP, which included the multi-channel EEG signal of 32 subjects. Each subject has 40 trials where each trial includes the EEG signals of 32 channels, each signal lasting for 60 s. In the leftmost image of Figure 2, we schematically show the raw EEG signal of the first 10 channels. After that, the power spectrum density (PSD, [14,31,33,39,40]) is extracted as a EEG frequency domain feature from the raw signals. The PSD is estimated with Welch’s method in MATLAB (R2016a) using a Hamming window and different time window sizes (1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 30 and 60 s) with no overlap as parameters. A number of $(32 \text{ channels} \times 60 \text{ s}/\text{TI})$ features are obtained per trial, where TI is the size of the time window. Using a one-second time window as an example, 1920 (32×60) features are obtained from a raw EEG signal. After that, the features of each subject are normalized to reduce inter-participant variability by scaling between 0 and 1, as is shown in Equation (1):

$$F'_i = \frac{F_{\max_i} - F_i}{F_{\max_i} - F_{\min_i}}, \quad (1)$$

where F'_i is the normalized value of the feature; F_{\max_i} , F_{\min_i} are the maximum and minimum value of the internal subject features; and F_i is the i th value in the feature sequence. The red points in the feature matrix are directly filled with the normalized feature values. The values of the gray points are calculated with the surrounding point values, and can be expressed as Equation (2):

$$V_{(m,n)} = \frac{V'_{(m+1,n)} + V'_{(m-1,n)} + V'_{(m,n+1)} + V'_{(m,n-1)}}{K}, \quad (0 \leq m, n \leq 8, m, n \in N), \quad (2)$$

where V is the value of the gray point (corresponding to $P_{(m,n)}$); and V' is the value of the point surrounding $P_{(m,n)}$. If the index of the surrounding point exceeds the range of 0 and 8, then the value

is 0. K is the number of non-zero elements in the numerator, and the default value of K is 1. After the feature matrix is filled, it is used as a base table to generate EEG MFI through the interpolation method. We generate the EEG MFI in MATLAB (R2016a, MathWorks, Boston, MA, USA, 2016). The code of the interpolation function and MFI generation method is presented in Appendix A. Using this code, the EEG MFI is constructed and saved as .png images with a size of 200×200 pixels. An enlarged EEG MFI is shown in Figure 3. As seen in Figure 3, the frequency domain characteristics are mapped to a two-dimensional plane according to the spatial distribution of the EEG electrodes. This MFI corresponds to a five-second time window. It displays Subject 1's spatial PSD feature of the first-time window. The color legend explains the range of the normalized PSD and the variation. We can see from it that the higher the PSD value is, the closer it is to the dark red end. The lower the PSD value is, the closer it is to the dark blue end. The higher PSD value indicates that the EEG signal contains more energy and the corresponding brain area is more active. With this point, we can find in Figure 3 that the FP1 electrode is with the highest PSD value, and the lowest value appears at the FC₆ (FC₆ is a tested point in 10-20 system, you can find it in left image of Figure 1) electrode.

In this study, in order to find out which time window size is more appropriate for emotion recognition, we build EEG MFI sequence with different length time windows. To illustrate the process of forming a MFI based on different time windows, we formalize the raw feature matrix into a four-dimensional matrix:

$$\mathbf{P}_{(electrode, sequence, trial, subject)}, \quad (3)$$

where the size of the matrix is $(32 \times 60 \times 40 \times 32)$. With the different time windows, it is possible to produce a different number of EEG MFIs. Assuming the number of EEG MFIs is N and the length of the time window is t , N equals sequence/ t . The pseudo-code of producing the specific feature matrix is expressed in Appendix B.

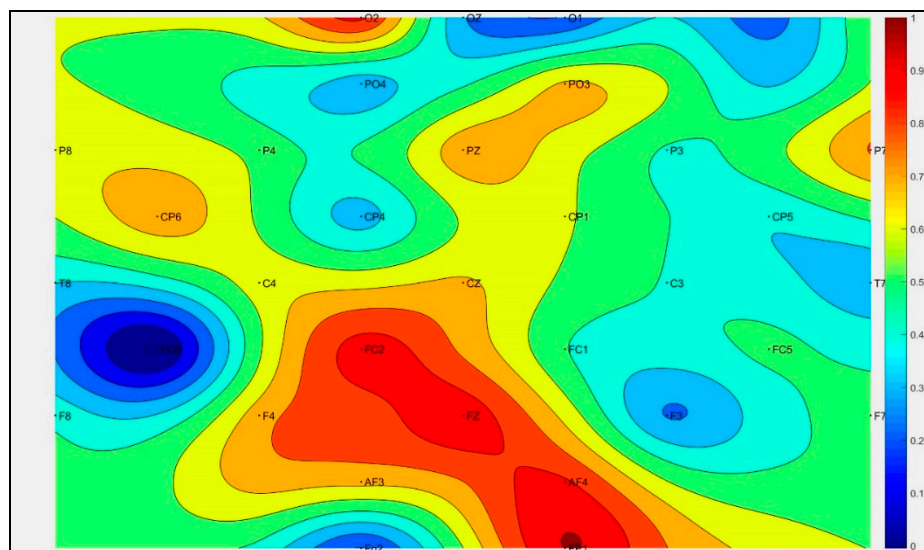


Figure 3. An enlarged EEG MFI with the names of the electrodes and contour lines.

Figure 4 displays the first five MFIs of Subject 1 with different time windows. Each row represents the MFIs with the same time window, and each line represents the MFIs with the same sequence order. Taking the first and the second row in Figure 4 as example, we can see that the first row represents the EEG variation over five seconds with five frames; however, the second row represents the same time variation with two frames. $MFI_{(1,5)}$ and $MFI_{(2,3)}$ are very similar. Accordingly, we can infer that the MFI sequence with a short time window provides more details about the variation than the MFI sequence with a long-time window. The meaning of the color in Figure 4 is the same as it in Figure 3.

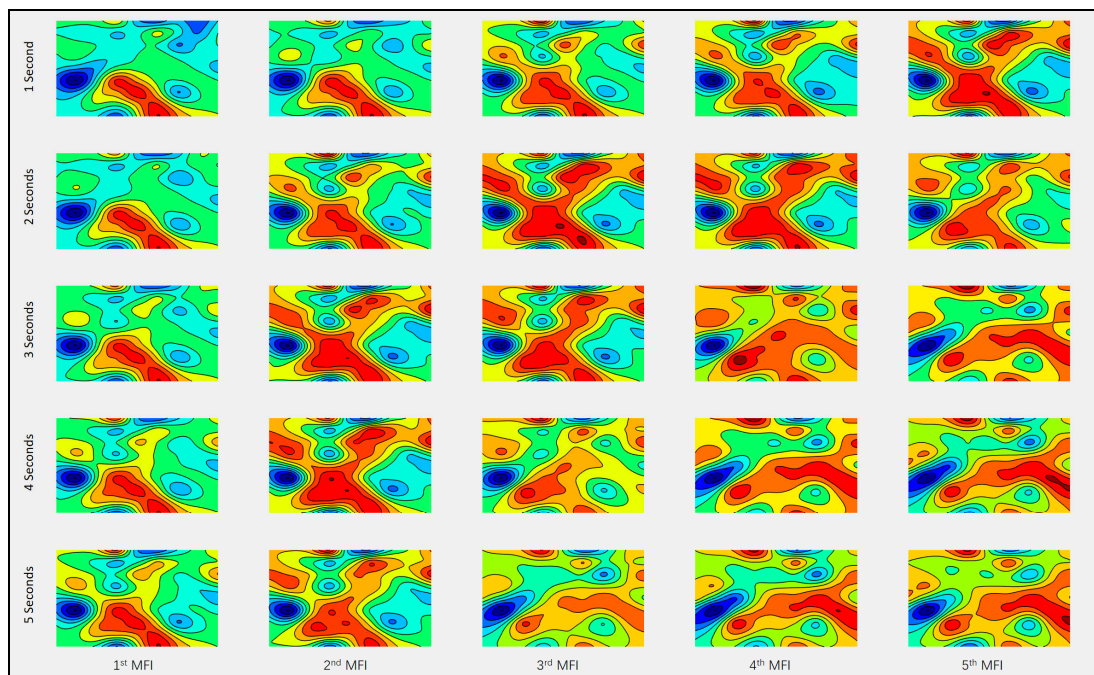


Figure 4. The MFI sequence of Subject 1 with different time windows.

3.2. The Construction of the Emotion Classification Labels

The classification method adopted in this paper is a supervisory machine learning method. Therefore, the corresponding classification labels of the EEG signal also need to be prepared in advance. The DEAP dataset contains the emotional evaluation values (including Valence, Arousal, Dominance, Like, and Familiarity) for the trials. In this paper, Valence and Arousal are extracted as emotional evaluation criteria to generate emotional labels. According to the different levels of Valence and Arousal, we divided the emotional two-dimensional plane into four quadrants. They are High Valence High Arousal (HVHA), High Valence Low Arousal (HVLA), Low Valence Low Arousal (LVLA), and Low Valence High Arousal (LVHA). Each quadrant corresponds to an emotion classification, as shown in Figure 5. According to the positive or negative deviation of the Valence and Arousal, we mapped each trial into the four quadrants to form an emotional classification label.

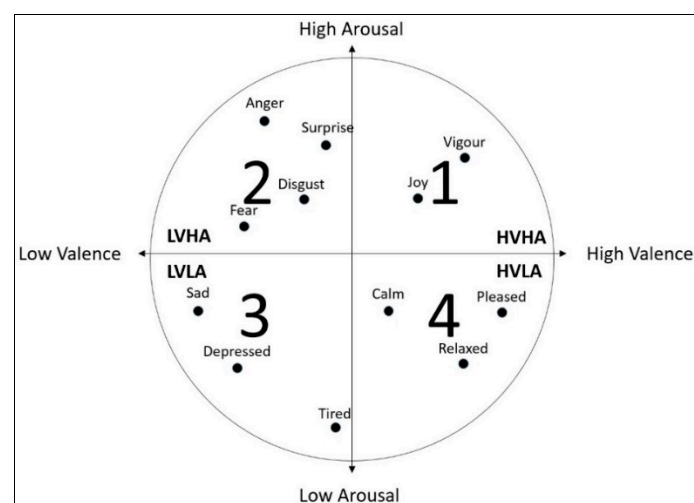


Figure 5. The Valence–Arousal dimension model of human emotion.

Table 3 shows the number of the different emotional samples mapped into the four quadrants. The number of samples contained in different emotional types is basically balanced, which ensured the balance of the neural network classification training.

Table 3. The number of samples in different emotion classifications ¹.

Emotion Labels	Number of the Samples
HVHA	348
HVLA	298
LVLA	282
LVHA	352
Total	1280

¹ HVHA, High Valence High Arousal; HVLA, High Valence Low Arousal; LVLA, Low Valence Low Arousal; LVHA, Low Valence High Arousal.

4. The Construction of the Hybrid Deep Neural Networks

We propose a hybrid deep learning model called Convolutional and LSTM Recurrent Neural Networks (CLRNN) to conduct emotion recognition tasks. This model is a composite of two kinds of deep learning structures: CNN and the LSTM recurrent neural network (LSTM RNN). The structure of the model is presented in Figure 6. The CNN is used to extract features from EEG MFI, and the LSTM RNN is used for modeling the context information of the long-term EEG MFI sequences. The features automatically extracted by the CNN reflect the spatial distribution of the EEG signals. In this work, two stacked convolutional layers are adopted as the basic structure of the CNN, which included two convolution layers, two max pooling layers, and a full connection layer. Given the dynamic nature of the EEG data, the LSTM RNN is a reasonable choice for modeling the emotion classification. Before connecting to the LSTM unit, a flattening operation is adopted to transform the final feature maps into a one-dimensional vector.

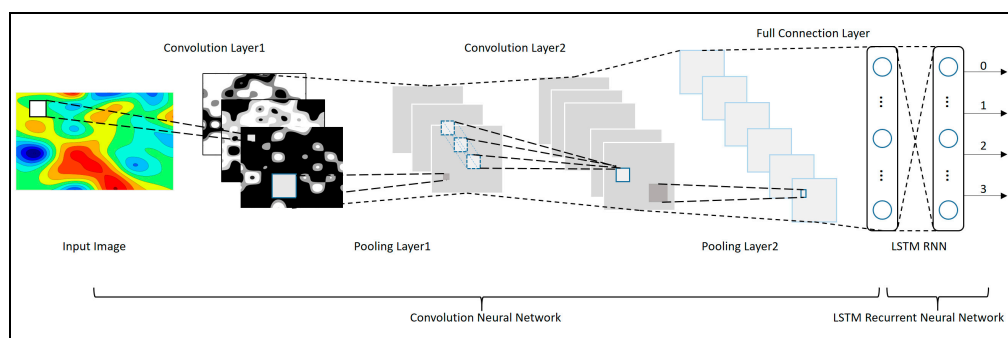


Figure 6. The structure of the hybrid deep neural networks used for emotion classification.

4.1. The Construction of Convolutional Neural Networks

The inputting MFI size of the networks is 200×200 pixels, and it contains three color channels. We set the number of convolutional filters as 30 in the first convolutional layer to extract 30 different kinds of correlation information, namely 30 different features. At the same time, to extract the multiple scale spatial characteristics of MFI, we use different size receptive fields in the first convolutional layer. The field sizes are 2×2 pixels, 5×5 pixels and 10×10 pixels, respectively. Corresponding to the different sizes of the field, the strides are 2, 5 and 10 pixels, respectively, without overlap between the strides. The activation function is ReLU. Following the first convolutional layer is a max pooling layer with pooling size of 2×2 , and the strides are 2. The second convolutional layer is set as 10 different filters with a size of 2×2 without overlap between strides. This setting helps to further fuse the information of a specific scale range from the prior features. Like the first convolutional layer, we add

a max pooling stage after this convolutional layer for information aggregation. Before connecting to the LSTM unit, a flatten operation is adopted to transform the final features into a one-dimensional feature vector. The configuration of the CNN described above is presented in Table 4. The dense layer in Table 4 is the layer that transforms the final features into a one-dimensional feature vector. In this layer, we set the output at 1/10 of the input to further compress the features and simplify the network. The LSTM RNN layer achieves a full connection to the dense layer. Next, the RNN output layer took ‘softmax’ as its activation function, and the output size is set to 4, corresponding to the four types of emotion states.

Table 4. The configurations of CNN. The parameters are denoted as <input size/receptive field size/pooling size> \times <number of kernels/channels/out size>.

Input Data	Convolutional Layer 1	Max Pooling Layer 1	Convolutional Layer 2	Max Pooling Layer 2	Dense Layer	LSTM RNN	RNN Output
<200 \times 200> \times 3	<2 \times 2> \times 30	<2 \times 2>	<2 \times 2> \times 10	<2 \times 2>	6250:625	625:625	4
	<5 \times 5> \times 30	<2 \times 2>	<2 \times 2> \times 10	<2 \times 2>	4000:400	400:400	
	<10 \times 10> \times 30	<2 \times 2>	<2 \times 2> \times 10	<2 \times 2>	1000:100	100:100	

4.2. The Construction of LSTM Recurrent Neural Networks

In the DEAP experiment, the stimulus intensity changes over 60 s. The emotion scores by the Subjects are often based on the most exciting part of the entire video. Therefore, we needed to model the context information for long-term sequences. As mentioned before, RNN is good at sequential modeling. However, a simple RNN must face the challenge of ‘gradient vanish or explode’ in back propagation when its dependencies are too long [47]. LSTM units have been adopted to replace the simple units of a traditional RNN. LSTM units combine gate mechanisms in their structures so that the key features of the timing data are effectively maintained and transmitted during the long-period calculation. The gate is able to forget the used information and the self-loop structure allows the gradient to flow for long durations [48].

A typical structure of a LSTM unit is illustrated in Figure 7. For comparison, Figure 7 shows the structure of two neural network units. The upper left corner of the figure is a simple recurrent neural network unit, and the LSTM unit is below the graph. As seen in Figure 7, the simple RNN unit only contains the feedback from the output to the input. However, the LSTM unit contains three gate structures, i.e., input gate, forget gate, and output gate, which determine what information from the prior step should be forgotten and what information in the current time step should be added into the main data flow. f_i , f_o and f_g are the activation function of the input data, output data, and gate, respectively. In this study, they are all sigmoid functions.

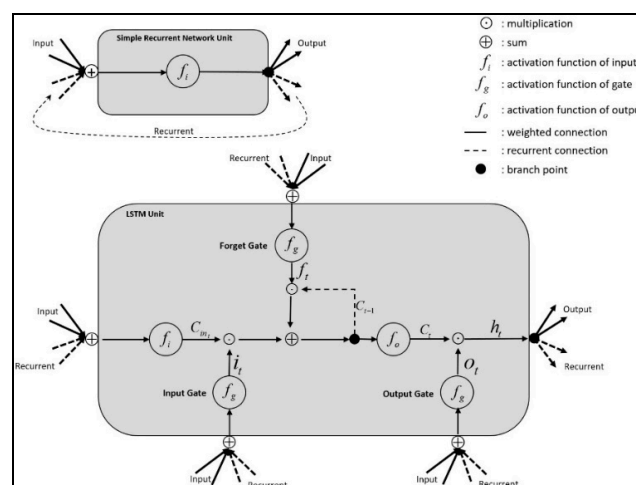


Figure 7. The Long Short-Term-Memory (LSTM) unit and simple recurrent network unit.

Different gates generate decision vectors to decide what candidate information will be selected. Using ‘Input Gate’ as an example, this generates vector \mathbf{i}_t with the hidden state \mathbf{h}_{t-1} from the prior LSTM cell and the current step’s input \mathbf{x}_t . The process of generating \mathbf{i}_t can be formalized as in Equation (4):

$$\mathbf{i}_t = f_g(\mathbf{w}_i \mathbf{x}_t + \mathbf{w}_i \mathbf{h}_{t-1} + \mathbf{b}_i), \quad (4)$$

where \mathbf{w}_i is the weighted matrix of the input function; and \mathbf{b}_i is the bias. The input candidate information $\tilde{\mathbf{C}}_t$ is also generated with \mathbf{h}_{t-1} and \mathbf{x}_t . $\tilde{\mathbf{C}}_t$ can be formalized as Equation (5):

$$\tilde{\mathbf{C}}_t = f_i(\mathbf{w}_c \mathbf{x}_t + \mathbf{w}_c \mathbf{h}_{t-1} + \mathbf{b}_c). \quad (5)$$

The final updating information is the multiplication of the candidate information by the decision vector $\tilde{\mathbf{C}}_t \times \mathbf{i}_t$. Another gate is the forget gate, which generates vector f_t to determine if the prior unit’s state \mathbf{C}_{t-1} should be reserved by multiplication $\mathbf{C}_{t-1} \times f_t$. The f_t can be formalized as Equation (6):

$$f_t = f_g(\mathbf{w}_f \mathbf{x}_t + \mathbf{w}_f \mathbf{h}_{t-1} + \mathbf{b}_f), \quad (6)$$

where f_t is scaled between 0 and 1 with the sigmoidal operation. The ‘0’ element causes the corresponding information in \mathbf{C}_{t-1} to be wiped out, while the ‘1’ means the corresponding information is allowed to pass.

The current unit state \mathbf{C}_t is a combination of \mathbf{C}_{t-1} and $\tilde{\mathbf{C}}_t$, and can be formalized as Equation (7):

$$\mathbf{C}_t = \mathbf{C}_{t-1} \times f_t + \tilde{\mathbf{C}}_t \times \mathbf{i}_t. \quad (7)$$

The output state of the LSTM unit is determined by the output gate. The output gate also generates a decision vector \mathbf{o}_t to decide the hidden state \mathbf{h}_t , and they can be formalized as Equations (8) and (9), as follows:

$$\mathbf{o}_t = f_g(\mathbf{w}_o \mathbf{x}_t + \mathbf{w}_o \mathbf{h}_{t-1} + \mathbf{b}_o) \quad (8)$$

$$\mathbf{h}_t = f_o(\mathbf{C}_t) \times \mathbf{o}_t. \quad (9)$$

In this study, the LSTM RNN is adopted to learn contextual information from the spatial features sequence extracted from the MFI.

4.3. The Construction of CLRNN with DL4J

DeepLearning4J is a java based toolkit for building, training and deploying Neural Networks [49]. In this study, DL4J is adopted as the framework to construct the CLRNN. We present the network’s configuration in Appendix C. The code in Appendix C is used to construct the network structure of the CLRNN. The size of the kernel in each layer is set according to the configuration given in Table 4. The setting of the learning rate for each layer changed in the tuning process of network training.

5. Results and Discussion

In this section, we present the process of the experiment and compare our method with the baselines to show the effectiveness of our methods.

5.1. Experiment Dataset and Settings

As mentioned earlier, we used the open dataset DEAP to verify the effectiveness of our method, which include EEG signals from 32 channels collected from 32 subjects. Each subject took 40 trials, and each trial lasted 60 s. The sampling frequency of the EEG signal was 512 Hz. With different time windows, we obtained EEG MFI sequences with a different number of EEG MFIs. For example, with a one-second time window, we obtained 2400 MFIs for one EEG MFI sequence. With a two-second time window, we obtained 1200 MFIs for one EEG MFI sequence. However, even with the shortest time

window, EEG MFIs are not enough for training a stable emotion recognition model with our method. For this reason, we adopt data augmentation strategies before training. We added “salt & pepper” noise to the MFIs in MATLAB with the command ‘imnoise()’. Image flipping or zooming is not used when augmenting the data. With this method, the original MFI set is expanded 20 times to ensure that we had at least 20,000 MFIs per subject for training. Sufficient training data helps a model with a large number of parameters to converge and generalize well. A five-fold cross-validation method is used to evaluate the performance of our approach, and the average performance of the 5-fold validation processes is taken as the experiment’s final results. We trained the model with different time windows to find out if the division of the EEG signal had an impact on the classification’s accuracy. The models are trained and tested in the Windows server environment, which included an Intel Xeon® V3 CPU (12×2.4 GHz) and 64 Gb RAM. No GPU acceleration is used in the experiment.

5.2. Baseline Methods

To illustrate its effectiveness, we compare our approach with the baseline methods and peer-reviewed studies. The selected baseline classifiers are commonly used in this field, including k-nearest neighbor (k-NN), Random Decision Forest, and Support Vector Machines (SVM). All baseline methods used a 5-fold cross-validation method for comparison with our method. The features trained in the baseline methods included the PSD, the C0 complexity, the power spectrum entropy, the Lyapunov index, and the correlation dimension. We trained the baseline classifiers in two ways: training in segments, and training in trials. The dimensions of each subject’s features matrix are (five kinds of features \times 32 channels) or (five kinds of features \times 60/length of the time window \times 32 channels). Principal component analysis is adopted to reduce the features’ dimensions. All training processes are tested in the MATLAB (R2016a) environment. Here, we briefly describe the details and parameter settings used in those methods.

k-NN: k-nearest neighbor algorithm (k-NN) is a non-parametric method used for classification. An object is classified by a majority vote of its neighbors. It is useful to assign weight to the contributions of the neighbors, so that nearer neighbors contribute more to the average than more distant ones. Therefore, the main parameters of the k-NN algorithm are the number of neighbors and the weighting scheme of giving each neighbor a weight. In this study, k is selected from the set ($k = \{5, 10, 15, 20, 25, 30\}$). Furthermore, the Chebyshev distance is adopted to calculate the distance between the object and the neighbors. The inverse of the distance gives the neighbor a weight, and the weight is used in the vote procedure.

Random Decision Forest: Random Decision Forest (RDF) is an ensemble learning method for classification, which constructs a multitude of decision trees at training time. The training algorithm applies bootstrap aggregating, or bagging, to the tree learners. It selects a random subset of the features during the learning process. The main parameter setting of RDF is the number of the learners. Here, the number of learners is selected from the set ($N = \{5, 10, 20, 30, 40, 50\}$).

SVM: SVM hyperparameters consisting of a regularization penalty parameter (C) and inverse of RBF kernel’s standard deviation ($\gamma = 1/\sigma$) were selected by grid search through cross-validation on the training set ($C = \{0.01, 0.1, 1, 10, 100\}$, $\gamma = \{0.1, 0.2, \dots, 1, 2, \dots, 10\}$). For the reason of multiple classes, the One-vs.-One strategy is employed during the SVM training.

CNN + RNN (without LSTM units): To show the memory effect of LSTM units in long-period data analysis, we designed a hybrid neural network structure including CNN and RNN without LSTM units. In this network structure, in addition to the RNN network layer not using the LSTM unit, the other network structure is the same as the CLRNN.

For peer-reviewed studies, we chose the studies listed in Table 2 for purposes of comparison.

5.3. Results and Discussion

In this section, we present the results of our experiments. Due to a variation of the parameters in the classification methods, we only present the best results obtained by each method. First,

a comparison of the classification accuracies between CLRNN and the baseline methods is presented in Figure 8. We present a boxplot of the mean emotion recognition accuracies with the different time windows for each subject in Figure 8. The comparison shows the effectiveness of our method. The average emotion classification accuracy of each subject with CLRNN is 75.21%, whereas the average accuracies of other classification methods are 69.58% with CNN + RNN, 67.45% with SVM, 45.47% with Random Decision Forest, and 62.84% with k-NN, respectively. The highest accuracy is obtained from Subject 4 with CLRNN, which is 90.54%.

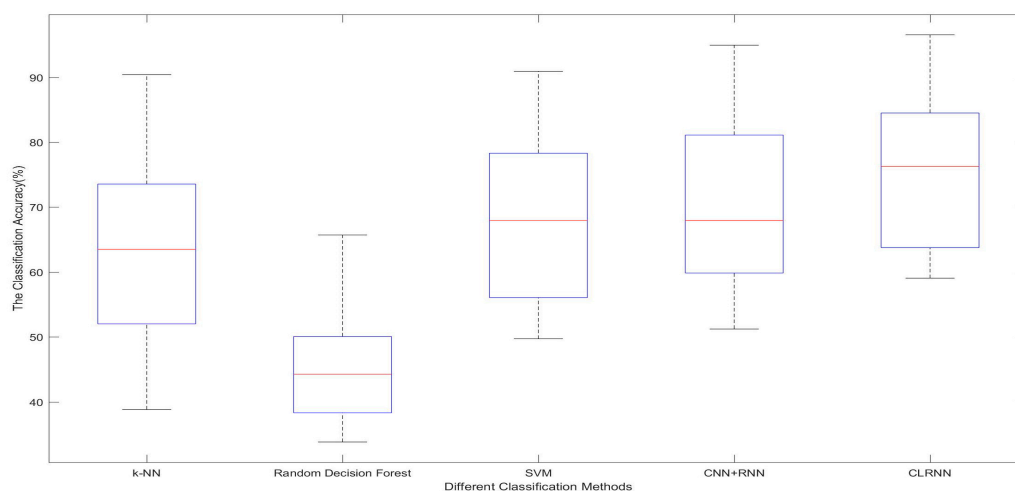


Figure 8. Emotion recognition accuracies with different classification methods.

After a comparison with the baseline methods, we chose relevant studies listed in Table 2 to compare with our method. The selection of the previous studies is based on two aspects: (1) the emotion analysis is based on EEG signals; and (2) the emotion label is produced by the scales of Valence and Arousal. We found that most studies in Table 2 classified emotions into two classes: Pleasant/Unpleasant or Positive/Negative. Some studies [7,43,44] classified emotion into three categories: Pleasant, Neutral, and Unpleasant. In our study, we classify emotion into four types (HVHA, HVLA, LVLA, and LVHA). Two emotion classification problems are relatively simple, and the highest accuracy reached 82%. Multiple (more than two) emotion classification problems are complex, and the accuracy of our method reaches 75.21%, which is higher than the results presented in [43,44]. The studies in [43,44] also employ DEAP as a dataset to recognize human emotions. This shows the effectiveness of our method. In addition, [41,43,44] and this study all employed DEAP as the dataset to undertake the emotion analysis. The performance of our method is better than the others. A similar research method is used in [41], which also built Neural Networks by CNNs and LSTM RNNs, with the difference being that the two-dimensional EEG feature images constructed in [41] ignored the spatial characteristics of EEG signals. In this paper, the spatial features of EEG signals are considered very important for emotion recognition. Through the experiments in this study, we proved the correctness of this point.

To further validate the effectiveness of our method, we investigated the effect of the time window size on the classification analysis. The MFIs with different time windows are trained and tested in CLRNN and CNN + RNN, respectively. For comparison, the features trained in the baseline method are also extracted from the raw EEG signals with the same time window size and are presented in Table 5, which shows the average of the emotional classification accuracy obtained by 32 subjects under different time windows. As seen from the results in Table 5, CLRNN showed sensitivity to the time window size. With an increase in the time window, the classification accuracy showed a decreasing trend. The accuracies of the classification from other methods did not change significantly with the

increase of the time window. This further confirmed that the LSTM unit played a role in capturing long-term critical features during the classification process.

Table 5. The classification results with different time window size and methods.

Classification Methods	1 s	2 s	3 s	4 s	5 s	6 s	8 s	10 s	12 s	15 s	20 s	30 s	60 s
k-NN	46.09	52.49	57.29	59.49	61.69	61.39	61.49	61.89	61.79	62.19	62.59	62.19	62.39
Random Decision Forest	39.03	38.17	39.38	40.19	38.53	36.43	44.53	45.38	45.88	46.58	46.68	46.78	46.98
SVM	65.11	65.11	66.11	64.21	63.01	64.31	61.01	63.01	62.61	62.41	63.01	63.41	63.21
CNN + RNN	61.76	59.02	58.07	60.34	61.43	62.13	61.13	59.1	60.1	60.9	61.1	61.5	62.1
CLRNN	74.73	75.21	75.13	74.32	73.25	70.37	67.23	65.01	57.3	60.2	62.1	60.6	61.8

The study in [45] analyzed the emotion classification with DEAP EEG signals from the perspective of time window size and wavelet features and obtained the highest accuracy using the wavelet entropy of three-second signal segments, which is similar to the results we obtained.

Further intuitive investigation is conducted with a graphical representation of the results from CLRNN and ‘CNN + RNN’, as illustrated in Figure 9. It can be seen from Figure 9 that with the same small time window (before 12 s) CLRNN had higher classification accuracies than CNN + RNN without LSTM units. After 12 s, the difference between the accuracies of the two methods is very small. For this phenomenon, our inference is that the change in the EEG signal presented with the MFIs is overshadowed as the time window becomes larger. Therefore, the MFI sequence corresponding to the large-sized time window does not reflect the change of emotion. To confirm this inference, we select MFI sequences from 32 subjects to seek corresponding evidence. After comparison, Subject 4 is chosen to present the variation, which is shown in Figure 10.

As seen in Figure 10, there are 12 MFI sequences, and each line corresponds to a time window size. From the first to the fifth line, each line contained the first 10 images of the MFI sequences. Starting from the sixth line, it contains the whole images of the sequence. Studies presented in [50,51] suggested that emotion is related to a group of structures in the center of the brain called the limbic system and other structures such as the prefrontal cortex [52], orbitofrontal cortex, and so on. Out of these areas, the correspondence between the prefrontal cortex and the EEG electrodes FP₁-FP₂ are more direct than others. Therefore, we focused on the area corresponding to FP₁-FP₂. We can see from Figure 10 that the MFI sequences corresponding to time windows 1 s to 3 s reveal more details about the activation in this area. However, starting from the MFI sequence corresponding to the 4 s time window, the activation information for this area is gradually reduced. This also corresponded to the case where emotion recognition accuracy decreased after a 4 s time window by CLRNN in Figure 9.

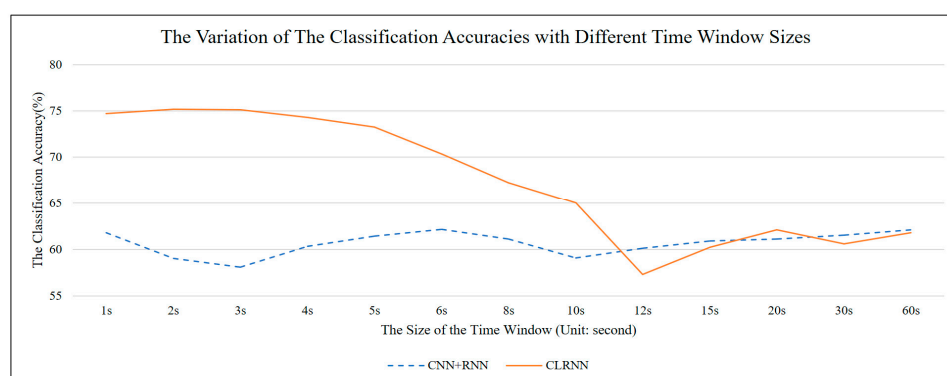


Figure 9. The comparison of the classification accuracies between CLRNN and CNN + RNN with different time window sizes.

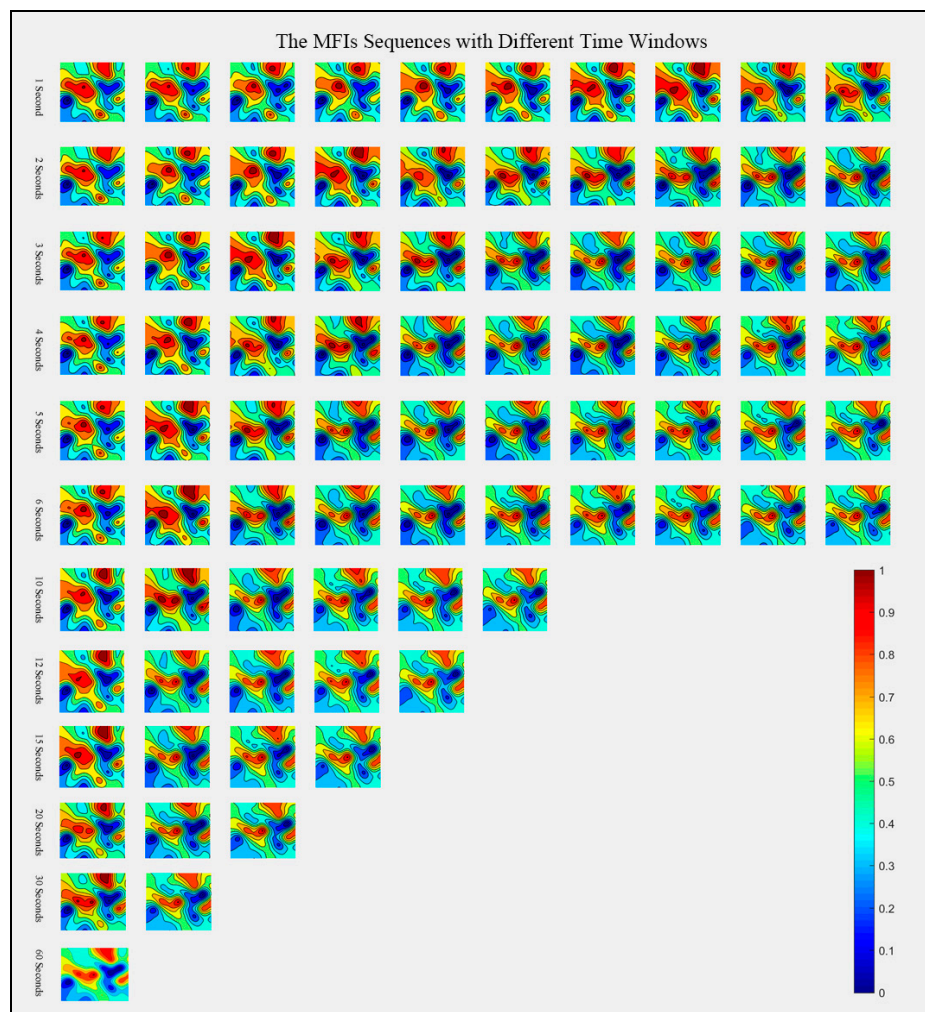


Figure 10. MFI sequences with different time windows corresponding to Subject 4 (corresponding to the emotion HVHA).

6. Conclusions

In this study, we try to improve the accuracy of classifying human emotion by EEG signals. The innovation of our methods involves two aspects. First, we propose a new method for the EEG feature extraction and representation. EEG frequency features (PSD) are extracted from different EEG channels and mapped to a two-dimensional plane to construct the EEG MFI. EEG MFI sequences are built from the raw EEG signal. The EEG MFI sequences fuse together the spatial, frequency domain, and time characteristics of the raw EEG signal. Another aspect is our proposal of a hybrid deep neural network that deals with the EEG MFI sequences and recognizes the emotions. The hybrid deep neural networks combined Convolution Neural Networks and Long-Short-Term-Memory Recurrent Neural Networks. In the hybrid structure, CNN is used to learn temporary image patterns from EEG MFI sequences, and LSTM RNN is used to classify human emotions.

With our method, empirical research is carried out with the DEAP dataset. We compare our results with those from the baseline methods and find that the emotion classification accuracy of our method reached 75.21%, which is higher than the accuracies from the baseline methods. In the baseline methods, we chose a ‘CNN + RNN’ Neural Network without LSTM unit to compare with our method. We find that the LSTM unit showed the time sensitivity. Furthermore, we reviewed the state of the art of human emotion recognition by EEG signals. Compared with similar studies, our study improves the classification accuracy.

Additionally, we analyzed the effects of different time windows on classification accuracy and found that time windows corresponding to two to three seconds achieved good classification accuracy, and the corresponding classification accuracy decreased from the time window division after four seconds. Given these results, we inferred that MFI sequences from a smaller time window represent more details of the variation of the EEG signal. We select Subject 4 to seek corresponding evidence in the MFI sequence and find that, with smaller size time windows, MFI reveals more details about activation in the FP_1 and FP_2 area.

Acknowledgments: This work is supported by the National Basic Research Program of China (2014CB744600), the National Natural Science Foundation of China (61420106005), and the International Science & Technology Cooperation Program of China (2013DFA32180).

Author Contributions: Youjun Li and Ning Zhong proposed the method of the construction of EEG MFI sequences; Youjun Li and Jiajin Huang proposed the framework of the hybrid neural networks; Haiyan Zhou contributed to the EEG feature extraction and analysis method; Youjun Li and Ning Zhong designed and performed the experiment; Youjun Li wrote the paper.

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

Appendix A

The code of the interpolation function and MFI generation method (used in MATLAB (R2016a)).

```
function build_EEG_MFI()
    % the feature matrix has been generated and stored in variable basic_matrix (81,3);
    x = basic_matrix(:,1); %Horizontal axis coordinates
    y = basic_matrix(:,2); % Vertical axis coordinates
    z = basic_matrix(:,3); % Feature values of the corresponding electrodes
    nx = linspace(min(x), max(x),1000);
    ny = linspace(min(y), max(y),1000);
    [xx,yy] = meshgrid(nx,ny);
    zz = griddata(x,y,z,xx,yy,'v4');
    contourf(yy,xx,zz,'linestyle','-', 'LineWidth',0.5);
    colormap('HSV');
    axis off;
    set(gcf,'PaperUnits','inches','PaperPosition',[0 0 2 2]);
    print(1,'-dpng','picWritePath','-r100');
end
```

Appendix B

The pseudo-code of producing the specific feature matrix (used in MATLAB (R2016a)).

```
function buildFeatureMatrix(integer theLengthofTheTimeWindow)
    % P is a feature Matrix of four-dimensional: P(electrodeNum,sequenceNum,trialNum,subjectNum)
    %theLengthofTheTimeWindow is set to 1,2,3,4,5, ... , 10;
    the_Num_of_MFIs=sequenceNum/theLengthofTheTimeWindow;
    for subjectNum=1:32
        for trialNum=1:40
            specific_P = P(:, :, subjectNum, trialNum);
            for r1=1:electrodeNum
                for r2=1:the_Num_of_MFIs
                    meanRawData(r1,r2)=mean(specific_P(r1,(r2-1)*theLengthofTheTimeWindow+
                        1:(r2-1)*theLengthofTheTimeWindow+theLengthofTheTimeWindow));
                end
            end
            generate_MFIs(meanRawData);
        end
    end
end
```

Appendix C

The code to construct the network structure of the CLRNN (built with dl4j).

```

Updater updater = Updater.ADAGRAD; // ADAGRAD function is taken as the updater
MultiLayerConfiguration conf = new NeuralNetConfiguration.Builder()
    .seed(12345)
    .regularization(true).l2(0.001) //l2 regularization on all layers
    .optimizationAlgo(OptimizationAlgorithm.STOCHASTIC_GRADIENT_DESCENT)
    .iterations(1)
    .learningRate(0.04)
    .list()
    .layer(0, new ConvolutionLayer.Builder(2, 2)
        .nIn(3) //3 channels: RGB
        .nOut(30)
        .stride(2, 2)
        .activation("relu")
        .weightInit(WeightInit.RELU)
        .updater(updater)
        .build()) //Output: (200-2+0)/2+1 = 100 -> 100*100*30
    .layer(1, new SubsamplingLayer.Builder(SubsamplingLayer.PoolingType.MAX)
        .kernelSize(2, 2)
        .stride(2, 2).build()) //Output:(100-2+0)/2+1 = 50
    .layer(2, new ConvolutionLayer.Builder(2, 2)
        .nIn(30)
        .nOut(10)
        .stride(2, 2)
        .activation("relu")
        .weightInit(WeightInit.RELU)
        .updater(updater)
        .build()) //Output: (50-2+0)/2+1 = 25 -> 25*25*10 = 6250
    .layer(3, new DenseLayer.Builder()
        .activation("relu")
        .nIn(6250)
        .nOut(100)
        .weightInit(WeightInit.RELU)
        .updater(updater)
        .gradientNormalization(GradientNormalization.ClipElementWiseAbsoluteValue)
        .gradientNormalizationThreshold(10)
        .learningRate(0.01)
        .build())
    .layer(4, new GravesLSTM.Builder()
        .activation("softsign")
        .nIn(100)
        .nOut(100)
        .weightInit(WeightInit.XAVIER)
        .updater(updater)
        .gradientNormalization(GradientNormalization.ClipElementWiseAbsoluteValue)
        .gradientNormalizationThreshold(10)
        .learningRate(0.001)
        .build())
    .layer(5, new RnnOutputLayer.Builder(LossFunctions.LossFunction.MCXENT)
        .activation("softmax")
        .nIn(100)
        .nOut(4) //4 possible emotion states
        .updater(updater)
        .weightInit(WeightInit.XAVIER)
        .gradientNormalization(GradientNormalization.ClipElementWiseAbsoluteValue)
        .gradientNormalizationThreshold(10)
        .build())

```

```

.inputPreProcessor(0, new RnnToCnnPreProcessor(200, 200, 2))
.inputPreProcessor(3, new CnnToFeedForwardPreProcessor(50, 50, 10))
.inputPreProcessor(4, new FeedForwardToRnnPreProcessor())
.pretrain(false).backprop(true)
.backpropType(BackpropType.TruncatedBPTT)
.tBPTTForwardLength(60/5)
.tBPTTBackwardLength(60/5)
.build();

```

References

1. Mandryk, R.L.; Inkpen, K.M.; Calvert, T.W. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behav. Inf. Technol.* **2006**, *25*, 141–158. [[CrossRef](#)]
2. Healey, J.A.; Picard, R.W. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* **2005**, *6*, 156–166. [[CrossRef](#)]
3. Katsis, C.D.; Katertsidis, N.; Ganiatsas, G.; Fotiadis, D.I. Toward emotion recognition in car-racing drivers: A biosignal processing approach. *IEEE Trans. Syst. Man Cybern.* **2008**, *38*, 502–512. [[CrossRef](#)]
4. Katsis, C.D.; Katertsidis, N.S.; Fotiadis, D.I. An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders. *Biomed. Signal Process. Control* **2011**, *6*, 261–268. [[CrossRef](#)]
5. Verschuere, B.; Ben-Shakhar, G.; Meijer, E. Memory Detection: Theory and Application of the Concealed Information Test. In *Psychopathy and the Detection of Concealed Information*; Verschuere, B., Ben-Shakhar, G.M., Meijer, E., Eds.; Cambridge University Press: Cambridge, UK, 2011; pp. 215–230.
6. Picard, R.W.; Vyzas, E.; Healey, J. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 1175–1191. [[CrossRef](#)]
7. El Ayadi, M.; Kamel, M.S.; Karray, F. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognit.* **2011**, *44*, 572–587. [[CrossRef](#)]
8. Venkatesh, Y.V.; Kassim, A.A.; Yuan, J.; Nguyen, T.D. On the simultaneous recognition of identity and expression from BU-3DFE datasets. *Pattern Recognit. Lett.* **2012**, *33*, 1785–1793. [[CrossRef](#)]
9. Arnrich, B.; Setz, C.; La Marca, R.; Troster, G.; Ehlert, U. What does your chair know about your stress level? *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 207–214. [[CrossRef](#)] [[PubMed](#)]
10. Cacioppo, J.T.; Berntson, G.G.; Larsen, J.T.; Poehlmann, K.M. The psychophysiology of emotion. In *Handbook of Emotion*; Lewis, M., Haviland-Jones, J.M., Eds.; Guilford Press: New York, NY, USA, 2000; pp. 173–191.
11. Koelstra, S.; Muhl, C.; Soleymani, M.; Jong-Seok, L.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. Deap: A database for emotion analysis using physiological signals. *IEEE Trans. Affect. Comput.* **2012**, *3*, 18–31. [[CrossRef](#)]
12. Kim, M.-K.; Kim, M.; Oh, E.; Kim, S.-P. A review on the computational methods for emotional state estimation from the human EEG. *Comput. Math. Methods Med.* **2013**, *2013*, 1–13. [[CrossRef](#)] [[PubMed](#)]
13. Ansari-Asl, K.; Chanel, G.; Pun, T. A channel selection method for EEG classification in emotion assessment based on synchronization likelihood. In Proceedings of the 15th European Signal Processing Conference, Poznan, Poland, 3–7 September 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 1241–1245.
14. Horlings, R.; Datcu, D.; Rothkrantz, L.J.M. Emotion recognition using brain activity. In Proceedings of the 9th international conference on computer systems and technologies and workshop for PhD students in computing, Gabrovo, Bulgaria, 12–13 June 2008; ACM: New York, NY, USA, 2008.
15. Murugappan, M.; Ramachandran, N.; Sazali, Y. Classification of human emotion from EEG using discrete wavelet transform. *J. Biomed. Sci. Eng.* **2010**, *3*, 390–396. [[CrossRef](#)]
16. Petrantonakis, P.C.; Hadjileontiadis, L.J. Emotion recognition from EEG using higher order crossings. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 186–197. [[CrossRef](#)] [[PubMed](#)]
17. Petrantonakis, P.C.; Hadjileontiadis, L.J. Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Trans. Affect. Comput.* **2010**, *1*, 81–97. [[CrossRef](#)]
18. Khalili, Z.; Moradi, M.H. Emotion detection using brain and peripheral signals. In Proceedings of the Biomedical Engineering Conference, Cairo, Egypt, 18–20 December 2008; IEEE: Piscataway, NJ, USA, 2009; pp. 1223–1226.

19. Mu, L.; Lu, B.-L. Emotion classification based on gamma-band EEG. In Proceedings of the Annual International Conference of the IEEE, Minneapolis, MN, USA, 3–6 September 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1223–1226.
20. Liu, Y.; Sourina, O. EEG-based dominance level recognition for emotion-enabled interaction. In Proceedings of the IEEE International Conference on Multimedia and Expo, Melbourne, Australia, 9–13 July 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1039–1044.
21. Rozgic, V.; Vitaladevuni, S.N.; Prasad, R. Robust EEG emotion classification using segment level decision fusion. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 1286–1290.
22. Daunizeau, J.; Lee, Y.-Y.; Hsieh, S. Classifying different emotional states by means of EEG-based functional connectivity patterns. *PLoS ONE* **2014**, *9*, e95415.
23. Bashivan, P.; Rish, I.; Yeasin, M.; Codella, N. Learning representations from EEG with deep recurrent-convolutional neural networks. In Proceedings of the International Conference on Learning Representations 2016, San Juan, PR, USA, 2–4 May 2016.
24. Yin, Z.; Wang, Y.; Liu, L.; Zhang, W.; Zhang, J. Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination. *Front. Neurobot.* **2017**, *11*, 19. [[CrossRef](#)] [[PubMed](#)]
25. Chanel, G.; Karim, A.-A.; Thierry, P. Valence-arousal evaluation using physiological signals in an emotion recall paradigm. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Montreal, QC, Canada, 7–10 October 2007; IEEE: Piscataway, NJ, USA, 2008; pp. 2662–2667.
26. Nie, D.; Wang, X.-W.; Shi, L.-C.; Lu, B.-L. EEG-based emotion recognition during watching movies. In Proceedings of the 5th International IEEE/EMBS Conference on Neural Engineering, Cancun, Mexico, 27 April–1 May 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 667–670.
27. Zheng, W.-L.; Dong, B.-N.; Lu, B.-L. Multimodal emotion recognition using EEG and eye tracking data. In Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 5040–5043.
28. Zheng, W.-L.; Zhu, J.-Y.; Lu, B.-L. Identifying stable patterns over time for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* **2017**, *PP*, 1. [[CrossRef](#)]
29. Thammasan, N.; Moriyama, K.; Fukui, K.; Numao, M. Continuous music-emotion recognition based on electroencephalogram. *IEICE Trans. Inf. Syst.* **2016**, *99*, 1234–1241. [[CrossRef](#)]
30. Murugappan, M.; Rizon, M.; Nagarajan, R.; Yaacob, S. Inferring of human emotional states using multichannel EEG. *Eur. J. Sci. Res.* **2010**, *48*, 281–299.
31. Kroupi, E.; Yazdani, A.; Ebrahimi, T. EEG correlates of different emotional states elicited during watching music videos. In Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction, Memphis, TN, USA, 9–12 October 2011; Affective Computing and Intelligent Interaction (ACII): Berlin, Germany, 2011; pp. 457–466.
32. Hadjilimitriou, S.K.; Hadjileontiadis, L.J. Toward an EEG-based recognition of music liking using time-frequency analysis. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 3498–3510. [[CrossRef](#)] [[PubMed](#)]
33. Reuderink, B.; Mühl, C.; Poel, M. Valence, arousal and dominance in the EEG during game play. *Int. J. Auton. Adapt. Commun. Syst.* **2013**, *6*, 45–62. [[CrossRef](#)]
34. Lahane, P.; Sangaiah, A.K. An approach to EEG based emotion recognition and classification using kernel density estimation. *Procedia Comput. Sci.* **2015**, *48*, 574–581. [[CrossRef](#)]
35. Paul, S.; Mazumder, A.; Ghosh, P.; Tibarewala, D.N.; Vimalarani, G. EEG based emotion recognition system using MFDFA as feature extractor. In Proceedings of the International Conference on Robotics, Automation, Control and Embedded Systems (RACE), Chennai, India, 18–20 February 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–5.
36. Li, X.; Qi, X.Y.; Sun, X.Q.; Xie, J.L.; Fan, M.D.; Kang, J.N. An improved multi-scale entropy algorithm in emotion EEG features extraction. *J. Med. Imaging Health Inform.* **2017**, *7*, 436–439.
37. Soleymani, M.; Koelstra, S.; Patras, I.; Pun, T. Continuous emotion detection in response to music videos. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition and Workshops, Santa Barbara, CA, USA, 21–25 March 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 803–808.
38. Klem, G.H.; Luders, H.O.; Jasper, H.H.; Elger, C. The ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol. Suppl.* **1999**, *52*, 3–6. [[PubMed](#)]

39. Brown, L.; Grundlehner, B.; Penders, J. Towards wireless emotional valence detection from EEG. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, USA, 30 August–3 September 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 2188–2191.
40. Frantzidis, C.A.; Bratsas, C.; Papadelis, C.L.; Konstantinidis, E.; Pappas, C.; Bamidis, P.D. Toward emotion aware computing: An integrated approach using multichannel neurophysiological recordings and affective visual stimuli. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 589–597. [[CrossRef](#)] [[PubMed](#)]
41. Schaaff, K.; Schultz, T. Towards emotion recognition from electroencephalographic signals. In Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, Amsterdam, The Netherlands, 10–12 September 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1–6.
42. Hosseini, S.A.; Khalilzadeh, M.A.; Naghibi-Sistani, M.B.; Niazmand, V. Higher order spectra analysis of EEG signals in emotional stress states. In Proceedings of the Second International Conference on Information Technology and Computer Science, Kiev, Ukraine, 24–25 July 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 60–63.
43. Chung, S.Y.; Yoon, H.J. Affective classification using Bayesian classifier and supervised learning. In Proceedings of the 12th International Conference on Control, Automation and Systems, Jeju Island, Korea, 17–21 October 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1768–1771.
44. 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 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Shenzhen, China, 15–18 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 352–359.
45. Candra, H.; Yuwono, M.; Rifai, C.; Handojoseno, A.; Elamvazuthi, I.; Nguyen, H.T.; Su, S. Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine. In Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 7250–7253.
46. 10–20 System (EEG). Available online: [https://en.wikipedia.org/wiki/10-20_system_\(EEG\)](https://en.wikipedia.org/wiki/10-20_system_(EEG)) (accessed on 10 September 2017).
47. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [[CrossRef](#)] [[PubMed](#)]
48. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
49. DEEPLARNING4J. Available online: <https://deeplearning4j.org/> (accessed on 10 September 2017).
50. Panksepp, J. A role for affective neuroscience in understanding stress: The case of separation distress circuitry. *Psychobiol. Stress* **1990**, *54*, 41–57.
51. Papez, J.W. A proposed mechanism of emotion. *Arch. Neurol. Psychiatry* **1937**, *38*, 725–743. [[CrossRef](#)]
52. Davidson, R.J.; Sutton, S.K. Affective neuroscience: The emergence of a discipline. *Curr. Opin. Neurobiol.* **1995**, *5*, 217–224. [[CrossRef](#)]

