

Research Article

Construction and Analysis of Emotion Computing Model Based on LSTM

Huiping Jiang , **Rui Jiao** , **Zequn Wang** , **Ting Zhang** , and **Licheng Wu** 

Brain Cognitive Computing Lab, School of Information and Engineering, Minzu University of China, Beijing 100081, China

Correspondence should be addressed to Licheng Wu; wulicheng@tsinghua.edu.cn

Received 17 September 2020; Revised 25 November 2020; Accepted 21 January 2021; Published 3 February 2021

Academic Editor: Ning Cai

Copyright © 2021 Huiping Jiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The electroencephalogram (EEG) is the most common method used to study emotions and capture electrical brain activity changes. Long short-term memory (LSTM) processes the temporal characteristics of data and is mostly used for emotional text and speech recognition. Since an EEG involves a time series signal, this article mainly studied the introduction of LSTM for emotional EEG recognition. First, an ALL-LSTM model with a four-layered LSTM network was established in which the average accuracy rate for emotional classification reached 86.48%. Second, four EEG characteristics were extracted via the wavelet transform (WT) using the LSTM-based sentiment classification network. The experimental results showed that the best average classification accuracy of these four features was 73.48%. This was 13% lower than in the ALL-LSTM model, indicating that inappropriate feature extraction methods could destroy the timing of EEG signals. LSTM can be used to thoroughly examine EEG signal timing and preprocessed EEG data. The accuracy and stability of the ALL-LSTM model are significantly superior to those of the WT-LSTM model. The result showed that the process of emotion generation based on EEG is sequential. Compared with EEG emotion extraction using WT, the raw EEG signal's timing is more suitable for the LSTM network.

1. Introduction

Suppose a high level of human-computer interaction is to be achieved. In that case, it is essential for computers to effectively recognize human emotions, which is significantly useful for realizing a brain-computer interface and intelligent machines.

People expect computers that are easier to control and anticipate a gradual change from human-operated computers to computer-aided people, signaling a transition from passive cognitive to active. The concept of affective computing was proposed by Professor Picard of the MIT Media Lab in 1997. She indicated [1] that sentiment calculation involves certain techniques to classify and interpret emotions according to specific data. Zhou [2] believes that the purpose of emotional computing is to establish a harmonious human-machine environment by providing computers with the ability to recognize, understand, express, and adapt to human emotions, equipping computers with higher and comprehensive intelligence.

Scientific evidence shows that the appearance and development of emotions occur parallel to the brain's evolution, while brain development corresponds with the differentiation and development of facial expressions [3]. Together, the nervous and endocrine systems determine the physiological changes in the human body, the signals of which are challenging to control artificially. Many methods are employed for emotion recognition, such as those based on facial expressions and physiological signals.

Using physiological signals for the recognition of emotions usually yields accurate and objective results. Furthermore, this technique aids in the safety improvement of equipment used by reducing the security risks associated with emotional factors. As a specific physiological modality, EEG signals are exceptionally valuable in emotional classification, and this method has been extensively studied. The EEG detection instrument is inexpensive and exhibits a high time resolution and a bearable space resolution. Besides these advantages, an EEG obtains more detailed, complex information in a noninvasive manner for emotion

recognition. The data cannot be deliberately modified or concealed, making EEG-based emotion recognition more effective and reliable [4].

Human-computer emotional interaction can render many computer applications more convenient and feasible. The computer can use human physiological signals to make judgments without the need for cumbersome behavioral responses. This is of considerable significance to people with disabilities involving the facial muscular system or limbs.

Many machine-learning and pattern-recognition algorithms are applied to EEG-based emotion recognition, but the generation of emotions remains a complex cognitive activity. The mechanism and process of emotion generation are still being investigated, and applying EEG signals for emotion calculation shows significant potential.

Since EEG signals are nonstationary and highly random, it is challenging to extract EEG features related to a particular cognitive task. An essential feature of the extraction method is to minimize the loss of the raw signal and simplify the raw dataset. Therefore, feature extraction aims to reduce the complexity of the application to render information processing more cost-efficient. Since Dietch first used a Fourier transform for EEG analysis in 1932, classical methods, such as frequency domain analysis, time-domain analysis, and WT, were introduced [5, 6]. Because WT is more suitable for analyzing nonstationary signals, as well as the signal in the time and frequency domain, it can be used to resolve the contradiction between the time and frequency resolutions [7].

The EEG classification challenge is essentially a pattern-recognition problem. The current methods used for classifying EEG signals include linear discriminant analysis, support vector machine (SVM), and deep learning models [8, 9]. Deep learning is a general term for this type of neural network learning algorithm depth and has attracted significant attention in recent years [10]. Deep learning models, such as the autoencoder (AE), deep belief networks (DBN), convolutional neural networks (CNN), and recurrent neural networks (RNN), are widely used [11–13]. An unsupervised DBN was applied for the depth level feature extraction from fused observation signals. Experiments involving a public multimodal physiological signal dataset show that these models significantly increase the emotion recognition rate accuracy [14]. A novel computer model [15] is presented for the EEG-based screening of depression using a CNN. The algorithm attained 93.5% and 96.0% accuracy using EEG signals from the left and right hemispheres, respectively. Results reveal that the EEG signals from the right hemisphere are more distinctive than those from the left hemisphere in depression. A compact CNN is introduced for EEG-based brain-computer interface (BCI) [16], allowing for EEG feature extraction. EEGNet can better generalize across paradigms than the reference algorithms when only limited data is available across all tested programs while achieving comparable high performance. These techniques and others have been applied to explore EEG in machine learning and deep learning, achieving positive results. However, EEG signals are composed of multilead signals and contain important

time-frequency information. Without sufficient time and frequency domain information, it is difficult to obtain a good classification result. The recurrent structure of the RNN can be used to obtain contextual information about the time series [17]. However, when the training sequence is too long, the traditional RNN faces the problem of gradient disappearance or explosion due to its structural design. Therefore, this paper proposes a new emotion recognition model based on the LSTM network to resolve this issue. Compared with standard RNN, LSTM performs better in longer sequences and can be applied widely in the technology field. LSTM-based systems can perform image analysis, speech recognition, and disease prediction [18, 19]. However, establishing an LSTM emotion model based on EEG and its application in emotion recognition requires further study.

The emotion recognition method based on EEG generally proceeds as follows. First, the EEG signals induced by specific images or videos corresponding to emotions are read. Second, models are established by learning the EEG samplings. Finally, the models are applied to the real systems.

In conclusion, one of the most critical factors of emotion recognition based on EEG is acquiring EEG ground truth data for training the models, substantially affecting the accuracy rate and stationary ability regardless of which model is adopted. Furthermore, according to the temporal characteristics of emotion induction, it can improve emotion classification accuracy by constructing an emotion recognition system with a sequential effect.

This paper aims to establish a corresponding neural network based on the experimental EEG data, fully utilizing the sequential information implicit in the EEG signals, and building a suitable deep learning neural network. The method proposed in this paper mainly solves two problems: one is the effect of the feature extraction of the preprocessed EEG signal on emotion recognition before establishing a related model and the other is the influence of the model on emotion recognition compared with existing research.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 discusses the overall framework design, experimental dataset, experimental environment, and classification. Sections 4 and 5 discuss the results for different models and compare them.

2. Related Work

2.1. Emotion Features Based on the EEG Signal. The EEG signals exhibit various frequencies. Neuroscientists have divided them into frequency bands, each of which is responsible for specific brain activity. The different brainwaves and the activity responsible for them are as follows:

Delta (0.5–4 Hz): its amplitude is about 0–200 μV , which only occurs during sleep, deep anesthesia, hypoxia, or brain lesions.

Theta (4–8 Hz): its amplitude is about 100–150 μV , which appears during drowsiness and corresponds to daydreaming, drowsiness, or sleep.

Alpha (8–12 Hz): its amplitude is about 5–20 μV , corresponding to the resting state of the brain.

Beta (12–30 Hz): it is associated with active, task-oriented, busy or anxious thinking, and active concentration.

Gamma (>30 Hz): it occurs when different populations of neurons work together to perform demanding cognitive or motor functions.

The methods for EEG signal feature extraction are relatively mature. The common emotional features include the following three categories: the time-domain features, such as the mean, standard deviation, skewness, peak amplitude, variance, skewness, and kurtosis. The second is the frequency domain features, including the features extracted via the Fourier transform and those extracted via the parameter model (such as AR, Ma, ARMA, and the harmonic signal model). Finally, there are the time-frequency features, such as the short-time Fourier transform, WT, and nonlinear dynamic features.

The WT decomposes the input signals into various constituting small range frequency bands. This is done by obtaining the approximation and detail coefficients via multiple-level decomposition.

The key to the efficient extraction of EEG features is to choose the appropriate wavelet base. Standard wavelet bases include Daubechies (dbN) wavelet, Symlets (symN) wavelet, and Coiflet (coifN) wavelet. There is no unified standard for the selected wavelet base and it primarily relies on the classification accuracy. During early research, EEG emotion classification involving different wavelet bases was shaped according to CNN [20]. The results show that the Sym8 wavelet could better classify emotion based on the raw EEG signal. Therefore, this study used Sym8 wavelet for further experimentation.

After the WT of the EEG signal, the wavelet coefficients of each layer of the frequency band were obtained. Still, the wavelet coefficients cannot be sent directly to the classifier as features and require further processing to extract the EEG features. The features selected in this paper include the band energy (E), the band energy ratio (REE), the logarithm of the

band energy ratio (LREE), and the differential entropy (DE) and are described below.

The E refers to the energy of each frequency band after WT and is obtained by square-summing the coefficients of each frequency band. The solution formula is shown as follows:

$$E_i = \sum_{j=1}^{n_i} d_{ij}^2, \quad (1)$$

where E_i is the energy of the i -th band, n_i is the number of coefficients decomposed by the i -th layer, and d_{ij} is the j -th wavelet coefficient of the i -th layer.

The REE refers to the ratio of each layer of energy to the total energy and is expressed as follows:

$$\text{REE}_i = \frac{E_i}{\sum_{j=1}^n E_j}, \quad (2)$$

where REE_i is the band energy ratio of the i -th band and n is the number of bands.

The LREE for each band is based on 10 and is expressed as follows:

$$\text{LREE}_i = \log_{10} \text{REE}_i, \quad (3)$$

where LREE_i represents the logarithm of the energy ratio of the i -th band.

If the signal obeys a different distribution, the DE is solved differently. It is assumed that the acquired EEG signal, X , is affected by the Gaussian distribution, as shown in equations (4)–(6):

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-((x-m)^2/2\sigma^2)}, \quad -\infty < x < \infty, \quad (4)$$

$$m = E(x) = \int_{-\infty}^{\infty} x p(x) dx, \quad (5)$$

$$\sigma^2 = E[(x-m)^2] = \int_{-\infty}^{\infty} (x-m)^2 p(x) dx, \quad (6)$$

The solution process of the DE is shown as follows:

$$\begin{aligned} h(x) &= - \int_{-\infty}^{\infty} p(x) \log p(x) dx = - \int_{-\infty}^{\infty} p(x) \log p(x) \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-((x-m)^2/2\sigma^2)} \right) dx \\ &= \log \sqrt{2\pi\sigma^2} \int_{-\infty}^{\infty} p(x) dx + \log e \int_{-\infty}^{\infty} p(x) \frac{(x-m)^2}{2\sigma^2} = \log \sqrt{2\pi\sigma^2} + \frac{1}{2} \log e = \frac{1}{2} \log(2\pi e \sigma^2). \end{aligned} \quad (7)$$

This formula indicates that the key to solving DE is acquiring the EEG signal variance, which is approximately the same as the average of the energy of the EEG signal in each band. In practical applications, the E value is commonly used as a logarithm instead of DE. The simplified formula for DE is expressed by equation (8):

$$\text{DE}_i = \log_{10} E_i, \quad (8)$$

where DE_i represents differential entropy.

2.2. Deep Learning for EEG Analysis. Over the past few years, traditional machine learning technology (i.e., nondeep

learning algorithm) has been the only feasible EEG analysis option. It continues to be widely used in combination with various feature extraction and feature selection algorithms [21–24].

As a relatively new trend, the deep learning algorithm has been applied in medical image and signal processing due to the improvement and availability of computing power and big data. In most cases, its performance exceeds the rates that have been previously achieved with traditional machine learning techniques [25].

Many methods have been proposed for studying appropriate computational models for emotion recognition using EEG signals. Various deep learning structures have been used to classify EEG signals to solve different recognition tasks. Generally, most existing EEG research based on deep learning can be summarized into two categories. The first is based on EEG signals input to the network. The second type is based on features extracted from EEG signals as input to the network.

An EEG analysis task requires the developed model to capture private information from EEG signals. Traditional machine learning methods need to design and extract the features of EEG signals manually. The redundancy of the features is exceedingly high and does not consider the temporal dynamics of the EEG signals crucial for emotion recognition.

Tripathi et al. [26] proposed an emotion recognition method based on CNN from EEG signals in the Database for Emotion Analysis using Physiological Signals (DEAP) dataset. They explored two different neural models: a simple deep neural network and a CNN. The performance of the latter is 4.96% higher than that in state-of-the-art techniques.

Shawky et al. [27] presented a three-dimensional CNN approach for recognizing emotions from multichannel EEG signals. They developed a data enhancement phase to improve the performance of their 3D CNN model. They achieved 87.44% accuracy for valence and 88.49% for arousal.

Moon et al. [28] applied CNN to recognize emotion based on EEG. They employed brain connectivity features to explain the synchronous activation of different brain regions, an approach that has not been used in previous studies. Therefore, their method effectively captures the asymmetric brain activity patterns, playing a vital role in emotion recognition.

LSTM [29–31] is one form of RNN that overcomes the problem of exploding and vanishing gradients. The building blocks of LSTM include a cell, an input gate, an output gate, and a forget gate. The cell is responsible for handling long-term dependency while the three gates regulate the flow of values between the different layers of the LSTM network.

The innovation of LSTM networks compared to traditional RNNs is the inclusion of “gates” to solve the vanishing gradient problem and allow the algorithm to control more precisely what information needs to be retained in its memory and what must be removed [32, 33]. By controlling the learning rate with the three gates (i.e., input gate, forget gate, and output gate), the LSTM network can better adjust

to large data series sequences than RNNs and other deep learning techniques. Considering that EEG signals are essentially highly dynamic, nonlinear time-series data, LSTM networks are better than CNN in isolating the temporal characteristics of brain activity during different states as reported in various applications, such as emotion recognition, confusion estimation, and estimation prediction [26, 34–36]. Despite their inherent advantages in EEG analysis, LSTM models have not been examined combined with emotion feature extraction.

This paper analyzes an emotion recognition framework based on the LSTM. First, the multichannel EEG signal is divided into multiple segments, and the time domain, frequency domain, and nonlinear dynamic features are extracted from each segment of the signal to form a feature sequence along with time, respectively. Each feature sequence consists of characteristics representing specific feature information of the signal. Second, an LSTM neural network is used to obtain the time dynamic information from various feature sequences and make the final emotion prediction.

3. Methods

3.1. Framework Design. Figure 1 shows that the framework includes three parts: source signal processing, feature extraction, and sentiment classification.

3.2. Experimental Dataset. Here, 12 videos were selected as emotion-evoking stimuli to cover the entire emotional spectrum. Six of 12 videos were excerpts from movies and were chosen based on the preliminary study. During the initial investigation, the participants self-assessed their emotions by reporting their arousal feeling (ranging from calm to excited/activated) and valence (ranging from unpleasant to pleasant) on a nine-point scale. Sam Manikins were shown to facilitate the self-assessments of valence and arousal [37]. Ultimately, six videos between 110 s and 120 s long were selected to be shown. Psychologists recommend videos from 1 min to 10 min long to elicit a single emotion [38]. Here, the video clips were kept as short as possible to avoid multiple emotions or habituation to the stimuli [39] while keeping them long enough to observe the effect—data collection. The stimulus file is shown in Table 1.

Twenty younger adults (11 women), mainly students at the Minzu University of China (mean age: 21.4 years, range: 20–23), participated in the experiment. They were paid 150 RMB per hour for their participation. Participants were right-handed as assessed by a German version of the Edinburgh handedness inventory [40] and had a normal or corrected-to-normal vision. None of them reported neurological or psychological disorders. All participants were fully aware of the purpose of the study. The study was approved by the Local Ethics Committee (Minzu University of China, Beijing, ECMUC2019008CO).

3.3. Experiment Environment. In the cognitive brain laboratory, the electrode Synamps2 amplifier and Scan4.5 software developed by the Neuroscan Company, a cap with

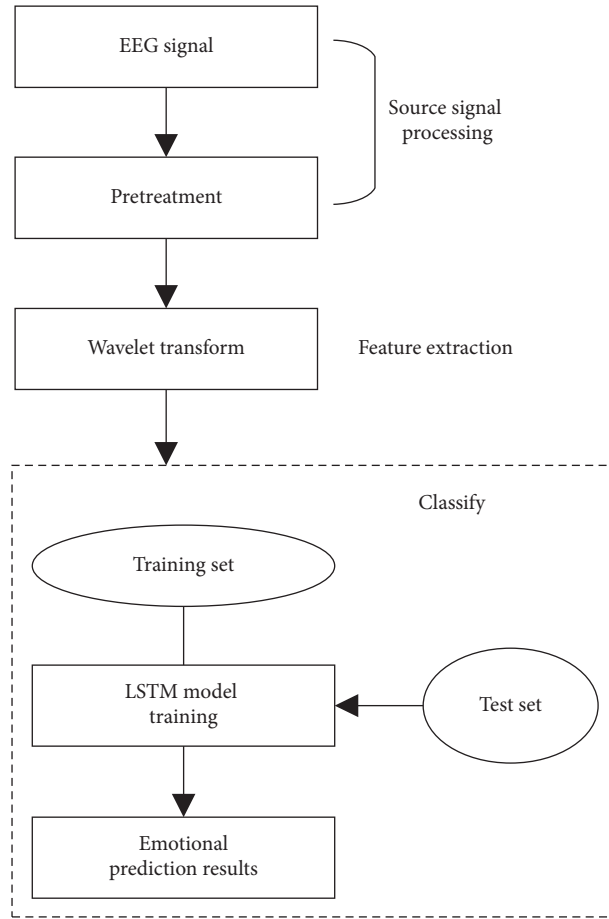


FIGURE 1: Framework design.

TABLE 1: The movie list.

Movie slices sources	Label
Martial arts biography	Positive
Love apartments	Positive
Pet funny video	Positive
Death is coming	Negative
The chainsaw cry	Negative
Teacher's grace	Negative

64 electrodes, and a computer (labeled as computer 1) were used to collect the EEG signals, while a webcam camera and another computer (labeled as computer 2) were used for documenting the facial expressions. A dedicated server was employed to generate the induction files developed with the E-prime application software while coordinating computer 1 and computer 2 to obtain the EEG records and facial expressions simultaneously.

The Neurscan-64 system was selected for EEG acquisition, while the electrode cap collected 64-channel EEG signals. The EEG sampling frequency was set to 1000 HZ, which fulfilled the requirements of rapidly changing EEG signals. The electrode distribution of the electrode cap was based on the currently used 10/20 system electrode placement method. Figure 2 shows the specific experimental process.

Furthermore, to obtain a better mood in the wake of the formal experiment, the video, as well as positive and negative video capture emotions samples, was randomly presented. Each test lasted about 25 min in total, and the screen displayed a 3000 ms gaze point “+” to prompt the participant to focus, immediately playing a stimulus video. The video was displayed for about 3 min. After the video had completed playing, the subject had to provide feedback regarding the subjective feelings after watching the material by pressing a button. The participants had a choice between three alternatives: “positive,” “neutral,” and “negative.” After providing feedback, a black screen appeared for 7000 ms to clear the participant’s thoughts and reduce mutual interference between videos. After the experiment was completed, the EEG data samples, including the positive emotions and negative emotions, were mixed, and part of the EEG data was proportionally selected as the training set for the model, while the remaining part of the EEG data denoted the test set.

The EEG signal is fragile and extremely susceptible to the internal or external environment during the measurement process. This rendered the collected signal unreliable, while it was subject to interference by many electrical activities not originating from the brain. These interferences are known as artifacts. Common artifacts originate from electrooculograms, electrocardiograms, electromyography, and electrode

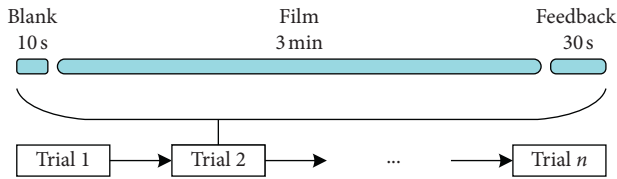


FIGURE 2: Emotional induction experimental process.

movements. The experimental environment used in the acquisition of EEG signals can be controlled artificially. However, it is challenging to artificially intervene in the human body's unconscious activities. Therefore, the unsatisfactory EEG area was deleted, while the electrooculogram artifacts and other interferences were removed. Digital filtering was performed to preprocess the EEG signals in preparation for the next step of feature extraction and EEG signal classification.

3.4. Classifier. The manuscript uses the LSTM network to classify the emotions obtained via the EEG. Two classification models were compared based on LSTM. One was to perform feature extraction and provide input to the LSTM network for classification, while the other used LSTM directly for classification.

4. Result

4.1. The Construction of an LSTM Model Based on EEG. The LSTM-based emotion classification model established in this paper consisted of four layers: the input layer, the LSTM layer, the fully connected layer, and the output layer. In establishing the LSTM network, different parameters determined different network structures. It was necessary to select the appropriate number of layers and determine the number of hidden nodes in each layer. These parameters directly determined the training speed of the network, the level of classification accuracy, and the stability of the network.

The advantage of deep learning is its colossal network scale. However, as the network scale expands, more computing resources are required. Too many nodes in the hidden layer may cause the training speed to decline or even overfit. The experiments indicated that when the number of hidden layer units was below a specific value, it became challenging to fit the model. It required the design of a more streamlined model structure on the premise of meeting the accuracy requirements. Therefore, as few nodes in the hidden layer and the number of LSTM layers as possible should be selected while ensuring accuracy. Due to the LSTM structure, the neuron's state continuously changed as the input increased, while the historical information of the data was saved, and the hidden layer could utilize the output of each step as the next input. Experiments were conducted on the single-layered and multilayered LSTM structures. After adjustment, it was found that the classification of the multilayered LSTM was superior. The multilayered LSTM sent the output values of the front-end LSTM as input to the back-end LSTM, while the LSTM could be stacked infinitely in a similar way. Finally, an LSTM-

based emotion classification model was established to classify the features of wavelet extraction. The model consisted of four layers, each of which exhibited 32 hidden nodes.

Because of the distinct individual differences in EEG, the EEG classification tasks in this paper are based on the EEG signals collected by individual people. The single-person EEG data is divided into a training set and a test set, where the model was trained using the former and tested using the latter.

Via continuous debugging, the Adam algorithm is used for parameter optimization and denotes an adaptive moment estimation method to calculate the adaptive learning rate for each parameter. In practical applications, the convergence speed of the network can be accelerated, which achieves excellent results. Compared with other adaptive learning rate algorithms, the convergence speed is faster, obtaining a more effective learning effect. The learning rate is set to 0.005. During the process of neural network training, regularization or Dropout is usually used to avoid overfitting. The model uses the Dropout method, with the parameter value set to 0.5. During the training process, this section uses the Batch technology. During the experiment, the batch was set to 16, 32, and 64, respectively. The analysis found that an appropriate increase in batch size could improve memory utilization and enhance the running speed, each finalizing a batch size of 64 training examples. Google's Tensor Flow framework implements an LSTM network model. The specific parameters are shown in Table 2.

4.2. ALL-LSTM. The LSTM model's structural properties allow it to learn the timing characteristics of the data, facilitating long-term memory. Therefore, the ALL-LSTM model does not perform artificial feature extraction from the EEG data but selects the preprocessed full-scale EEG information and sends it directly to the LSTM-based emotion classification model shown in Figure 3.

The ALL-LSTM emotion classification model consists of four layers. The first layer takes the preprocessed full EEG sequence as input. The second is the LSTM layer, which extracts contextual related features from the input EEG sequence, such as the time-domain information. The third is a fully connected layer, which is used to integrate the features extracted by the LSTM layer. It is a linear combination of the output of all LSTM units during the last time step. The function of this layer is to combine different feature-dynamic information learned from each LSTM unit. The output of this layer represents the input to the SoftMax layer to predict the emotional state. The fourth is the output layer, producing the recognized emotion category.

The dataset used in this paper is represented by the raw EEG data collected from subjects watching the stimulation material. Each subject's EEG data was divided into 10 ms (10 sampling points), obtaining 100440 EEG data from each. Each EEG data dimension collected via the 64-electrode cap presented a matrix of 64×10 . According to the LSTM principle, each column of the matrix (the voltage value collected by the 64-lead electrode) was selected as the data read in one step. Each row of the matrix

TABLE 2: LSTM emotion classification model specific parameter settings.

Name	Parameter
Learning rate	0.005
Input dimension	64*5, 64*10
Output category	2
Batch	64
Dropout	0.5
Hidden node number	32
LSTM layer number	4

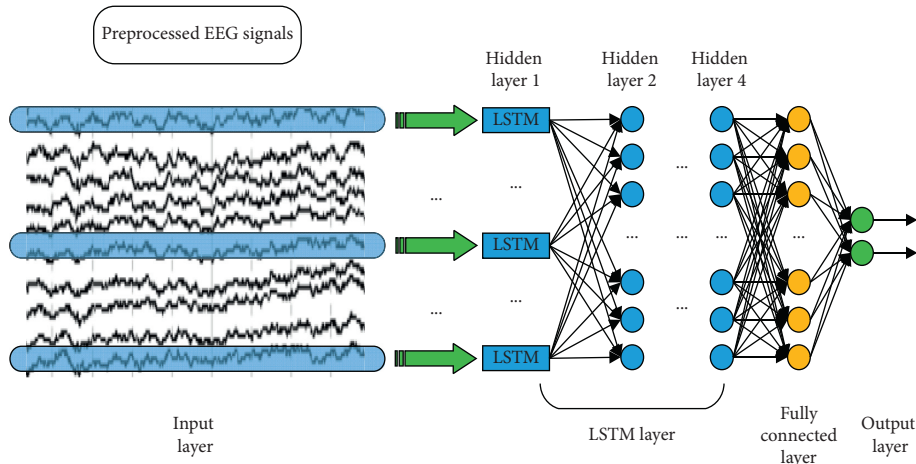


FIGURE 3: ALL-LSTM model.

(10 sampling points with a duration of 10 ms) was considered the time step number. One of the advantages of intercepting EEG data in this way is that the amount of data is sufficient since deep learning requires an abundant amount of data as a basis. Seventy-five thousand were selected as the training set, accounting for about 75%, and 25440 were selected as the test set, accounting for about 25%. The ratio of the training set to the test set was about 3:1.

All the EEG training data obtained from a single person was sent to the LSTM emotion classification model in batch polls for training, a process known as completing an epoch. After each epoch, the loss and accuracy rate of the training set based on the existing training model was provided. After 200 epochs, the model tended to converge, and the training was completed. The training process is shown in Figure 4, in which the abscissa represents the training number, and the ordinate denotes the accuracy during the training.

The test set was sent to the trained model. The final classification results of the eight subjects are shown in Table 3.

These analytical results indicate that the ALL-LSTM model classification's average accuracy is 86.48%, and the variance is 0.0039.

4.3. WT-LSTM. The WT-LSTM was established using EEG data after performing the WT emotion recognition process based on the LSTM emotion classification model. This paper examines four standard, efficient wavelet features, namely, E ,

REE, LREE, and DE. The same source data is used throughout the experiment and sent to the LSTM emotion classification model using the same parameters. The classification results corresponding to these four factors were obtained, and the EEG features that were most compatible with the LSTM network were identified. Therefore, the classification results obtained by changing only the feature parameters can clearly reflect the efficacy of the features.

Figure 5 shows that the sentiment classification model based on the WT-LSTM established in this section consists of four layers. First, the input layer considers the extracted wavelet features as input. The second layer represents the LSTM layer, extracting context-related features from the input information. The third layer is a fully connected layer, which is used to integrate the features extracted by the LSTM layer and convert the information from the LSTM into the desired output. The fourth layer is the output layer, which aims to output the recognized emotion categories.

The EEG data represent the dataset used in this article that collected the subjects watching the stimulation material simultaneously. Because the EEG signals recorded too many potential values, which were doped with some unwanted potential values, it was necessary to segment each person's EEG signal samples. However, since there is no clear conclusion regarding the time range of human emotional change, the length of each sample selected in this paper is 3000 ms (3000 sampling points). Each 3,000 ms was divided into one EEG data unit. Since the electrode cap used in the experiment was a 64-lead, the dimension of

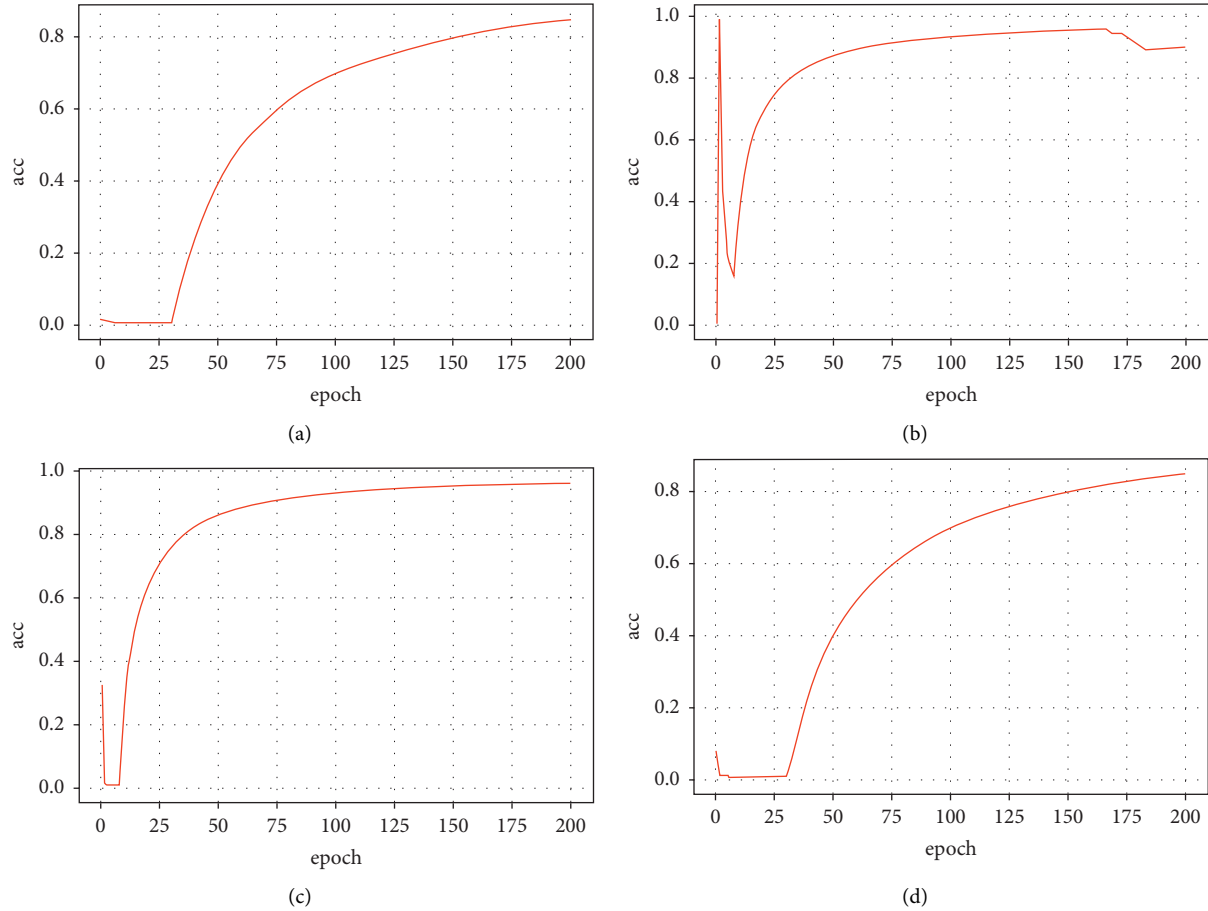


FIGURE 4: The random training process of four participants.

TABLE 3: ALL-LSTM model classification results.

Experimenter	Accuracy %
Subject 1	84.72
Subject 2	84.57
Subject 3	96.04
Subject 4	84.57
Subject 5	90.36
Subject 6	93.02
Subject 7	81.91
Subject 8	76.64
AVG	86.48
VAR	0.0039

each EEG data unit was 64×3000 . Of the total 348 EEG data units obtained after each participant's segmentation, 260 were randomly selected as the training set, accounting for about 75%, and 88 were chosen as the test set, accounting for about 25%, while the ratio of the data in the training set to the test set was approximately 3 : 1. After the five-layered WT, each raw EEG data dimension was reduced to a matrix of 64×5 . According to LSTM principles, this paper selected each column of the matrix as the data read in one step and considered each row of the matrix as the number of time steps.

The frequency E , frequency REE, frequency LREE, and DE were extracted as EEG features for training. Figure 6 shows a training process using the frequency LREE as a feature. The abscissa represents the training number in the figure, while the ordinate denotes the accuracy rate during training. Each image randomly shows the training process of four participants.

After 40 training epochs, the model tended to converge. The classification results of the eight subjects and the average classification accuracy and variance corresponding to the four characteristics are shown in Table 4.

It can be concluded from the above table that the classification accuracy of the four features of Subject 1 is low, and the classification accuracy of the four features of Subject 6 exceeds 80%. The classification accuracy is significantly affected by individual differences. Analysis of the four characteristics indicated an accuracy rate of 50% or less exhibiting one E , two REE, two DE, and zero LREE. An accuracy rate of 80% or more displays one E , one REE, four DE, and three LREE. An accuracy rate of 90% or more shows zero E , zero REE, zero DE, and one LREE. Furthermore, when the frequency LREE is used as the feature, the average classification accuracy is the highest, and the variance value is the lowest. Therefore, the use of LREE as a feature leads to

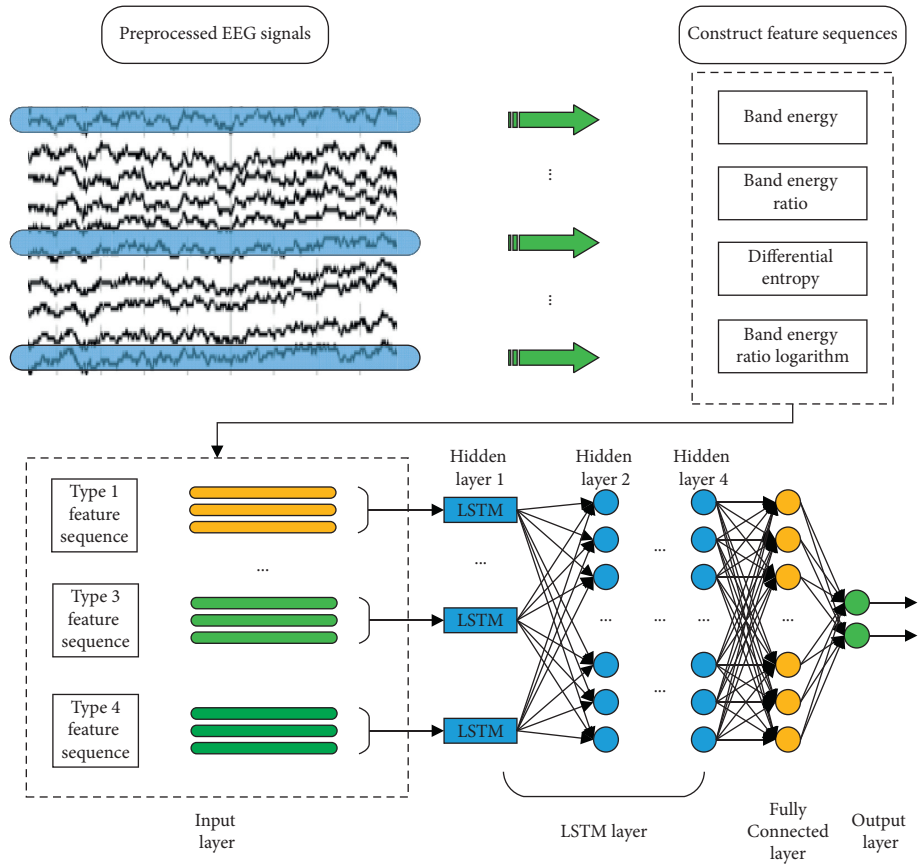


FIGURE 5: WT-LSTM model.

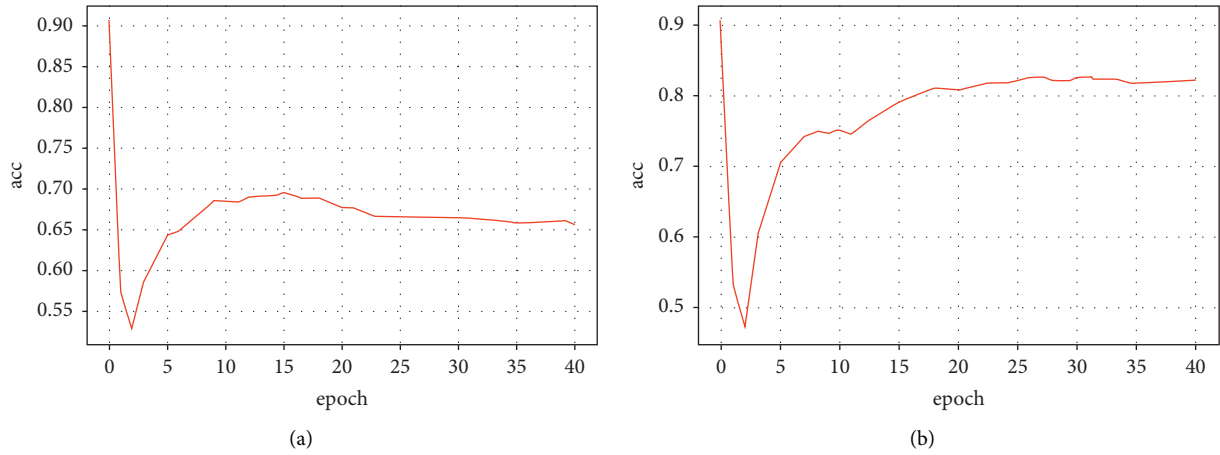


FIGURE 6: Continued.

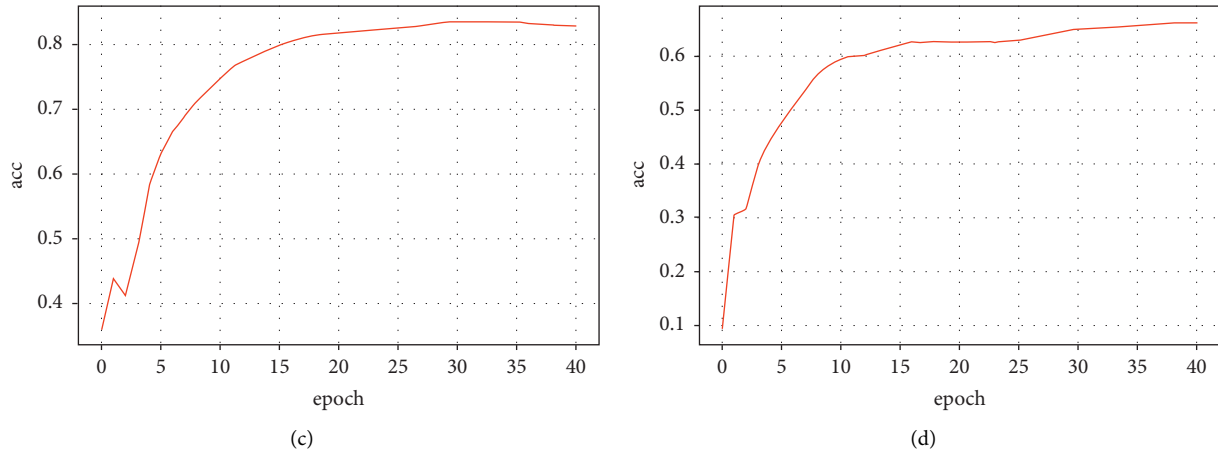


FIGURE 6: The training process of the LREE as an EEG feature.

TABLE 4: Test set classification results.

Sample	E (%)	REE (%)	DE (%)	LREE (%)
Subject 1	57.05	68.52	85.09	83.48
Subject 2	57.32	43.62	45.85	66.73
Subject 3	58.17	58.97	70.91	66.38
Subject 4	71.29	74.15	72.73	65.86
Subject 5	75.20	60.23	85.47	91.17
Subject 6	89.95	86.15	86.78	82.14
Subject 7	45.93	47.11	47.83	59.47
Subject 8	72.95	74.97	84.33	72.65
AVG	65.98	64.22	72.37	73.48
VAR	0.0193	0.0210	0.0284	0.0119

TABLE 5: Comparison of results of the two classification models.

Model	Average accuracy (%)	The variance of accuracy
ALL-LSTM	86.48	0.0039
WT-LSTM	73.48	0.0119

higher accuracy and higher stability when used for emotion classification.

These results indicate that when the LREE is selected as the feature for the WT-LSTM model, the classification effect is the best, and the average accuracy of the classification is 73.48%. The LREE is also the most suitable EEG feature of the four wavelet features in the emotion classification problem and LSTM network.

5. Discussion and Conclusion

In theory, the advantage of feature extraction for the WT-LSTM model is that it provides a straightforward solution for obtaining hidden information about the signal's frequency contents or brain area connectivity from the available channels, compared to the information gained by using the EEG signals as a time series. However, Table 5 shows that the best classification accuracy of the WT-LSTM

model with features extracted via WT is still significantly lower than that of the ALL-LSTM model.

The best average accuracy of the WT-LSTM model is about 13% lower than the average accuracy of the ALL-LSTM model, while the variance also increases. The results showed that the ALL-LSTM model's stability was slightly better than that of the WT-LSTM model in the emotion classification of these eight subjects.

This could be attributed to LSTM displaying a strong ability to use context. Feature extraction based on WT is currently based on more complex and mature feature extraction methods used during EEG emotion recognition. However, feature extraction via WT may destroy the timing of the EEG signal itself. Timing information is vital for emotional classification, and the LSTM model can make full use of the timing information implicit in the EEG data. The significant improvement in the classification ability of the ALL-LSTM model also shows that the method is feasible for EEG-based emotion recognition.

Furthermore, by adding more layers and memory units, better EEG signal representation could be learned when the LSTM network's size was substantially increased, compensating for the more extensive input size by directly providing the EEG signals. However, the computational cost of training larger LSTM networks increases rapidly, requiring extended training time or GPU arrays. Regardless of the computational cost, this method could need even more EEG data to effectively train the millions of network parameters.

Therefore, the goal is to train the LSTM network by learning the suitable emotion features, which can be realized by essentially simulating a more profound and more complex LSTM model without increasing the time and data training limitations. Therefore, the emotion recognition system can run under more suitable conditions.

Data Availability

The EEG data used to support the findings of this study were supplied by National Nature Science Foundation of China

under license and so cannot be made freely available. Requests for access to these data should be made to Huiping Jiang, jianghp@muc.edu.cn.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Huiping Jiang was supported by the National Nature Science Foundation of China (No. 61503423). This work was supported in part by the Leading Talent Program of State Ethnic Affairs Commission and Double First-Class Special Funding of MUC. The authors thank all the participants in this research and the technical support from FISTAR Technology Inc.

References

- [1] R. W. Picard, "Affective computing: challenges," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 55-64, 2003.
- [2] Q. Zhou, "Multi-layer affective computing model based on emotional psychology," *Electronic Commerce Research*, vol. 18, no. 1, pp. 109-124, 2018.
- [3] K. Bernhard, L. Suzana, K. Mathias et al., "Decision support with text-based emotion recognition: deep learning for affective computing," *Decision Support Systems*, vol. 115, pp. 24-35, 2018.
- [4] F. Wang, S.-H. Zhong, J. Peng, J. Jiang, and Y. Liu, "Data augmentation for EEG-based emotion recognition with deep convolutional neural networks," in *MultiMedia Modeling*, pp. 82-93, Springer, Cham, Switzerland, 2018.
- [5] T. Tanaka, T. Uehara, and Y. Tanaka, "Dimensionality reduction of sample covariance matrices by graph fourier transform for motor imagery brain-machine interface," in *Proceedings of the 2016 IEEE Statistical Signal Processing Workshop (SSP) IEEE*, Palma de Mallorca, Spain, June 2016.
- [6] B. K. Kim, E. C. Lee, B. Suhng, D. Ryu, and W. Lee, "Feature extraction using FFT for banknotes recognition in a variety of lighting conditions," in *Proceedings of the 2013 13th International Conference on Control, Automation and Systems (ICCAS 2013)*, pp. 698-700, Gwangju, South Korea, October 2013.
- [7] D. Wang, D. Q. Miao, and R. Z. Wang, "A new method of EEG classification with feature extraction based on wavelet packet decomposition," *Acta Electronica Sinica*, vol. 41, no. 1, pp. 193-198, 2013.
- [8] G. Akshansh, R. K. Agrawal, and B. Kaur, "Performance enhancement of mental task classification using EEG signal: a study of multivariate feature selection methods," *Soft Computing*, vol. 19, no. 10, pp. 2799-2812, 2015.
- [9] A. Subasi and M. Ismail Gursoy, "EEG signal classification using PCA, ICA, LDA and support vector machines," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8659-8666, 2010.
- [10] L. Deng, "The MNIST database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141-142, 2012.
- [11] M. Yanagimoto and C. Sugimoto, "Recognition of persisting emotional valence from EEG using convolutional neural networks," in *Proceedings of the 2016 IEEE 9th International Workshop on Computational Intelligence and Applications (IWCI/A)*, Hiroshima, Japan, November 2016.
- [12] T. Muhammed, U. B. Baloglu, Ö. Yildirim, and U. Rajendra Achary, "Application of deep transfer learning for automated brain abnormality classification using MR images," *Cognitive Systems Research*, vol. 54, pp. 176-188, 2019.
- [13] N. Sejjdel, K. Ramakrishnan, M. Losch, and S. Scholte, "Overlap in performance of CNN's, human behavior and EEG classification," *Journal of Vision*, vol. 16, no. 12, p. 501, 2016.
- [14] M. M. Hassan, M. G. R. Alam, M. Z. Uddin, S. Huda, A. Almogren, and G. Fortino, "Human emotion recognition using deep belief network architecture," *Information Fusion*, vol. 51, pp. 10-18, 2018.
- [15] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli, and D. P. Subha, "Automated EEG-based screening of depression using deep convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 161, pp. 103-113, 2018.
- [16] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018.
- [17] P. S. Huang, X. He, J. Gao, L. Deng, A. Acero, and L. Heck, "Learning deep structured semantic models for web search using clickthrough data," in *Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management*, San Francisco, CA, USA, October 2013.
- [18] Z. Ni, A. C. Yuksel, X. Ni, M. I. Mandel, and L. Xie, "Confused or not confused? disentangling brain activity from EEG data using bidirectional LSTM recurrent neural networks," in *Proceedings of the 8th ACM International Conference ACM*, Boston, MA, USA, August 2017.
- [19] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Transactions on Affective Computing*, vol. 7, no. 1, pp. 17-28, 2016.
- [20] B. Y. Zhang, H. P. Jiang, and L. Dong, "Classification of EEG signal by WT-CNN model in emotion recognition system," in *Proceedings of the 2017 IEEE 16TH International Conference on Cognitive Informatics & Cognitive Computing (ICCI CC)*, pp. 109-114, Oxford, UK, July 2015.
- [21] G. Ruffini, D. Ibañez, M. Castellano, S. Dunne, and A. Soria-Frisch, "EEG-driven RNN classification for prognosis of neurodegeneration in at-risk patients," in *25th International Conference on Artificial Neural Networks (ICANN 2016)*, pp. 306-313, Barcelona, Spain, October 2016.
- [22] U. Elif Derya, "Analysis of EEG signals by implementing eigenvector methods/recurrent neural networks," *Digital Signal Processing*, vol. 19, no. 1, pp. 134-143, 2009.
- [23] Z. Tang, C. Li, and S. Sun, "Single-trial EEG classification of motor imagery using deep convolutional neural networks," *Optik*, vol. 130, pp. 11-18, 2017.
- [24] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals," *Computers in Biology and Medicine*, vol. 100, pp. 270-278, 2018.
- [25] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," *Briefings in Bioinformatics*, vol. 18, no. 5, pp. 851-869, 2017.
- [26] S. Tripathi, S. Acharya, R. D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset," in *Proceedings of the Twenty-Ninth AAAI Conference on*

- Innovative Applications (IAAI-2017)*, San Francisco, CA, USA, February 2017.
- [27] E. Shawky, R. El-Khoribi, M. Shoman, and M. Wahby Shalaby, "EEG-based emotion recognition using 3D convolutional neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 8, pp. 329–337, 2018.
 - [28] S.-E. Moon, S. Jang, and J. Lee, "Convolutional neural network approach for EEG-based emotion recognition using brain connectivity and its spatial information," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2556–2560, Calgary, Canada, April 2018.
 - [29] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
 - [30] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch, "Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG," *Neurocomputing*, vol. 30, no. 1–4, pp. 201–218, 2000.
 - [31] A. Graves, *Supervised Sequence Labelling with Recurrent Neural Networks*, Springer, Berlin Germany, 2012.
 - [32] F. A. Gers, I. Galleria, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," *Journal of Machine Learning Research*, vol. 3, no. 1, pp. 115–143, 2003.
 - [33] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 2, pp. 107–116, 1998.
 - [34] R. DiPietro and G. D. Hager, "Deep learning: RNNs and LSTM," in *Handbook of Medical Image Computing and Computer Assisted Intervention*, pp. 503–519, Academic Press, Cambridge, MA, USA, 2020.
 - [35] M. M. Hasib, T. Nayak, and Y. Huang, "A hierarchical LSTM model with attention for modeling EEG non-stationarity for human decision prediction," in *Proceedings of the 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, Las Vegas, NV, USA, March 2018.
 - [36] S. Alex, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena*, vol. 404, 2020.
 - [37] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
 - [38] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers," *Cognition & Emotion*, vol. 24, no. 7, pp. 1153–1172, 2010.
 - [39] R. J. Barry and E. N. Sokolov, "Habituation of phasic and tonic components of the orienting reflex," *International Journal of Psychophysiology*, vol. 15, no. 1, pp. 39–42, 1993.
 - [40] R. C. Oldfield, "The assessment and analysis of handedness: the Edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, 1971.