Emotion Recognition From Multi-Channel EEG via Deep Forest

Juan Cheng, *Member, IEEE*, Meiyao Chen, Chang Li, *Member, IEEE*, Yu Liu, *Member, IEEE*, Rencheng Song, *Member, IEEE*, Aiping Liu, and Xun Chen, *Senior Member, IEEE*

***Abstract*—Recently, deep neural networks (DNNs) have been applied to emotion recognition tasks based on electroencephalography (EEG), and have achieved better performance than traditional algorithms. However, DNNs still have the disadvantages of too many hyperparameters and lots of training data. To overcome these shortcomings, in this article, we propose a method for multi-channel EEG-based emotion recognition using deep forest. First, we consider the effect of baseline signal to preprocess the raw artifact-eliminated EEG signal with baseline removal. Secondly, we construct $2D$ frame sequences by taking the spatial position relationship across channels into account. Finally, $2D$ frame sequences are input into the classification model constructed by deep forest that can mine the spatial and temporal information of EEG signals to classify EEG emotions. The proposed method can eliminate the need for feature extraction in traditional methods and the classification model is insensitive to hyperparameter settings, which greatly reduce the complexity of emotion recognition. To verify the feasibility of the proposed model, experiments were conducted on two public DEAP and DREAMER databases. On the DEAP database, the average accuracies reach to 97.69% and 97.53% for valence and arousal, respectively; on the DREAMER database, the average accuracies reach to 89.03%, 90.41%, and 89.89% for valence, arousal and dominance, respectively. These results show that the proposed method exhibits higher accuracy than the state-of-art methods.**

***Index Terms*—Deep neural networks (DNNs), emotion recognition, multi-channel EEG, deep forest, spatio-temporal information.**

Manuscript received November 10, 2019; revised April 20, 2020; accepted May 13, 2020. Date of publication May 19, 2020; date of current version February 4, 2021. This work was supported in part by the National Key R&D Program of China under Grant 2017YFB1002802, in part by the National Natural Science Foundation of China under Grants 61922075, 41901350, 61701160, and 61701158, and in part by the Fundamental Research Funds for the Central Universities under Grants JZ2019HGBZ0151 and JZ2020HGPA0111. *(Corresponding author: Chang Li.)*

Juan Cheng, Meiyao Chen, Chang Li, Yu Liu, and Rencheng Song are with the Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, China (e-mail: chengjuan@hfut.edu.cn; cmy@mail.hfut.edu.cn; changli@hfut.edu.cn; yuliu@hfut.edu.cn; rcsong@hfut.edu.cn).

Aiping Liu and Xun Chen are with the Department of Electronic Engineering & Information Science, University of Science and Technology of China, Hefei 230026, China (e-mail: aipingl@ustc.edu.cn; xunchen@ustc.edu.cn).

Digital Object Identifier 10.1109/JBHI.2020.2995767

I. Introduction

EMOTION is a complex psychological and physiological state [1], which affects people's cognition, behavior and interpersonal communication. The study of human emotions is a subject that has been under way for nearly a century and a half. With the increasing demand for human-computer interaction (HCI), whether the machine can correctly analyze the user's emotional state has become the key to interactive experience. Judging the human emotions must rely on long-term experience and personal thinking, which is a very complicated thing for humans, not to mention machines. Therefore, automatic emotion recognition is a challenging task that has attracted numerous attention [2]–[5].

Over the years, there have been several research directions on emotion recognition, including emotion recognition based on behavioral responses (such as facial expressions, vocal intonations and body postures) and emotion recognition based on physiological signals. Compared with behavioral responses, physiological signals can reflect dynamic changes in the central nervous system, which are difficult to hide and closer to people's real emotions [6]. Physiological signals such as electrooculogram (EOG), electrocardiogram (ECG) and electromyogram (EMG) are usually indirect responses caused by emotion with low recognition accuracy, while electroencephalography (EEG) signals have good time resolution [7], which can provide a direct and comprehensive means for emotion recognition with higher classification accuracy [8], [9]. Therefore, the EEG-based method is irreplaceable in the field of emotion recognition [10], [11].

Currently, there are two emotional models to construct emotional space, discrete model and dimensional model. According to the discrete model, the emotional space consists of a limited number of basic discrete emotions, such as happiness, sadness, surprise, fear, anger and disgust. The dimensional model defines emotions as points in a dimensional space. The more similar emotions are, the closer their coordinates are in the dimensional space. The researchers proposed a dimensional model of two or three dimensions: valence-arousal, valence-arousal-dominance [12], where the two-dimensional model is the most widely used, illustrated in Fig. 1. Valence refers to the positive degree of emotion, arousal refers to the intensity of emotion, and dominance refers to the degree of subjective control of the individual's emotional state. Considering that the dimensional model can show more feelings and is closer to human perception of feelings [13], we adopt the dimensional model in this study.

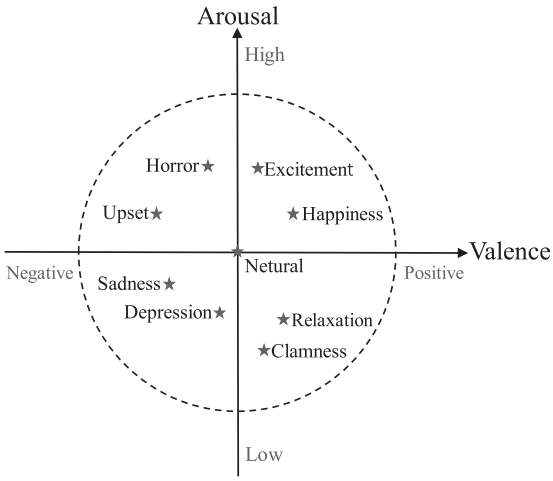


Fig. 1. The $2D$ valence-arousal model of emotions.

Two important points in the emotion recognition task are: (1) extraction of EEG features, and (2) construction of emotion classifiers [14]. Most traditional methods for extracting EEG features only focus on the information in time domain or/and frequency domain. Time-domain features are extracted by capturing the changes of time series with strong intuitiveness and clear physical meaning, including event related potentials (ERP) [15], statistical features (power, mean, standard deviation and the first difference) [16]–[19], Hjorth features [20], higher order crossings (HOC) [21], *etc*. The most commonly used frequency-domain features are power characteristics from different frequency bands. Usually, fast Fourier transform (FFT) [22] is utilized to transform the time-domain EEG signals into the frequency-domain, and Welch method is used to estimate corresponding power spectral density (PSD) [23]. Duan *et al.* [24] proposed the differential entropy (DE) to represent the state related to emotions, which proved to be more suitable for emotional classification than PSD [25]. For the construction of classifiers, traditional machine learning algorithms, typically as decision tree (DT), support vector machine (SVM), k-nearest neighbor (KNN) and multi-layer perceptron (MLP), are widely used and achieve good results. However, the generalization ability of these traditional methods is limited, which challenges the emotion recognition based on more complex EEG signals. In recent years, deep neural networks (DNNs) have developed rapidly and have been successfully adopted in many fields, (i.e.), computer vision, speech recognition, natural language processing, and biomedical signal processing. Many researchers have also successfully introduced them into the field of emotion recognition because they can automatically learn robust and abstract feature representations, achieving an improved performance than traditional algorithms. Alhagry *et al.* [26] proposed an end-to-end DNN-based method for identifying emotions from raw EEG signals. Long short-term memory-recurrent neural network (LSTM-RNN) was used to learn features from EEG signals, then the dense layer was used for classification. The recognition accuracies on the DEAP database were 85.45% and 85.65% for valence and arousal, respectively. Salama *et al.* [27] extracted $3D$ data representation from multi-channel EEG signals and send it to the proposed 3D convolutional neural network (3D-CNN) model for spatio-temporal feature extraction. They achieved the recognition accuracies of 87.44% and 88.49% for valence and arousal on DEAP database, respectively. Song *et al.* [28] proposed a novel dynamic graph convolutional neural network (DGCNN) that can dynamically learn the internal relationship between different EEG channels represented by an adjacency matrix to classify EEG emotions, and the accuracies were 86.23%, 84.54% and 85.02% for valence, arousal and dominance on DREAMER database, respectively.

Although DNNs show prominent advantages in EEG-based emotion recognition task, they still have some shortcomings. First, DNNs have lots of hyperparameters, and their learning ability depends heavily on the careful parameter adjustment. Second, DNNs require massive training data. However, in the field of EEG-based emotion recognition, large-scale labeled EEG databases are limited [29]. To solve the above problems, we adopt a deep forest model named multi-Grained Cascade Forest, termed as gcForest, and tailor it to multi-channel EEG-based emotion recognition task. The gcForest is a deep model proposed by Zhou *et al.* in 2017 [30], which has three similar characteristics as the DNNs: layer-by-layer processing, in-model feature transformation and sufficient model complexity. The gcForest algorithm has fewer hyperparameters and does not require backpropagation, as well as is robust to hyperparameter settings. In addition, its model complexity is data-dependent. It can terminate training adaptively, which makes it suitable for training with different sizes of data, not limited to large-scale one. For the field of EEG-based emotion recognition, where large-scale tag databases are limited due to high tagging cost, gcForest shows the advantage of better control of the training cost. Moreover, gcForest has been successfully utilized in the tasks of hyperspectral image classification [31], [32], metagenomic data classification [33], cancer subtypes classification [34], and protein self-interaction prediction [35].

In this paper, we propose a method for emotion recognition from multi-channel EEG via deep forest. First, we perform a pre-processing of baseline removal on the original artifact-eliminated EEG signals. Second, in order to utilize both the spatial relationship across channels and time information of the EEG signals, we map the pre-processed data of all channels at each moment to a $2D$ frame according to the electrode distribution, thereby obtaining a $2D$ frame sequence for each sample. Third, the scanning module in gcForest scans each frame in the sequence to obtain information across channels, and then we use the features obtained by all the frames. Finally, the feature vectors with spatial and temporal information are sent to cascade forest for classification. The contributions of this paper can be summarized as follows:

1) We tailor gcForest to multi-channel EEG-based emotional recognition according to the characteristics of multi-channel EEG, and a classification model that can fully use the spatial and temporal information of EEG signals is constructed.
2) Our method is data-driven, and the classification model does not rely on careful parameter adjustment, reducing

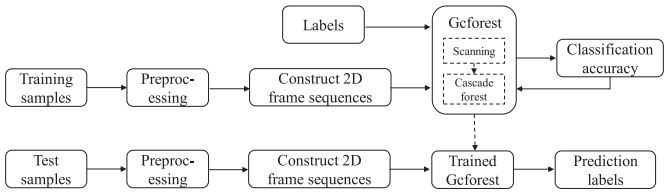


Fig. 2. Flowchart of the proposed method.

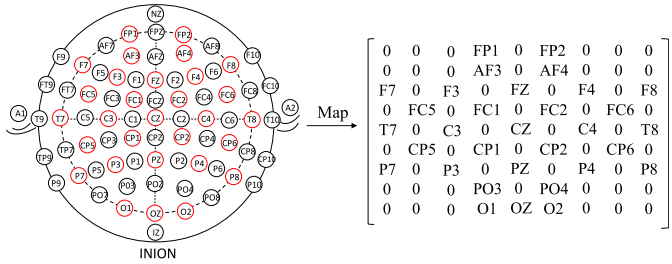


Fig. 3. The mapping of equivalence matrices according to the International 10–20 system [36].

the complexity of EEG emotion recognition. Besides, the model complexity of gcForest can be determined automatically for different size of training data, making it even suitable for small-scale training data.

3) Our method and seven other typical methods are validated on the DEAP and DREAMER databases. Our proposed method achieves the best performance except that of the valence dimension on DREAMER database, demonstrating the effectiveness of our method.

The rest of this paper is organized as follows: Section II introduces the formulation of the proposed method. Section III introduces the databsets used in the experiment, experimental setup and experimental results. Section IV gives profond discussions, and Section V concludes our study.

II. Methods

In this section, we will introduce the proposed method in detail, shown in Fig. 2. First, we introduce the preprocessing method of EEG data and the construction process of $2D$ frames. Then, we introduce the structure of two modules in gcForest: scanning and cascade forest. Finally, we introduce the specific process of tailoring gcForest to multi-channel EEG classification.

A. Pre-Processing

Most studies only use experimental signals without taking the effect of baseline signals (signals in relaxed state) into account. Since human EEG signals are unstable and susceptible to slight changes in surrounding environment, EEG signals produced by the same subject under the same stimulations are generally different. In addition, to a certain extent, the EEG signal produced by emotional material stimulation is affected by its emotional state before receiving stimulations. Thereby, the preprocessing of baseline removal can highlight the effects of stimulated emotions. It has been proved that the final classification results can be improved when adotping the amplitude difference between the experimental signal (induced EEG) and the rested EEG [37]. In this study, we divide the baseline signal and the experimental signal into K segments and I segments with a length of L, respectively. B_k and X_i represent the k-th baseline signal segment and the i-th experimental signal segment, respectively. Then we average all baseline signal segments and subtract this average value from each experimental signal segment. This step can be formulated as follow:

$$\overline{B} = \frac{\sum_{k=1}^{K} B_k}{K}, \tag{1}$$

$$X_i' = X_i - \overline{B}, \tag{2}$$

where $\overline{B}$ denotes the average of all baseline signal segments; X_i' denotes the i-th segment of final emotional state data.

B. The Construction of 2D Frame Sequences

The International 10-20 System is a recognized method for describing the placement of scalp electrodes during EEG acquisition. The "10" and "20" mean 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. We generalize the International 10-20 system together with the test electrodes used in the DEAP/DREAMER database to form a square matrix $(M \times M)$, where M equals to $max(P, Q)$. P is the number of electrode points included along the horizontal, while Q is the number of electrode points included along the vertical. The left side of Fig. 3 is a plain view of the International 10–20 system, where the EEG electrodes circled in red are the test points used in DEAP database, Fp1, AP2 and so on are the names of electrodes. Along the horizontal direction, the widest range of red dots is from T7 to T8 including nine electrodes. Vertically, the widest range of red dots is from Fp1 to O1 also including nine electrodes. Thereby, $P = Q = 9$ in this study and M is set to 9 for DEAP database. Similarly, M is also set to 9 for DREAMER database. Zeros represent channels that are not used in DEAP/DREAMER database.

The process of constructing $2D$ frame sequences is shown in Fig. 4. For a certain time index t, the EEG data from all channels is a $1D$ data vector $\boldsymbol{v}_t = [s_t^1, s_t^2, \ldots, s_t^N]^T$, where s_t^n is the pre-processed data for the n-th electrode channel and the acquisition system contains a total of N channels. For the DEAP and DREAMER databases, N is 32 and 14, respectively. During the time period $[t, t + L - 1]$, there are L $1D$ data vectors, and each of them contains N elements from all channels. For each $1D$ data vector, Z-score normalization [38] is used to normalize all elements by the following equation:

$$s_t^{n\prime} = \frac{s_t^n - \mu_{\boldsymbol{v}_t}}{\sigma_{\boldsymbol{v}_t}}, \tag{3}$$

where s_t^n denotes an element from one channel of the vector, $\mu_{\boldsymbol{v}_t}$ denotes the mean of all elements, and $\sigma_{\boldsymbol{v}_t}$ denotes the standard deviation of these elements.

In the EEG electrode map, each electrode is physically adjacent to multiple electrodes. In order to maintain spatial information between multiple adjacent channels, an $1D$ data vector

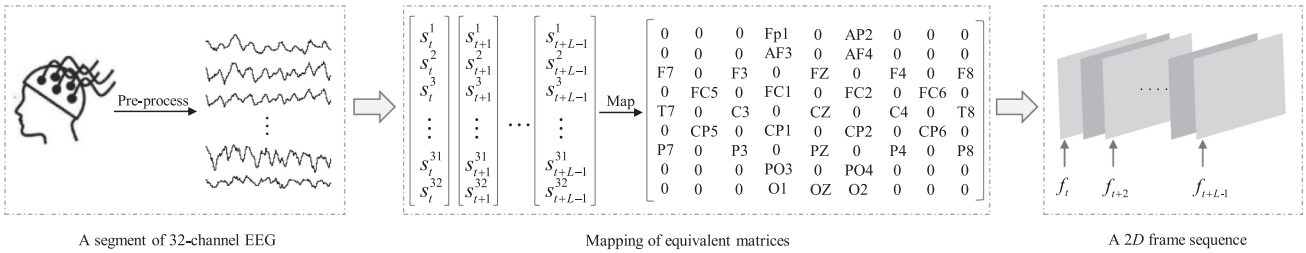


Fig. 4. The construction of $2D$ frame sequences.

is transformed into a $2D$ data frame according to the electrode map. The corresponding $2D$ data frame f_t of the $1D$ data vector $\boldsymbol{v}_t{}' = [s_t^{1'}, s_t^{2'}, \ldots, s_t^{N'}]^T$ at time index t is denoted as follows:

$$f_t = \begin{bmatrix} 0 & 0 & 0 & s_t^{1'} & 0 & s_t^{17'} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_t^{2'} & 0 & s_t^{18'} & 0 & 0 & 0 \\ s_t^{4'} & 0 & s_t^{3'} & 0 & s_t^{19'} & 0 & s_t^{20'} & 0 & s_t^{21'} \\ 0 & s_t^{5'} & 0 & s_t^{6'} & 0 & s_t^{23'} & 0 & s_t^{22'} & 0 \\ s_t^{8'} & 0 & s_t^{7'} & 0 & s_t^{24'} & 0 & s_t^{25'} & 0 & s_t^{26'} \\ 0 & s_t^{9'} & 0 & s_t^{10'} & 0 & s_t^{28'} & 0 & s_t^{27'} & 0 \\ s_t^{12'} & 0 & s_t^{11'} & 0 & s_t^{16'} & 0 & s_t^{29'} & 0 & s_t^{30'} \\ 0 & 0 & 0 & s_t^{13'} & 0 & s_t^{31'} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_t^{14'} & s_t^{15'} & s_t^{32'} & 0 & 0 & 0 \end{bmatrix}. \quad (4)$$

By this transformation, for the time period $[t, t+L-1]$, the normalized $1D$ vector sequence $[\boldsymbol{v}_t{}', \boldsymbol{v}_{t+1}{}', \ldots, \boldsymbol{v}_{t+L-1}{}']$ is converted into a $2D$ frame sequence $[f_t, f_{t+1}, \ldots, f_{t+L-1}]$ containing L frames.

C. Cascade Forest Structure

Inspired by the idea of layer-by-layer procedure of deep neural networks, cascade forest adopts a cascade structure of level-by-level procedure. Each level of the cascade forest receives feature information processed by the previous level and outputs its feature information to the next level. Each level is an ensemble of decision tree forests, and deep forest is an ensemble of some levels. To encourage diversity that is critical to the ensemble construction, each level contains two different types of forests: random forests [39] and completely-random forests [40]. The difference between a completely-random forest and a random forest lies in the strategy of selecting the splitted features of the tree. The completely-random forest selects one feature for split at each node randomly, while a random forest selects d features as candidates (d is the number of input features) and selects the splitting feature with the best $gini$ value. Each level contains m forests, each forest contains p decision trees, and each tree grows until each leaf node contains only instances of the same class or no more than q instances. Given an instance, the leaf nodes to which it belongs may contain training samples of the same class or different classes. Each forest can generate an estimation of the class distribution by calculating the percentage of different classes of training samples in the leaf node into which the instance falls, and then obtaining the average of all trees in the same forest. The estimated class distribution forms a class vector, and the class vectors of all forest outputs in each level will be concatenated and then joined to the original data as the next level input. As shown in the Fig. 5, assuming that each level contains two completely random forests and two random forests, for binary classification, each forest in each level will generate a two-dimensional class vector. Therefore, each level will generate 8 enhanced features.

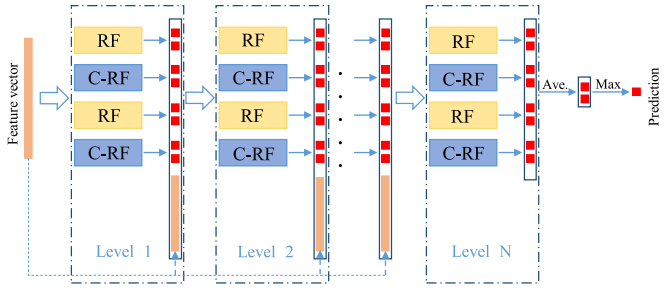


Fig. 5. Cascade forest structure (Suppose to predict two classes of emotions).

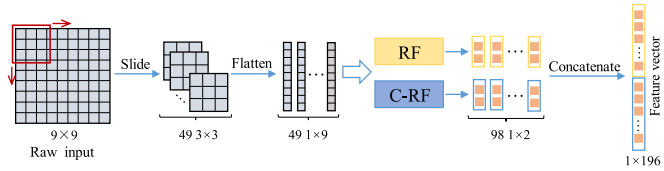


Fig. 6. Scanning structure (Suppose to predict two classes of emotions).

D. Scanning Structure

The scanning module is effective to process data with spatial relationships, so it is very suitable for processing multi-channel EEG data. As shown in the Fig. 6, a sliding window is used to scan the original input, then feature vectors are generated. Each feature vector will be assigned the same label as the original training sample, and then be regarded as an instance. All instances will be used to train a complete random forest, which generates class vectors. These class vectors are concatenated as the transformed features. For instance, assuming that the size of the original input is 9×9, the size of the sliding window is 3×3, and the sliding step is 1. For binary classification, each forest generates 49 two-dimensional class vectors, that is, a 196-dimensional transformed feature vector is finally generated.

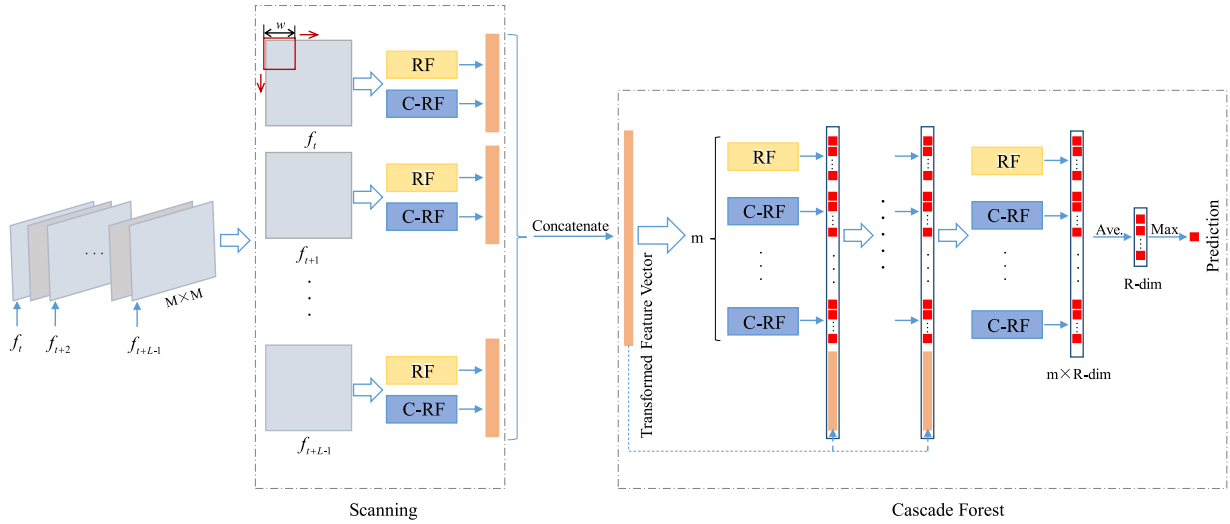


Fig. 7. GcForest for multi-channel EEG signals classification.

E. Multi-Channel EEG Signals Classification

The specific process of gcForest for emotion recognition from multi-channel EEG is described as below, shown in Fig. 7. According to the characteristic that the scanning module can process spatial information, we first send the EEG data into the scanning module to extract the spatial information across different EEG channels, and then feed the output of the scanning module into the cascade forest for binary classification.

According to aforementioned description, we transform each sample into a $2D$ frame sequence so that the size of the input with R emotions for scanning module is $S \times L \times M \times M$, where S represents the number of samples, L represents the number of frames contained in each frame sequence, and $M \times M$ represents the size of one frame. The size of the sliding window is $w \times w$, the sliding length is l, and the sliding window divides each frame into $((M-w)/l+1)((M-w)/l+1)$ sub-blocks. Each sub-block is flattened into a vector and sent to a completely random forest and a random forest, then a $((M-w)/l+1)((M-w)/l+1) \times 2 \times R$-dimensional transformed feature vector is generated. We concatenate these feature vectors generated from the L frames together as the final feature vector of this EEG segment, and then it will be sent to the cascade forest.

For each sample, m R-dimensional class vectors are generated by each level of cascaded forest. The m vectors are averaged to obtain a R-dimensional estimated distribution, and the sample will be classified as the one with the largest value. A validation set is utilized to verify the accuracy of the model with the current number of levels. If the current accuracy is higher than before, the m R-dimensional class vectors will be concatenated as a $m \times R$-dimensional enhanced feature with the transformed feature generated by scanning module, and then become the input of the next level. The number of layers stops increasing and the training stops until the accuracy no longer increases. This is different from DNNs, we don't need a fixed number of epochs. This means that each subject's model is not exactly the same and each subject's model can obtain the best result already available. Meanwhile, the training cost is effectively controlled.

TABLE I
DATABASE DESCRIPTION

Data format for each participant			
DEAP	array	array shape	array content
	Data	40×32×8064	videos×channels×data
	Labels	40×2	videos×label (V, A)
DREAMER	ExperData	18×14×25472 (M)	videos×channels×data
	BaseData	18×14×7808	videos×channels×data
	Labels	18×3	videos×label (V, A, D)

V denotes valence, A denotes arousal, and D denotes dimonance.

III. EXPERIMENTS

In this section, we first introduce the two public databases used in this study. Then, we explain the relevant experimental settings in detail. Finally, we present and analyze the experimental results on DEAP and DREAMER databases, respectively.

A. Databases

To test the performance of our method, we use two public databases widely used in EEG-based emotion recognition, DEAP [41] and DREAMER [42]. Table I shows the form of raw EEG data that we use in the two databases.

In DEAP database, 32 healthy participants (16 males, 16 females) participated in the experiment. Each participant was asked to watch 40 one-minute music videos. Their EEG signals were recorded from 32 electrodes that their positions according to the International 10-20 system. The sample rate of the raw EEG signal was 512 Hz. Participants rated the valence, arousal, dominance and liking on a continuous scale between 1 and 9 after watching each video. In the pre-processed version, EEG signals were down-sampled to 128 Hz, and EOG artifacts were removed with a blind source separation technique such as independent component analysis (ICA). Data recorded for each participant consists of 40 segments of EEG data and corresponding labels.

Each segment of EEG data contains 60 s experimental signals and 3 s baseline signals in a relaxed state.

In DREAMER database, 32 healthy participants (14 males, 9 females) participated in the experiment. 18 film clips were shown to every participant and each film clip targets one of nine emotions: entertainment, excitement, happiness, calmness, anger, disgust, fear, sadness, and surprise. The length of the film clips was between 65 to 393 s (M = 199 s). Their EEG signals were recorded with the sample rate of 128 Hz from 14 electrodes that their positions according to the International 10-20 system, and most eye artifacts were removed with linear phase FIR filters. Participants rated the valence, arousal and dominance from 1 to 5 after watching each film. Data recorded from each participant consists of three parts, 18 experimental signal segments, 18 baseline signal segments corresponding to relaxed state and 18 corresponding labels.

B. Experimental Design

Emotional recognition tasks based on EEG can be divided into subject-dependent and subject-independent ones. In this paper, we focused on subject-dependent EEG-based emotion recognition task.

Selection of window length: [43] indicates that the duration of emotion is about 0.5-4 s, and the experiments in [44] show that EEG data divided into 1 s segments can obtain the highest classification accuracy. Thus, we set the window length to 1 s.

Data processing: In DEAP database, for each subject, the raw EEG data is segmented into 2400 samples with the size of 32×128 using the non-overlapping window. We construct a $2D$ time frame sequence for each sample, and eventually 2400 samples with the size of $128 \times 9 \times 9$ can be obtained for each subject. In DREAMER database, for each subject, the raw EEG data is segmented into 3728 samples with the size of 14×128 using the same widowing technique. We also construct a $2D$ time frame sequence for each sample, and eventually 3728 samples with the size of $128 \times 9 \times 9$ are obtained for each subject.

Label processing: We label each segment of the signal divided by the sliding window with the same label as the whole segment of the signal. In DEAP database, we divide the score of 1-9 into two binary classification problems with a threshold of 5: high/low arousal and high/low valence (low: ≤ 5, high: > 5). In DREAMER database, we divide the 1-5 score into three binary classification problems with a threshold of 3: high/low arousal, high/low valence, and high/low dominance (low: ≤ 3, high: > 3).

Parameters settings in gcForest: GcForest has six parameters: m, n, $q1$, $q2$, w, and l; m represents the number of forests that each level contains, n represents the number of trees that each forest contains, $q1$ and $q2$ represent the number of instances that each tree contains when the leaf node stop growing in scanning module and cascade forest module respectively, $w \times w$ represents the size of the sliding window of the scanning module, and l represents the step length of the sliding window. Through a lot of experiments, we found that a large range of changing of $q1$ and $q2$ has a slight influence on the results, so we set them as 20 and 10, consistent with the default settings. For $w \times w$, the maximum setable value is 9×9, which is the same as the size of a $2D$ frame fed into scanning module. Usually, the 3×3 kernel is widely uesd in CNN. However, in this study, the non-zero elements in $2D$ frames constructed by gcForce are sparse. In order to capture sufficient spatiol information, we set $w \times w$ to 6×6. The default setting for l is 1, and the running time of the proposed model will decrease with the increase of l. Through experiments, it is found that when l is 3, the recognition accuracy is close to that when l is 1. In order to achieve a balance between information richness and computing resources, we set l to 3. m and n were determined by the classification results varying with the change of the corresponding numbers. For DEAP database, we experimentally set them to 220 and 8, respectively; For DREAMER database, we experimentally set them to 260 and 24, respectively. The detailed analysis will be released in the following section.

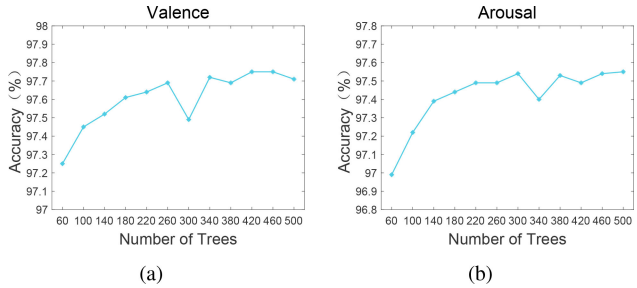


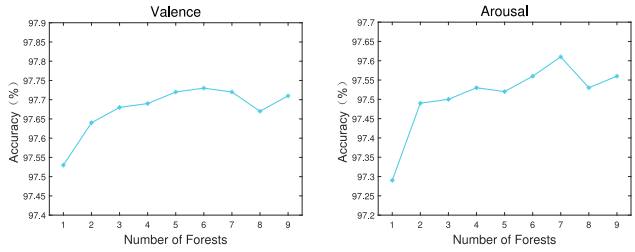
Fig. 8. Accuracy varies with the number of trees in gcForest on DEAP: (a) Valence, (b) Arousal.

Division of training/test sets: 10-fold cross-validation [45] is implemented for both databases.

C. Result on DEAP

For DEAP database, we have verified the proposed method on valence and arousal dimensions in order to facilitate comparison with previous work.

For the setting of m and n: the number of trees in each forest and the number of forests in each level of cascade forest, we have observed the impact of the number change on the classification results through experiments. The classification accuracy on DEAP database varies with the change of number of trees in each forest, shown in Fig. 8. When the number of trees in each forest is small, accuracy increases obviously with the increase of number of trees; when the number of trees increasesd to 180, the increase of accuracy becomes relatively gentle; when the number of trees is large, the accuracy will fluctuate slightly with the increase of trees. Besides, it can be seen from Fig. 8 that the number change of trees has a slight influence on the classification accuracy, and the difference between the maximum and minimum values on valence and arousal dimensions are only 0.50% and 0.56%, respectively. The classification accuracy on DEAP database varies with the change of number of forests in cascade forest as shown in Fig. 9, and each increase of '1' on the horizontal axis coordinates means adding a random forest and a completely random forest. When the horizontal axis increases from 1 to 2, the accuracy increases sharply; after that, as the number of



(a) (b)

Fig. 9. Accuracy varies with the number of forests in gcForest on DEAP: (a) Valence, (b) Arousal.

TABLE II
AVERAGE ACCURACIES AND STANDARD DEVIATIONS (%) OF DIFFERENT METHODS ON DEAP DATABASE

Method	Valence	Arousal
DT	68.28±4.12	71.16±6.12
SVM	86.60±6.98	87.43±6.62
MLP	87.73±6.30	88.88±5.08
Conti-CNN	89.45±4.51	90.42±3.72
CNN-RNN	89.92±3.01	90.81±2.99
CRAM	85.54±7.86	83.65±7.34
DGCNN	92.55±3.53	93.50±3.93
Ours	**97.69±1.22**	**97.53±1.52**

forests increases, the accuracy slowly increases and tends to be stable. The difference between the maximum and minimum values on valence and arousal dimensions are only 0.20% and 0.32%, respectively. To strike a balance between classification accuracy and computational resources, we set the number of trees in each forest to 220 for DEAP database, respectively; the number of forests in each level of cascade forest to 8 (4 random forests and 4 completely random forests). Thus, recognition accuracies of our method for valence and arousal dimensions on DEAP database are 97.69% and 97.53%, respectively.

We compare our method with seven other typical methods, namely DT [46], SVM [47], MLP [36], Conti-CNN [36], CNN-RNN [37], CRAM [48], and DGCNN [28]. Conti-CNN takes $3D$ EEG cubes as the input, aiming to extract both spatial and frequency information. CNN-RNN is a hybrid neural network that classifies emotional states by learning spatial-temporal representation of $2D$ EEG frames. CRAM encodes the spatio-temporal information of the temporal slices by a specifically designed convolutional network, then extractes the attentive temporal dynamics of EEG temporal slices for successful emotions classification. These seven methods use the same pre-processing step of baseline removal, the same slice length and the same division of training/test set as our method, which ensures the fairness of comparison experiment. Except CNN-RNN and CRAM, the other five methods need to extract DE features from four frequency bands. For DT, SVM, MLP, and DGCNN, their input are 2400 feature vectors with the size of 128×1. For Conti-CNN, its input is 2400 $3D$ data with the size of $4 \times 9 \times 9$, for CNN-RNN, its input is 2400 $3D$ data with the size of $128 \times 9 \times 9$, and for CRAM, its input is 2400 $2D$ data with the size of 128×32. Table II shows the mean accuracies and standard deviations of 32 subjects with these methods. The results show that compared with the other seven methods, our

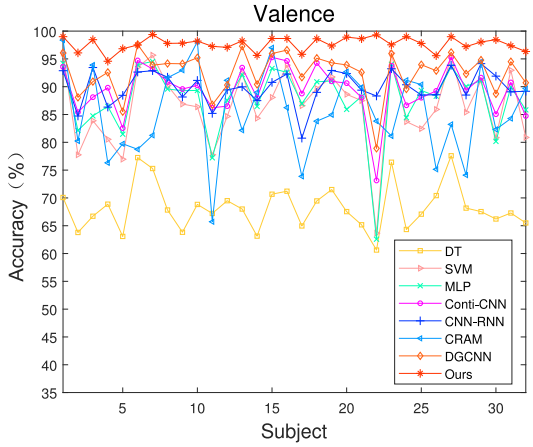


Fig. 10. Performance comparison of each subject using different methods for valence on DEAP database.

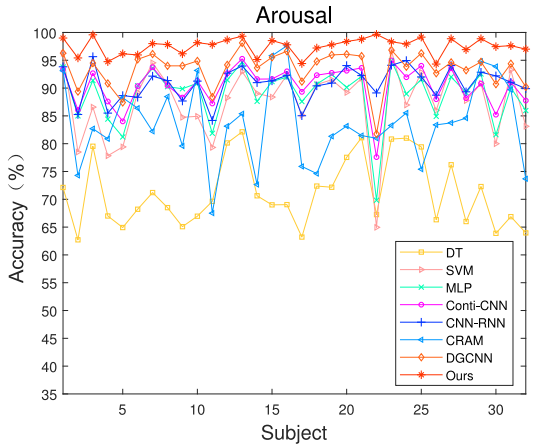


Fig. 11. Performance comparison of each subject using different methods for arousal on DEAP database.

method achieves the best performance for both dimensions, with an improvement of 5.14% and 4.03% compared to the second best results. In terms of standard deviation, among the four DNN-based methods, CRAM has much larger standard deviations than those of the three other methods. The standard deviations corresponding to valence and arousal experiments of our method are the smallest compared with all other methods, demonstrating the high stability. In a word, our method has significant superiority in both average accuracy and standard deviation. Fig. 10 and Fig. 11 show the recognition accuracy of the two dimensions of valence and arousal with the eight methods for each subject, and the accuracy for each subject is an average of ten folds. It can be seen that our method has achieved the best performance for each subject, demonstrating the effectiveness of our method.

D. Result on DREAMER

For DREAMER database, we have verified the proposed method in valence, arousal and dominance dimensions.

The classification accuracy on DREAMER database varies with the change of the number of trees in each forest and the

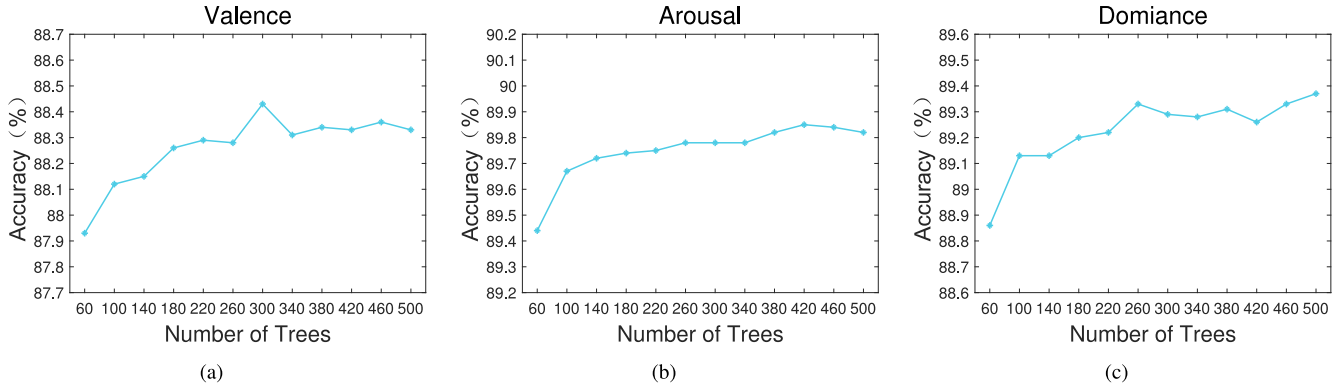


Fig. 12. Accuracy varies with number of trees in gcForest on DREAMER: (a) Valence, (b) Arousal, and (c) Dominance.

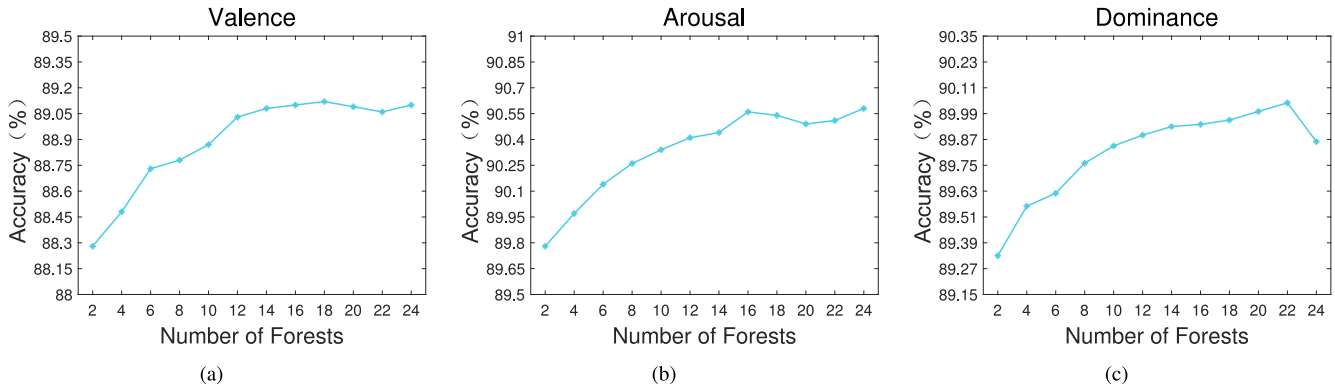


Fig. 13. Accuracy varies with number of forests in gcForest on DREAMER: (a) Valence, (b) Arousal, and (c) Dominance.

number of forests in each level of cascade forest, shown in Fig. 12 and 13, respectively. The variation trend of accuracy is consistent with that of DEAP database. It can be seen from Fig. 12 that the difference between the maximum and minimum values on vanlence, arousal and dominance dimensions are only 0.50%, 0.41% and 0.51%, respectively; and it can be seen from Fig. 13 that the difference between the maximum and minimum values on vanlence, arousal and dominance dimensions are only 0.84%, 0.80% and 0.71%, respectively. To strike a balance between classification accuracy and computational resources, we set the number of trees in each forest to 260; the number of forests in each level of cascade forest to 24 (12 random forests and 12 completely random forests). Thus, recognition accuracies of our method on valence, arousal and dominance dimensions on DREAMER database are 89.03%, 90.41% and 89.89% respectively.

Similar to DEAP database, our method was also compared with the above-mentioned seven methods. For DT, SVM, and DGCNN, their input are 3728 feature vectors with 56 features. For Conti-CNN, its input is 3728 $3D$ data with size of $4 \times 9 \times 9$, for CNN-RNN its input is 3728 $3D$ data with size of $128 \times 9 \times 9$, and for CRAM, its input is 2400 $2D$ data with the size of 128×14. Table III shows the mean accuracy and standard deviation of 23 subjects with these methods. The results show that our method has obtained comparable recognition accuracy with DGCNN, and much higher recognition accuracy than other methods. Different from DEAP, the traditional machine learning algorithm SVM performes better than three DNN-based methods. Compared with the excellent SVM, our method is still about 3% higher in accuracy. In terms of standard deviations, our method has obtained comparable results with the comparison methods. Figs. 14, 15, and 16 show the recognition accuracy of each subject on the three dimensions of valence, arousal and dominance with these methods, and the accuracy of each subject is an average of ten folds. It can be seen that our method has obtained the highest recognition accuracy for about half of the subjects.

TABLE III
AVERAGE ACCURACIES AND STANDARD DEVIATIONS (%) OF DIFFERENT METHODS ON DREAMER DATABASE

Method	Valence	Arousal	Dominance
DT	75.53±6.71	75.74±6.44	76.40±5.68
SVM	87.14±5.20	87.03±4.88	87.18±4.87
MLP	83.64±5.97	83.71±5.39	83.90±5.32
Conti-CNN	84.54±5.00	84.84±4.86	85.05±4.96
CNN-RNN	79.93±6.65	81.48±6.33	80.94±5.66
CRAM	79.71±8.20	79.61±8.26	77.99±8.08
DGCNN	**89.59±5.13**	88.93±3.93	88.63±5.13
Ours	89.03±5.56	**90.41±5.33**	**89.89±6.19**

E. Small-Scale Training Data Experiment

In order to verify that the proposed method is different from deep neural networks that rely on a large amount of training data,

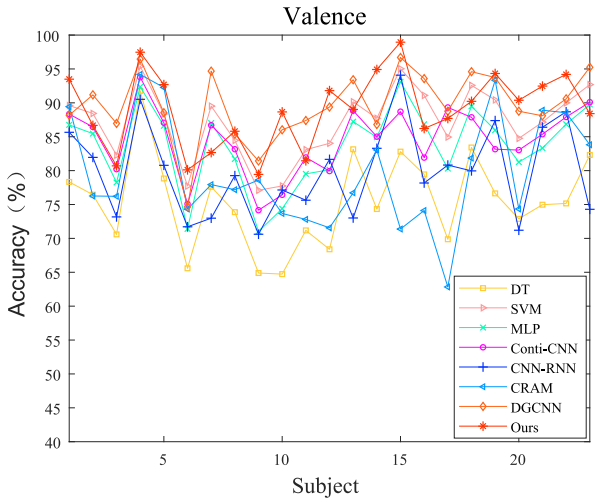


Fig. 14. Performance comparison of each subject using different methods for valence on DREAMER database.

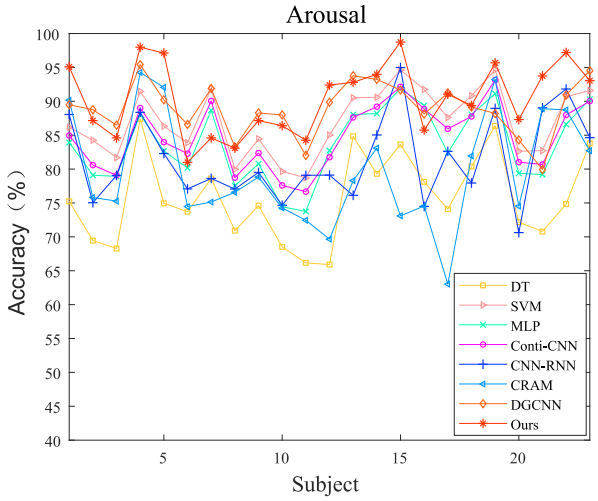


Fig. 15. Performance comparison of each subject using different methods for arousal on DREAMER database.

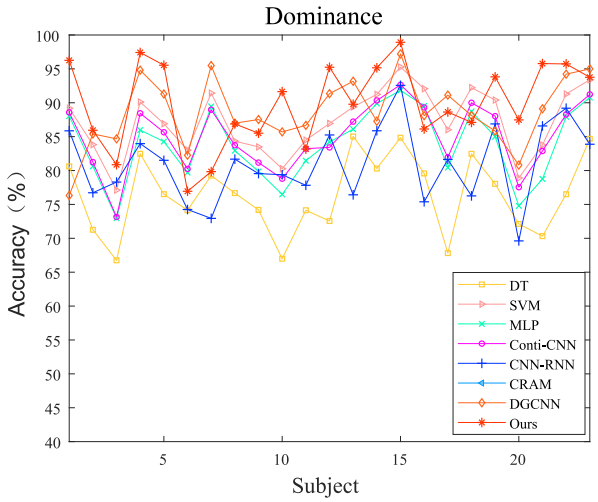


Fig. 16. Performance comparison of each subject using different methods for dominance on DREAMER database.

we perform small training data experiments with the proposed method and all comparison methods on both databases. We set the number of training data to 5%, 10%, 20%, 30%, 40%, and 50% of the total samples to observe the performance of each method. Table IV and V show the accuracies and standard deviations of each method on DEAP and DREAMER databases as the size of the training data changes, respectively. The training data for each experiment is randomly selected from the entire sample of the subject, and each result presented is the average result of all subjects repeating the experiment ten times. We can observe that no matter deep learning based or traditional methods, the recognition accuracy declines to a certain extent with the reduction of the proportion of training data. For DEAP, the proposed method performs best under all conditions. Moreover, when the proportion of training data is 30%, the accuracy of the proposed method is higher than these of all comparison methods under the condition of 10-fold cross validation (90%). For DREAMER, when the proportion of training data is 40% and 50%, the recognition accuracy for valence of the proposed method is only 0.49% and 0.66% less than those of DGCNN, respectively; under other conditions, the proposed method has the highest accuracy. These verify the effectiveness of the proposed method under small-scale training data conditions.

F. Running Time

In our experiment, DT, SVM, and the proposed method are trianed on a INTEL i7-7800X CPU, MLP, Conti-CNN, CNN-RNN, CRAM, and DGCNN are trained on a NVIDIA GPU with TensorFlow framework. The training time of each method is shown in Table VI. It can be seen that feature-driven methods have advantages over data-driven methods in terms of training time. The training time of the proposed method using CPU is only half of that of data-driven method CRAM. In addition, gcForest is a parallel ensemble method, which shows a significant superiority for large-scale parallel implementation.

IV. Discussions

EEG-based emotion recognition is a hot issue in the field of HCI in recent years, and many researchers have proposed effective classification models to achive the good results. However, there are still some limitations. For instance, most methods require artificial feature extraction before being sent to classifiers, which not only limits the usage of original information, but also ignores the spatial characteristics of brain regions. This study aims to find a more effective and simpler method that is data-driven to realize the purpose of multi-channel EEG-based emotion recognition. In this paper, gcForest model is tailored to multi-channel EEG-based emotion recognition from the perspective of fully considering the spatial-temporal information.

We compare the proposed method with seven other methods, among which DT, SVM and MLP are traditional machine learning algorithms, Conti-CNN, CNN-RNN, CRAM, and DGCNN are models based on DNN. Experimental results in Table II and III verify that the proposed model is more effective. Our method has higher recognition accuracy on both DEAP and DREAMER databases, especially with a huge lead on DEAP. We also conduct

TABLE IV
AVERAGE ACCURACIES AND STANDARD DEVIATIONS (%) OF SMALL-SCALE SAMPLE EXPERIMENT WITH EACH METHODS ON DEAP DATABASE

Train/Test sample		DT	SVM	MLP	Conti-CNN	CNN-RNN	CRAM	DGCNN	Ours
5% / 95%	valence	58.63±3.50	63.80±5.80	59.19±3.68	69.89±5.29	62.81±3.22	65.95±5.24	69.19±5.54	**77.44±4.58**
	arousal	61.92±6.30	68.65±7.88	62.27±5.84	72.00±6.20	65.67±4.58	68.92±7.23	71.49±6.81	**80.03±4.25**
10% / 90%	valence	60.84±3.79	68.45±6.07	63.09±4.03	75.12±5.17	68.54±3.25	71.63±5.50	74.84±5.30	**87.40±3.70**
	arousal	63.85±6.28	72.00±7.17	66.75±6.18	76.66±5.33	70.92±4.31	73.42±5.85	76.27±5.91	**88.04±3.63**
20% / 80%	valence	62.95±3.80	74.21±6.71	76.77±5.62	80.46±4.83	74.73±2.88	76.79±5.77	80.59±5.36	**92.89±2.94**
	arousal	66.21±6.28	76.69±6.65	79.16±5.98	82.09±4.63	76.76±3.96	77.41±5.98	82.02±5.40	**93.06±2.86**
30% / 70%	valence	64.36±3.71	77.68±7.08	82.44±6.11	83.47±4.76	78.52±3.02	79.51±6.17	84.81±5.37	**94.89±2.09**
	arousal	67.46±6.32	79.65±6.65	84.00±5.63	84.13±4.58	80.33±3.54	79.57±6.14	85.50±4.93	**95.11±2.50**
40% / 60%	valence	65.37±3.75	80.28±7.49	84.68±5.89	85.37±4.49	81.39±2.95	81.49±6.54	85.35±5.40	**95.85±1.83**
	arousal	68.49±6.13	81.81±6.69	86.46±5.33	86.39±4.47	83.19±3.29	80.81±6.52	85.50±4.93	**95.85±2.23**
50% / 50%	valence	66.26±3.83	82.13±7.27	86.07±5.59	86.79±4.56	85.38±3.12	82.98±6.87	88.52±5.17	**96.30±1.80**
	arousal	69.32±6.02	83.30±6.79	87.84±4.88	87.63±4.24	86.57±3.40	81.57±6.96	86.69±4.41	**96.34±2.14**

TABLE V
AVERAGE ACCURACIES AND STANDARD DEVIATIONS (%) OF SMALL-SCALE SAMPLE EXPERIMENT WITH EACH METHODS ON DREAMER DATABASE

Train/Test sample		DT	SVM	MLP	Conti-CNN	CNN-RNN	CRAM	DGCNN	Ours
5% / 95%	valence	67.59±5.53	77.07±5.70	65.87±5.62	75.00±6.07	67.84±6.78	71.30±6.69	75.99±5.49	**77.62±6.59**
	arousal	67.71±6.09	77.00±5.76	67.38±6.18	75.40±5.74	69.45±7.01	72.37±7.38	76.88±6.02	**78.91±6.41**
	diminance	68.56±5.43	77.74±6.38	67.94±5.80	75.79±6.02	68.52±6.45	72.01±6.55	76.60±6.18	**78.25±6.06**
10% / 90%	valence	70.00±5.62	80.33±5.70	79.40±5.86	78.24±5.72	69.43±6.58	73.37±7.31	80.64±5.94	**81.42±6.48**
	arousal	69.73±6.27	80.21±5.67	79.43±5.52	78.34±5.62	71.71±7.11	74.71±7.25	80.15±6.14	**82.83±5.99**
	diminance	70.46±5.71	80.65±5.80	77.79±5.66	78.89±5.77	71.76±6.42	75.29±6.93	80.79±5.40	**81.71±5.95**
20% / 80%	valence	71.97±6.03	82.81±5.66	82.31±5.77	81.20±5.66	72.96±6.83	75.20±7.76	83.37±5.05	**84.19±6.25**
	arousal	72.04±6.42	82.92±5.39	82.41±5.40	81.41±5.50	74.13±6.90	76.95±8.98	84.23±5.49	**85.89±5.87**
	diminance	72.60±5.60	83.20±5.44	82.78±5.48	81.49±5.38	73.43±6.51	76.53±6.91	84.14±5.40	**84.91±6.27**
30% / 70%	valence	73.14±6.21	84.19±5.48	83.58±5.53	86.64±5.61	77.43±6.83	76.34±7.77	84.98±5.02	**85.59±5.99**
	arousal	73.03±6.44	84.14±5.34	83.56±5.15	82.80±5.55	79.03±6.85	77.40±9.54	86.03±4.62	**87.28±5.73**
	diminance	73.73±5.62	84.43±5.19	83.91±4.97	82.85±5.45	78.36±6.24	77.70±7.25	84.84±4.82	**86.47±6.26**
40% / 60%	valence	73.60±6.31	85.08±5.46	84.15±5.30	83.44±5.49	79.26±6.76	76.84±7.74	**87.05±5.41**	86.56±5.90
	arousal	73.78±6.53	85.04±5.68	84.00±5.02	83.54±5.20	80.67±6.76	78.65±9.04	86.97±4.91	**88.22±5.56**
	diminance	74.48±5.83	85.17±5.10	84.28±5.03	83.69±5.31	80.23±6.41	78.62±7.44	84.11±7.04	**87.46±6.23**
50% / 50%	valence	74.40±6.44	85.77±5.34	84.16±5.28	84.04±5.64	80.42±6.52	77.09±7.45	**87.89±4.98**	87.23±5.81
	arousal	74.43±6.53	85.56±5.21	84.25±4.92	84.24±5.28	81.93±6.30	78.99±9.55	88.58±4.45	**88.95±5.44**
	diminance	75.06±5.84	85.75±5.10	84.31±5.12	84.30±5.21	81.18±6.25	78.40±7.87	85.46±6.70	**88.20±6.26**

TABLE VI
RUNNING TIME OF VARIOUS METHODS ON BOTH DATABASES

Method		DT	SVM	MLP	Conti-CNN	CNN-RNN	CRAM	DGCNN	Ours
time (s)	DEAP	0.2928	0.7508	33.7596	12.6405	656.3955	1494.8705	7.0225	693.4861
	DREAMER	0.4548	1.4115	52.3274	20.6040	602.0744	2340.8621	10.2529	1307.406

small-scale training data experiments on the two databases, and the results are shown in Table IV and V. From Table II and Table IV, we can observe that the proposed method achives the best performance under all circumstances. When the proportion of training data is 10%, 20%, 30%, 40%, and 50%, the advantages of the proposed method are more obvious than that of the 10-fold cross validation (90%). Specifically, for valence, when 10-fold cross-validation is performed, the accuracy of the proposed method is 8.24%, 7.77%, 12.15%, and 5.14% higher than those of the other four DNNs-based methods respectively; and when the proportion of training data is 10%, they become 12.28%, 18.86%, 15.77%, and 12.56%. From Table III and Table V, we can observe that the proposed method achieves similar results with DGCNN when 10-fold cross-validation is performed; the accuracy of the proposed method is 1.63%, 2.03%, and 1.65% higher than those of it while the proportion of training data is 5% for valence, arousal, and dominance, respectively. These indicate that the proposed method has more obvious advantages in the case of small-scale training data. Compared with DNNs, this method does not require a large number of training samples, which is very meaningful for some fields where large-scale data is limited. In addition, we demonstrate the influence of the changes of two important parameters (the number of trees in each forest and the number of forests in each level of cascade forest) in our model on the recognition accuracy. It can be seen from Fig. 8, Fig. 9, Fig. 12 and Fig. 13 that the proposed method has a slight impact on the recognition accuracy with the change of parameters. This indicates that the proposed method is insensitive to parameter setting, which is different from DNNs-based methods that require careful parameter adjustment. Moreover, Table VI shows the training time for each method. The time consumption of traditional methods are generally lower than those of DNN-based methods. Among DNN-based methods, the running time of the proposed method is moderate, which indicates that the proposed method can achieve a better balance between computational resources and recognition accuracy.

During the experiment, we find that Conti-CNN and CNN-RNN, which outperformed SVM on DEAP, did not perform well

on DREAMER. In particular, CNN-RNN is about 6% lower in accuracy than SVM. In addition, our method has a relatively smaller advantage on DREAMER compared with that on DEAP. Among all the eight methods, Conti-CNN, CNN-RNN and our proposed method all can capture spatial information from EEG segments by the step of constructing $2D$ frames according to the spatial distribution of electrodes. The difference lies in whether the elements in the frames are EEG data or EEG features. Specifically, CNN-RNN uses original EEG signals and considers the positions of electrodes to form $2D$ EEG frames, Conti-CNN uses DE features and considers the position of electrodes to form $3D$ EEG cubes that generated from $2D$ frames, and our proposed method constructs $2D$ EEG frames using original EEG signals. One possible reason may be that the channel number of DEAP database is 32, while that of DREAMER is only 14, which indicates that when the channel number is larger, each $2D$ frame has less zero elements, and more spatial information can be provided using sub-block extraction technique, i.e., the scanning process of our method and the convolution process of CNN. Therefore, our method may be more suitable for EEG data with a sufficient number of channels.

V. Conclusion

In this paper, we propose a method of gcForest for multi-channel EEG-based emotion recognition, which reduces the complexity of emotion recognition compared with DNNs. Scanning module in gcForest is used to capture the spatial information of EEG signal, and the cascade forest abstracts the feature vectors with spatial-temporal information generated by the scanning module and classifies them. Subject-dependent experiments on DEAP and DREAMER databases have been conducted. Experimental results have indicated that our method achieves a better recognition performance than other state-of-the-art methods with a moderate consumption of computational resources. Meanwhile, the recognition results on both databases vary with parameters adjustment are within 1% for all dimensions, demonstrating that our method is robust to parameter settings. Additionally, our method also achieves an outstanding performance under small-scale training data conditions. Currently, this study is limited to subject-dependent classification task. In the future, we will focus on subject-independent emotion classification tasks based on deep forest with domain adaptation methods.

References

[1] J. J. Gross and R. A. Thompson, "Emotion regulation: Conceptual foundations." *Handbook of emotion regulation*, New York: Guilford Press, 2017.

[2] L. Piho and T. Tjahjadi, "A mutual information based adaptive windowing of informative eeg for emotion recognition," *IEEE Trans. Affect. Comput.*, 2018, doi: 10.1109/TAFFC.2018.2840973.

[3] H. Chao, H. Zhi, L. Dong, and Y. Liu, "Recognition of emotions using multichannel eeg data and dbn-gc-based ensemble deep learning framework," *Comput. Intell. Neurosci.*, vol. 2018, 2018, Art. no. 9750904.

[4] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil, "Review and classification of emotion recognition based on EEG brain-computer interface system research: a systematic review," *Appl. Sci.*, vol. 7, no. 12, 2017, Art. no. 1239.

[5] Y. Zhao, X. Cao, J. Lin, D. Yu, and X. Cao, "Multimodal emotion recognition model using physiological signals," 2019, *arXiv:1911.12918*.

[6] W. Heller, "Neuropsychological mechanisms of individual differences in emotion, personality, and arousal," *Neuropsychology*, vol. 7, no. 4, pp. 476–489, 1993.

[7] C. Li, W. Tao, J. Cheng, Y. Liu, and X. Chen, "Robust multichannel eeg compressed sensing in the presence of mixed noise," *IEEE Sensors J.*, vol. 19, no. 22, pp. 10 574–10 583, Nov. 2019.

[8] Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi, "Real-time movie-induced discrete emotion recognition from eeg signals," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 550–562, Oct./Dec. 2018.

[9] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "EEG-based emotion recognition during watching movies," in *Proc. 5th Int. IEEE/EMBS Conf. Neural Eng.*, 2011, pp. 667–670.

[10] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," *IEEE Trans. Syst., Man, Cybern.-Part A: Syst. Humans*, vol. 41, no. 6, pp. 1052–1063, Nov. 2011.

[11] R. Khosrowabadi, C. Quek, K. K. Ang, and A. Wahab, "Ernn: A biologically inspired feedforward neural network to discriminate emotion from eeg signal," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 609–620, Mar. 2014.

[12] Y. Liu, O. Sourina, and M. K. Nguyen, "Real-time eeg-based emotion recognition and its applications," in *Transactions on Computational Science XII*, Berlin, Germany: Springer, 2011, pp. 256–277.

[13] P. Ozel, A. Akan, and B. Yilmaz, "Synchrosqueezing transform based feature extraction from eeg signals for emotional state prediction," *Biomed. Signal Process. Control*, vol. 52, pp. 152–161, 2019.

[14] Y. Li, J. Huang, H. Zhou, and N. Zhong, "Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks," *Appl. Sci.*, vol. 7, no. 10, 2017, Art. no. 1060.

[15] C. A. Frantzidis, C. Bratsas, C. L. Papadelis, E. Konstantinidis, C. Pappas, and P. D. Bamidis, "Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli," *IEEE Trans. Inf. Technol. Biomedicine*, vol. 14, no. 3, pp. 589–597, May 2010.

[16] K. Takahashi and A. Tsukaguchi, "Remarks on emotion recognition from multi-modal bio-potential signals," in *Proc. SMC'03 Conf. IEEE Int. Conf. Syst., Man Cybern. Conf. Theme-Syst. Secur. Assurance (Cat. No. 03CH37483)*, 2003, vol. 2, pp. 1654–1659.

[17] X.-W. Wang, D. Nie, and B.-L. Lu, "Eeg-based emotion recognition using frequency domain features and support vector machines," in *Proc. Int. Conf. Neural Inf. Process*, 2011, pp. 734–743.

[18] Y. Liu and O. Sourina, "Real-time fractal-based valence level recognition from eeg," in *Transactions on Computational Science XVIII*, Berlin, Germany: Springer, 2013, pp. 101–120.

[19] M. Murugappan, N. Ramachandran, and Y. Sazali, "Classification of human emotion from eeg using discrete wavelet transform," *J. Biomed. Sci. Eng.*, vol. 3, no. 04, pp. 390–396, 2010.

[20] B. Hjorth, "Eeg analysis based on time domain properties," *Electroencephalography Clin. Neurophysiology*, vol. 29, no. 3, pp. 306–310, 1970.

[21] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from eeg using higher order crossings," *IEEE Trans. Inf. Technol. Biomedicine*, vol. 14, no. 2, pp. 186–197, Mar. 2010.

[22] H. J. Nussbaumer, "The fast fourier transform," in *Fast Fourier Transform and Convolution Algorithms*, Berlin, Germany: Springer, 1981, pp. 80–111.

[23] J. F. D. Saa and M. S. Gutierrez, "Eeg signal classification using power spectral features and linear discriminant analysis: A brain computer interface application," in *Proc. Eighth Latin Amer. Caribbean Conf. Eng. Technol*, 2010, pp. 1–7.

[24] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for eeg-based emotion classification," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng.*, 2013, pp. 81–84.

[25] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[26] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on eeg using lstm recurrent neural network," *Emotion*, vol. 8, no. 10, pp. 355–358, 2017.

[27] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. W. Shalaby, "Eeg-based emotion recognition using 3d convolutional neural networks," *Int. J. Adv. Comput. Sci. Appl*, vol. 9, no. 8, pp. 329–337, 2018.

[28] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, 2018, doi: 10.1109/TAFFC.2018.2817622.
[29] J. Chen, P. Zhang, Z. Mao, Y. Huang, D. Jiang, and Y. Zhang, "Accurate eeg-based emotion recognition on combined features using deep convolutional neural networks," *IEEE Access*, vol. 7, pp. 44 317–44 328, 2019.
[30] Z.-H. Zhou and J. Feng, "Deep forest: Towards an alternative to deep neural networks," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 3553–3559.
[31] X. Liu, R. Wang, Z. Cai, Y. Cai, and X. Yin, "Deep multigrained cascade forest for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8169–8183, Oct. 2019.
[32] X. Cao, R. Li, Y. Ge, B. Wu, and L. Jiao, "Densely connected deep random forest for hyperspectral imagery classification," *Int. J. Remote Sens.*, vol. 40, no. 9, pp. 3606–3622, 2019.
[33] Q. Zhu, Q. Zhu, M. Pan, X. Jiang, X. Hu, and T. He, "The phylogenetic tree based deep forest for metagenomic data classification," in *Proc. IEEE Int. Conf. Bioinf. Biomedicine*, 2018, pp. 279–282.
[34] Y. Guo, S. Liu, Z. Li, and X. Shang, "Bcdforest: a boosting cascade deep forest model towards the classification of cancer subtypes based on gene expression data," *BMC Bioinf.*, vol. 19, no. 5, pp. 1–13, 2018.
[35] Z.-H. Chen, L.-P. Li, Z. He, J.-R. Zhou, Y. Li, and L. Wong, "An improved deep forest model for predicting self-interacting proteins from protein sequence using wavelet transformation," *Frontiers Genetics*, vol. 10, 2019, Art. no. 90.
[36] Y. Yang, Q. Wu, Y. Fu, and X. Chen, "Continuous convolutional neural network with 3d input for EEG-based emotion recognition," in *Proc. Int. Conf. Neural Inf. Process.*, 2018, pp. 433–443.
[37] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, "Emotion recognition from multi-channel eeg through parallel convolutional recurrent neural network," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2018, pp. 1–7.
[38] E. I. Altman *et al.*, "Predicting financial distress of companies: revisiting the z-score and zeta models," *Stern School Bus., New York Univ.*, vol. 5, pp. 9–12, 2000.
[39] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[40] F. T. Liu, K. M. Ting, Y. Yu, and Z.-H. Zhou, "Spectrum of variable-random trees," *J. Artif. Intell. Res.*, vol. 32, pp. 355–384, 2008.
[41] S. Koelstra *et al.*, "Deap: A database for emotion analysis; using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan.–Mar. 2012.
[42] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 1, pp. 98–107, Jan. 2018.
[43] V. Gupta, M. D. Chopda, and R. B. Pachori, "Cross-subject emotion recognition using flexible analytic wavelet transform from eeg signals," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2266–2274, Mar. 2019.
[44] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from eeg data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.
[45] A. W. Moore, "Cross-validation for detecting and preventing overfitting," *School Comput. Sci. Carneigie Mellon Univ.*, 2001. http://www.autonlab.org/tutorials/overfit10.pdf
[46] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
[47] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, 1999.
[48] D. Zhang, L. Yao, K. Chen, and J. Monaghan, "A convolutional recurrent attention model for subject-independent eeg signal analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 5, pp. 715–719, May 2019.