Check for updates

Exploring EEG Features in Cross-Subject Emotion Recognition

Xiang Li[1], Dawei Song[2,3]*, Peng Zhang[1]*, Yazhou Zhang[1], Yuexian Hou[1] and Bin Hu[4]*

[1] Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin, China, [2] School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China, [3] School of Computing and Communications, The Open University, Milton Keynes, United Kingdom, [4] School of Information Science and Engineering, Lanzhou University, Lanzhou, China

Recognizing cross-subject emotions based on brain imaging data, e.g., EEG, has always been difficult due to the poor generalizability of features across subjects. Thus, systematically exploring the ability of different EEG features to identify emotional information across subjects is crucial. Prior related work has explored this question based only on one or two kinds of features, and different findings and conclusions have been presented. In this work, we aim at a more comprehensive investigation on this question with a wider range of feature types, including 18 kinds of linear and non-linear EEG features. The effectiveness of these features was examined on two publicly accessible datasets, namely, the dataset for emotion analysis using physiological signals (DEAP) and the SJTU emotion EEG dataset (SEED). We adopted the support vector machine (SVM) approach and the "leave-one-subject-out" verification strategy to evaluate recognition performance. Using automatic feature selection methods, the highest mean recognition accuracy of 59.06% (AUC = 0.605) on the DEAP dataset and of 83.33% (AUC = 0.904) on the SEED dataset were reached. Furthermore, using manually operated feature selection on the SEED dataset, we explored the importance of different EEG features in cross-subject emotion recognition from multiple perspectives, including different channels, brain regions, rhythms, and feature types. For example, we found that the Hjorth parameter of mobility in the beta rhythm achieved the best mean recognition accuracy compared to the other features. Through a pilot correlation analysis, we further examined the highly correlated features, for a better understanding of the implications hidden in those features that allow for differentiating cross-subject emotions. Various remarkable observations have been made. The results of this paper validate the possibility of exploring robust EEG features in cross-subject emotion recognition.

Keywords: EEG, emotion recognition, feature engineering, DEAP dataset, SEED dataset

1. INTRODUCTION

Emotion recognition as an emerging research direction has attracted increasing attention from different fields and is promising for many application domains. For example, in human-computer interaction (HCI), recognized user emotion can be utilized as a kind of feedback to provide better content to enhance the user experiences in e-learning, computer games, and information retrieval (Mao and Li, 2009; Chanel et al., 2011; Moshfeghi, 2012). Moreover, psychologists have verified the important roles that emotion plays in human health. Difficulties in the regulation of negative

emotions may cause various mood disorders, such as stress and depression (Gross and Muñoz, 1995), which may influence people's health (O'Leary, 1990). Hence, emotion recognition techniques also contribute to developing e-services for mental health monitoring. In particular, cross-subject emotion recognition (i.e., depression prediction based on a person's physiological data, with a classifier learnt from the training data from a group of patients who have been diagnosed as depression or not) has been considered an important task for its generality and wider applicability, compared with the intra-subject emotion recognition.

Electroencephalogram (EEG) measurements reflect the neural oscillations of the central nervous system (CNS) and are directly related to various higher-level cognitive processes (Ward, 2003), including emotion (Coan and Allen, 2004). EEG-based emotion recognition has shown a greater potential compared with the facial expression- and speech-based approaches, as the internal neural fluctuations cannot be deliberately concealed or controlled. However, a main issue confronted in this research area is how to improve the cross-subject recognition performance. The performances of current recognition systems are largely limited by the poor generalizability of the EEG features in reflecting emotional information across subjects. For example, Kim (2007) studied the bimodal data fusion method and utilized LDA to classify emotions. Using this method, the best obtained recognition accuracy on all three subjects' data was 55%, which was far lower than the best result of 92% obtained on a single subject's data. Zhu et al. (2015) adopted differential entropy as the emotion-related feature and the linear SVM as the classifier. The authors verified the recognition performance on intra-subject and cross-subject experimental settings respectively. The average recognition accuracy was 64.82% for cross-subject recognition tasks, which was also far lower than the results of 90.97% obtained in the intra-subject settings.

In the literature, there has been some related work that attempted to tackle this problem and to identify robust EEG features in cross-subject emotion recognition. For example, Li and Lu (2009) examined the recognition performance using ERD/ERS features extracted from different frequency bands and found that 43.5–94.5 Hz, the higher gamma band, was the optimal frequency band related to happiness and sadness. Lin et al. (2010) extracted DASM features and summarized the top 30 subject-independent features by measuring the ratio of between- and within-class variance, and found that the frontal and parietal electrode pairs were the most informative on emotional states. However, no significant difference between different frequency bands was observed in this work. Soleymani et al. (2012) performed cross-subject emotion recognition tasks on EEG and eye gaze data. The power spectral density (PSD) for EEGs was extracted. The most discriminative features for arousal were in the alpha band of the occipital electrodes, while those for valence were in the beta and gamma bands of the temporal electrodes. Kortelainen and Seppänen (2013) extracted the PSD from different frequency bands, and the best cross-subject classification rate for valence and arousal was obtained on the feature subset in the 1–32 Hz band. Zheng et al. (2016) focused on finding stable neural EEG patterns across subjects and

sessions for emotion recognition. The authors found that EEGs in lateral temporal areas were activated more for positive emotions than negative emotions in the beta and gamma bands and that subject-independent EEG features stemmed mostly from those brain areas and frequency bands.

In the aforementioned existing work, however, only few kinds of features were examined and why those robust features contribute to cross-subject emotion recognition was not studied. Hence, in this work, we aim at a more comprehensive and systematic exploration of a wider range of EEG features. Specifically, we extracted nine kinds of time-frequency domain features and nine kinds of dynamical system features from EEG measurements. Through automatic feature selection, e.g., recursive feature elimination (RFE), we verified the effectiveness and performance upper bounds of those features in cross-subject emotion recognition. Furthermore, through manual selection of features from different aspects, e.g., different EEG channels, we studied the importance of different aspects in cross-subject emotion recognition. We further conducted a correlation analysis to better understand the implications of those features for differentiating cross-subject emotions. The support vector machine (SVM), a state of the art classifier, was used in all the experiments.

2. MATERIALS AND METHODS

The procedure of the proposed methodology is illustrated in **Figure 1**. We adopted a "leave-one-subject-out" verification strategy. Each time we left one subject's data out as the test set and adopted the other subjects' data as the training set. The feature selection was conducted on the training set, and then, the performance was evaluated on the test set. This procedure was iterated until each subject's data had been tested. This strategy can eliminate the risk of "overfitting."

2.1. Experimental Data

We conducted our analysis using two publicly accessible datasets, namely, DEAP (dataset for emotion analysis using physiological signals) (Koelstra et al., 2012) and SEED (SJTU emotion EEG dataset) (Zheng et al., 2016). DEAP includes 32-channel EEG data collected from 32 subjects (17 male, 27.2 ± 4.4 years). The subjects' emotions were induced through one-minute-long music video clips. After each stimulus, the subjects rated their emotional experience on a two-dimensional emotional scale proposed by Russell (1980). The two dimensions are arousal (ranging from relaxed to aroused) and valence (ranging from unpleasant to pleasant). The higher a specific rating is, the more intense the emotion is in a specific dimension. The SEED dataset contains 62-channel EEG data collected from 15 subjects (7 male, 23.27 ± 2.37 years), and each subject participated in the experiment three times. The subjects' emotions are induced through 15 film clips, and each film clip lasts for approximately 4 min. Three classes of emotions (positive, neutral, negative) are evaluated, and each class has five corresponding film clips. In this study, we utilized only the trials of positive and negative emotions to evaluate the features' ability to differentiate between these two emotions. For
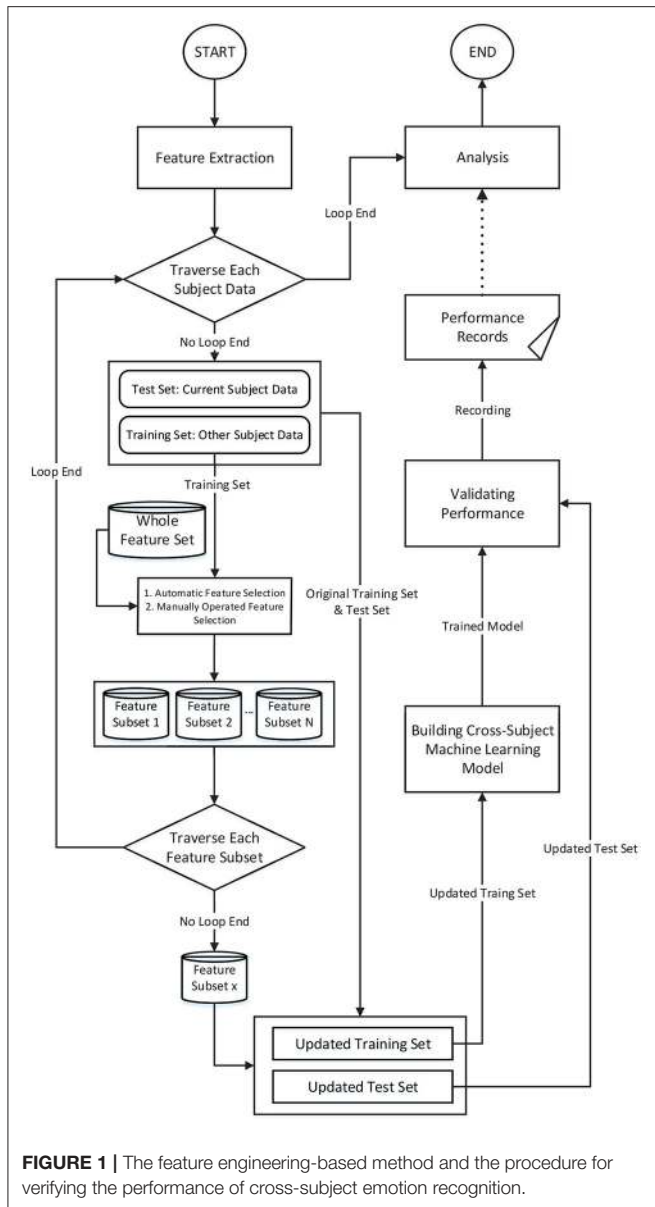
FIGURE 1 | The feature engineering-based method and the procedure for verifying the performance of cross-subject emotion recognition.

FIGURE 2 | (A) The data normalization method for one subject's multi-channel signals. **(B)** The sliding window-based feature extraction method for one EEG signal (taking one 12-s signal as an example). The mean of the calculated values in all sliding windows was adopted as the feature.

beta rhythm (16–31 Hz), and gamma rhythm (>32 Hz). We attempted to investigate the importance of these different rhythms in reflecting subjects' emotions. We excluded the delta rhythm (<4 Hz), as this rhythm is traditionally regarded as being correlated only with sleep. The four target rhythms were extracted through a custom finite impulse response (FIR) bandpass filter with a Hanning window. Secondly, we conducted data normalization as shown in **Figure 2A**. The extracted rhythm data for each subject were normalized channel by channel across all the trials. This procedure helped to remove subject bias and to generate more comparable features between subjects while allowing the variability of different channels to be preserved.

2.2.2. Label Preprocessing
For DEAP, we divided the subject trials into two classes according to their corresponding ratings on the valence dimension. A rating higher than 5 indicated a positive class, whereas a rating lower than 5 indicated a negative class. Hence, for valence, the two classes were high valence (positive) and low valence (negative). For SEED, the trials have already been categorized into three emotional classes (positive, neutral, negative); hence, we do not need to perform label preprocessing. For consistency, we studied only the positive and negative samples in SEED. The emotion recognition capability was evaluated using binary classification tasks.

2.3. Feature Extraction
In this work, we explored the robustness of a wider range of EEG features in cross-subject emotion recognition. Specifically, we extracted nine kinds of time-frequency domain features and nine kinds of dynamical system features from EEG measurements, as listed in **Table 1**. Extracting features based on some domain knowledge can provide a concise representation of the original data and materials. In this work, after preprocessing the data, we calculated the features for each of the four rhythms with a 4-s sliding window and a 2-s overlap, and then, the mean of the

consistency with the DEAP dataset, we used the one-minute-long data extracted from the middle part of each trial in SEED.

2.2. Data Preprocessing
2.2.1. EEG Preprocessing
As a kind of neurophysiological signal, EEG data are high dimensional and contain redundant and noisy information. In this work, after data acquisition, the raw data was firstly pre-processed, such as by removing the electrooculogram (EOG) and electromyogram (EMG) artifacts and downsampling the raw data to reduce the computational overhead in feature extraction. Two additional preprocessing procedures were needed before feature extraction, namely, rhythm extraction and data normalization. The multi-channel EEG is typically regarded as a reflection of brain rhythms. We first filtered out the four target rhythms, namely, the theta rhythm (4–7 Hz), alpha rhythm (8–15 Hz),
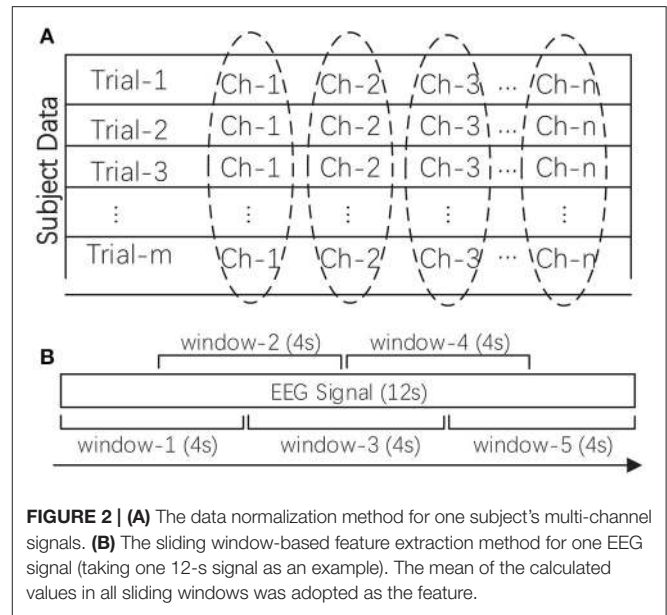
TABLE 1 | This table lists the two main categories of EEG features that we extracted.

Feature type	Extracted features
Time-frequency domain features	1. Peak-Peak Mean. 2. Mean Square Value. 3. Variance. 4. Hjorth Parameter: Activity. 5. Hjorth Parameter: Mobility. 6. Hjorth Parameter: Complexity. 7. Maximum Power Spectral Frequency. 8. Maximum Power Spectral Density. 9. Power Sum.
Non-linear dynamical system features	10. Approximate Entropy. 11. C0 Complexity. 12. Correlation Dimension. 13. Kolmogorov Entropy. 14. Lyapunov Exponent. 15. Permutation Entropy. 16. Singular Entropy. 17. Shannon Entropy. 18. Spectral Entropy.

The features were extracted for four rhythms. For the DEAP dataset, the total number of features extracted for one trial is 2304. For the SEED dataset, the total number of features extracted for one trial is 4464.

feature values extracted from those sliding windows was adopted as the trial's feature. The sliding window-based feature extraction methods are illustrated in **Figure 2B**. For DEAP, the number of features extracted for one trial is: $((9 + 9) \times 32) \times 4 = 2304$. For SEED, the number of features extracted for one trial is: $((9 + 9) \times 62) \times 4 = 4464$. All features were normalized before further analysis.

The details and reasons for selecting these candidate features are elaborated below:

2.3.1. Time-Frequency Domain Features

Nine kinds of features in the time and frequency domains of each signal were considered. The *peak-to-peak mean* is the arithmetic mean of the vertical length from the very top to the very bottom of the time series. The *mean squared value* is the arithmetic mean of the squares of the time series. *Variance* measures the degree of dispersion of the time series. After transforming the time series into the frequency domain through Fourier transform, we calculated the sum of the power spectral, and we further extracted the maximum power spectral density along with its corresponding frequency value. Three *Hjorth parameters* that can reflect characteristics of activity, mobility, and complexity were also extracted according to the work by Hjorth (1970): the *activity* parameter reflects the information of the signal power, the *mobility* parameter is an estimation of the mean frequency, and the *complexity* reflects the bandwidth and the change in frequency. The Hjorth parameters are considered suitable for analyzing non-stationary EEG signals.

2.3.2. Non-linear Dynamical System Features

We also extracted nine kinds of features that can reflect the characteristics of non-linear dynamical systems. Researchers have found that human brain manifests many characteristics specifically belonging to non-linear and chaotic dynamical systems; thus, the EEG signal is inherently complex, non-linear, non-stationary, and random in nature (Stam, 2005; Sanei and Chambers, 2013). *Approximate entropy (ApEn)* is a non-linear measure of the regularity of a signal; the more regular a signal is, the smaller the ApEn will be (Pincus et al., 1991). *C0 Complexity*

is adopted to measure the amount of the stochastic components, which assumes that a signal consists of a regular part and a stochastic part (Lu et al., 2008). *Correlation dimension* determines the number of dimensions (independent variables) that can describe the dynamics of the system and reflects the complexity of the process and the distribution of system states in the phase space (Khalili and Moradi, 2009). The *Lyapunov exponent* is used to measure the aperiodic dynamics of a chaotic system. This feature can capture the separation and evolution of the system's initial states in the phase space. The positive Lyapunov exponent indicates the chaos in the system (Übeyli, 2010). The *Kolmogorov entropy* is also a metric of the degree of chaos and measures the rate at which information is produced by the system as well as the rate at which information is lost by the system (Aftanas et al., 1997). Note that ApEn is closely related to Komolgorov entropy. The calculation of Komolgorov entropy is greatly influenced by the noise and dimensionality of the data. The complexity of neural activity can also be measured using the symbolic dynamic theory, in which a time series can be mapped to a symbolic sequence, from which the *permutation entropy (PE)* can be derived. The largest value of PE is 1, which indicates that the time series is completely random, while the smallest value of PE is 0, which indicates that the time series is completely regular (Li et al., 2007). *Singular spectrum entropy* is calculated by a singular value decomposition (SVD) of the trajectory, which is obtained by reconstructing the one-dimensional time series into a multi-dimensional phase space. This feature reflects the uncertainty and complexity of the energy distribution and is an indicator of event-related desynchronization (ERD) and event-related synchronization (ERS) (Zhang et al., 2009). *Shannon entropy* is a classical quantification of uncertainty and is frequently used to measure the degree of chaos in the EEG signal. *Power spectral entropy* is based on the Shannon entropy and measures the spectral complexity of the system. After the Fourier transform is performed, the signal is transformed into a power spectrum, and the information entropy of the power spectrum is called the power spectral entropy (Zhang et al., 2008).

2.4. Automatic Feature Selection

In this work, we first try to determine the upper bound of the performance of the proposed features. Hence, we choose to utilize some automatic feature selection techniques. Five different automatic feature selection methods were used to extract the most informative EEG features from the whole candidate set. Specifically, the whole features are re-ranked according to a pre-defined ranking criteria, e.g., based on the degree of correlation between the feature and the target class or based on the value of the feature weight, and then, the features above a pre-defined threshold are selected (Huang et al., 2006; Maldonado and Weber, 2008).

Two typical automatic feature selection techniques are the filter-based strategy and the wrapper-based strategy (Guyon and Elisseeff, 2003). The former is independent of any pattern recognition algorithm and filters out a specific number of features according to some statistical properties of the features. The classical filter-based strategy includes the chi-squared (χ^2), mutual information, and *F*-test methods. The wrapper-based strategy, on the other hand, cooperates with a specific pattern

recognition algorithm. A widely adopted wrapper-based strategy is the recursive feature elimination (RFE) method. We also considered a more efficient L1-norm penalty-based feature selection method, which has been widely used in recent years.

The details of these feature selection methods are elaborated as follows:

2.4.1. Chi-Squared-Based Feature Selection (χ^2)

The Chi-squared test is a classical statistical hypothesis test method for testing the independence of two variables or to investigate whether the distribution of one variable differs from that of another. This work is concerned with the former, formulated as below:

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}, \tag{1}$$

where r and c are the number of categories in the two random variables, $O_{i,j}$ is the number of observations of type i, j, and the $E_{i,j}$ is the expected frequency of type i, j. In our work, a higher χ^2 value indicates a higher correlation between a feature variable and the target classes.

2.4.2. Mutual Information-Based Feature Selection (MI)

The mutual information metric is derived from probability theory and information theory. This metric is adopted to measure the mutual dependence (shared information) between the feature variables and the target classes. It is closely linked to the concept of entropy, which defines how much information is contained in a variable. Mutual information can be expressed as follows:

$$
\begin{aligned}
I(X; Y) &= H(X) + H(Y) - H(X, Y) \\
&= \sum_x P(x) \log \frac{1}{P(x)} + \sum_y P(y) \log \frac{1}{P(y)} \\
&\quad - \sum_{x,y} P(x, y) \log \frac{1}{P(x, y)} \\
&= \sum_{x,y} P(x, y) \log \frac{p(x, y)}{P(x)p(y)},
\end{aligned}
\tag{2}
$$

where $H(X)$ and $H(Y)$ are the marginal entropy of X and Y respectively, and $H(X, Y)$ is the joint entropy of X and Y.

2.4.3. ANOVA F-Value-Based Feature Selection (AF)

F-test is a representative version of the analysis of variance (ANOVA). It is typically used to test whether the means of multiple populations are significantly different. In feature selection, ANOVA can measure the "F-ratio" of the between-class variance (as in Equation 4) over within-class variance (as in Equation 5). The "F-ratio" indicates the degree of class separation, as formulated in Equation (3). The higher a feature variable's F-ratio is, the better this feature is in differentiating different classes.

$$F_{ratio} = \frac{\sigma_{between}^2}{\sigma_{within}^2} \tag{3}$$

where the between-class variance and within-class variance are:

$$\sigma_{between}^2 = \frac{\sum_{j=1}^{J} (\bar{x}_j - \bar{x})^2 N_j}{J - 1} \tag{4}$$

and

$$\sigma_{within}^2 = \frac{(\sum_{j=1}^{J} \sum_{i=1}^{N_j} (x_{i,j} - \bar{x})^2) - (\sum_{j=1}^{J} (\bar{x}_j - \bar{x})^2 N_j)}{N - J}, \tag{5}$$

respectively. J is the number of classes, N_j is the number of measurements in the jth class, $\bar{x}_j$ is the mean of the jth class, $\bar{x}$ is the overall mean, and $x_{i,j}$ is the ith measurement of the jth class.

2.4.4. Recursive Feature Elimination (RFE)

As first introduced in Guyon et al. (2002) for gene selection, RFE is a wrapper-based method that judges the importance of features using an external machine learning algorithm. It adopts a sequential backward elimination strategy. First, the algorithm is trained on the initial whole set of features and assigns weight to each of the features. Then, a pre-defined number of features with the lowest-ranking absolute weights are pruned from the current feature set. This procedure recursively repeats for several steps until the desired number of selected features is reached. The pseudo code of the RFE is illustrated below in Algorithm 1. In this work, we used SVM with a linear kernel as the ranking method, in which the RFE utilized $\|w\|$ as the ranking criteria for the importance of the features.

Algorithm 1: Pseudo Code for Recursive Feature Elimination (RFE) Algorithm

---

Input:
Training set: T
Feature set: $F = \{f_1, f_2, ..., f_p\}$
Ranking method: $M(T, F)$
Desired feature number: q
Number of feature to eliminate in each step: k
Output:
Final ranking feature set: $R = \{f_{r1}, f_{r2}, ..., f_{rp}\}$
Final selected feature set: $F = \{f_1, f_2, ..., f_q\}$

1 **Initialization;**
2 Steps: S = (p-q)/k;
3 **for** $i = 1 \rightarrow S$ **do**
4   Rank set F according to $M(T, F)$;
5     $L_k \longleftarrow$ Last ranked k features in F;
6     $R[p - i*k + 1 : p - (i - 1)*k] \longleftarrow L_k$;
7     $F \longleftarrow F - L_k$;
8 **end**

2.4.5. L1-Norm Penalty-Based Feature Selection (L1)

This method introduces a L1-norm regularization term into the objective function to induce the sparsity by shrinking the weights toward zero. Regularization is usually adopted in case that the size of the training set is smaller relative to the dimensionality of the features. This process favors small parameters of the model to prevent overfitting (Ng, 2004). It is natural in feature selection settings for features with weights of zero to be eliminated from

the candidate set. Some researchers have indicated that the L1-norm-based method is better than the L2-norm-based method, especially when there are redundant noise features (Zhu et al., 2004). In this work, we adopted an SVM with an L1-norm penalty to select the important features. The formulation of the objective function is as follows:

$$\min_{\omega_0, \omega} \sum_{i=1}^{n} [1 - y_i(\omega_0 + \sum_{j=1}^{q} \omega_j x_{i,j})] + C\|\omega\|_1 \qquad (6)$$

$\|\omega\|_1$ is the L1-norm term, in which the ω represents the model weights. The parameter "C" controls the trade-off between the loss and penalty.

2.5. Manually Operated Feature Selection

The upper bound of the performance of the proposed features can be verified using the previous automatic feature selection methods. We chose to further evaluate the performance and the importance of the features from different aspects, including different electrodes, locations, rhythms, and feature types. Haufe et al. (2014) indicated that the interpretation of the parameters in backward methods (multivariate classifiers) may lead to the wrong conclusions in neuroimaging data modeling. Hence, in this work, we did not conduct analyses based on the selected features or the corresponding feature weights of the automatic feature selection methods. However, we adopted a simple "searchlight" approach in which we manually selected features from different aspects and evaluated the performances independently.

The performances of automatic and manually operated feature selection were verified by applying a linear SVM. The codes for the data preprocessing, feature extraction, and cross-subject verification processes with different feature selection methods, as well as the extracted features, can be accessed at the following web page: https://github.com/muzixiang/EEG_Emotion_Feature_Engineering.

3. RESULTS

3.1. Overall Evaluation

We first determined the upper bound of the performance of the proposed features using all of the mentioned automatic feature selection methods. In the experiment, considering the computational overhead of different methods as well as the adequacy of the experiments, we employed different settings for different methods. For filter- and RFE-based methods, we set the step size for the number of selected features to 10. Hence, the number of steps for DEAP and SEED are 230 and 446, respectively. For the L1-norm penalty-based method, we set 100 different values for penalty parameter "C" ranging from 0.01 to 1 with a step size of 0.01. We adopted a "leave-one-subject-out" verification strategy, and the performance was evaluated by the mean recognition accuracy metric. **Figures 3A–C, 4A–C** illustrate the performance of the automatic feature selection methods with different settings on the DEAP and SEED dataset, respectively.

For DEAP, when no feature selection method was utilized, the recognition performance was 0.5531 (std:0.0839). The best result of 0.5906 (std: 0.0868) was obtained with the L1-norm penalty-based method when the value of "C" is 0.08. For SEED, when no feature selection method was utilized, the recognition performance was 0.7844 (std:0.1119). The best result of 0.8333 (std: 0.1016) was obtained with the RFE-based method when the number of selected features is 130. **Table 2** shows the p-values calculated through one-way ANOVA test between the method with best performance and other methods. For a better comparison between those methods, as shown in **Figures 3D, 4D**, we also produced ROC curves. Different feature selection methods were compared by analyzing their ROC curves and the Areas under the ROC curves (AUC). The results showed that the L1-norm penalty-based method outperformed other methods on both DEAP (AUC = 0.605) and SEED (AUC = 0.904). Moreover, the L1-norm penalty-based method incurred a lower computational cost than the other methods. Hence, considering both effectiveness and efficiency, the L1-norm penalty-based feature selection method is recommended to verify the upper bound of the recognition performance when a large amount of features are provided.
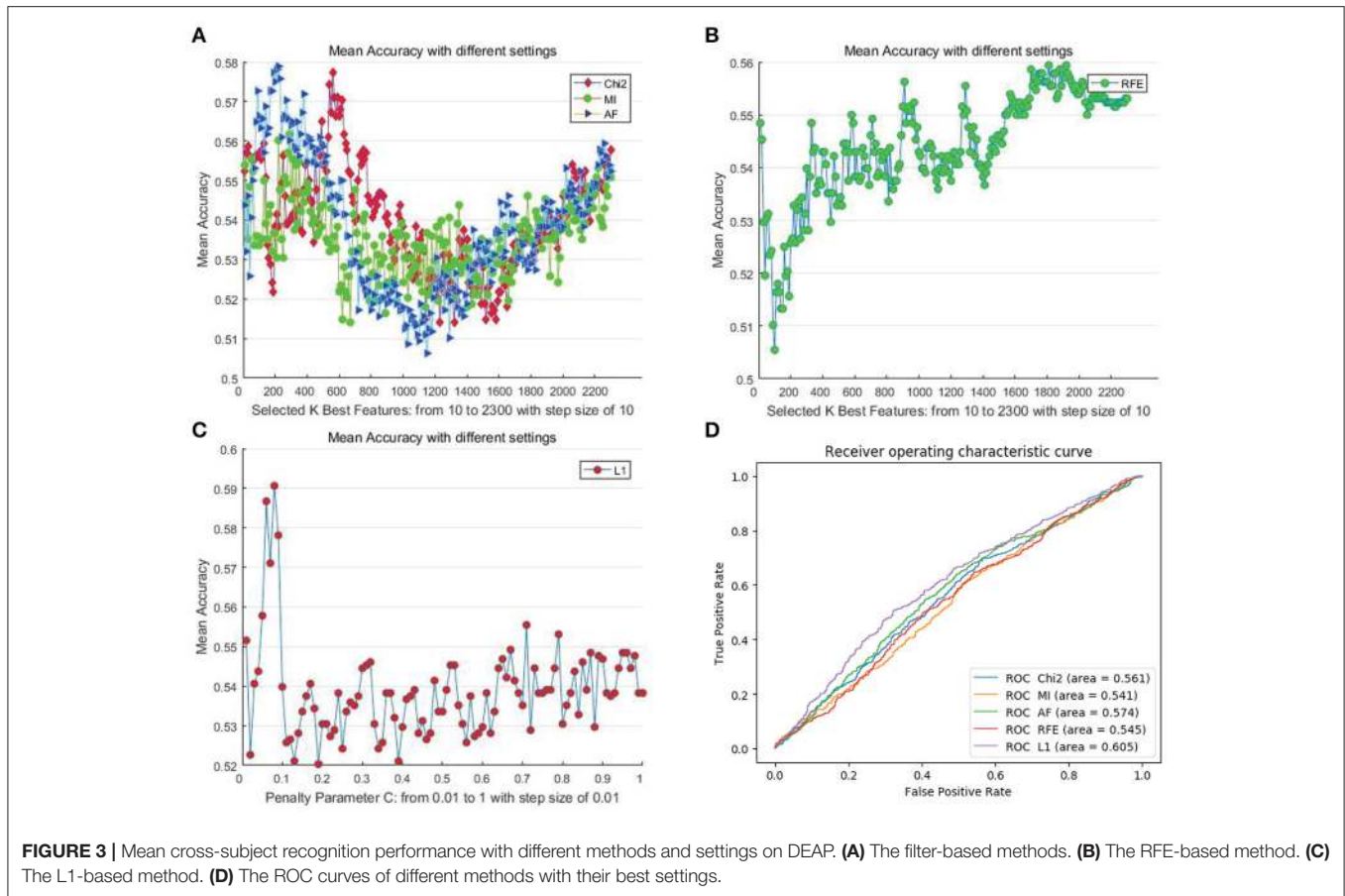
The results demonstrate the effectiveness of our proposed EEG features in cross-subject emotion recognition, especially on the SEED dataset. The performance on DEAP is significantly inferior to that on SEED. This is possibly due to the relatively low quality of the data and the emotional ratings of trials in the DEAP dataset. Hence, we chose to conduct further evaluation only on the SEED dataset.

3.2. Evaluation From Different Perspectives

We also explored the importance of different EEG features in cross-subject recognition from multiple perspectives, including different channels, brain regions, rhythms, and feature types.

Figure 5A illustrates the performance of each individual channel. We ranked the performance of these channels and labeled the top one-sixth of the channels on the diagram of the 10–20 international system of electrode placement. The channels on the bilateral temporal regions achieved higher mean accuracies for cross-subject emotion recognition. As shown in **Figure 5B**, we also evaluated the performance of different regions, including the left-right anterior regions, left-right posterior regions, left-right hemispheres, and anterior-posterior hemispheres. The partition of the regions is illustrated in the figure, and the channels in the cross regions were eliminated when evaluating the performance of the sub-regions. We found that the left anterior region achieved better performance compared to the right anterior region, left posterior region, and the right posterior region, especially when the information in the beta band was utilized. The left hemisphere performed better than the right hemisphere in each band except for the gamma band. Furthermore, the information from the anterior hemisphere enhanced recognition performance in each band more than that from the posterior hemisphere.

Validating the performance of different EEG rhythms was also of interest to us. As we can see in **Figure 5C**, the individual beta rhythm achieved the best performance, and the higher-frequency

FIGURE 3 | Mean cross-subject recognition performance with different methods and settings on DEAP. **(A)** The filter-based methods. **(B)** The RFE-based method. **(C)** The L1-based method. **(D)** The ROC curves of different methods with their best settings.

beta rhythm and gamma rhythm bands performed better than the lower-frequency theta rhythm and alpha rhythm bands. When the data on all four rhythms were concatenated, the performance was greatly promoted.

The main research objective of this paper is to verify the effectiveness of the proposed features. Thus, we also evaluated the performances of each kind of feature. As shown in **Figure 5D**, the information on linear features No. 5 (Hjorth parameter: mobility), No. 6 (Hjorth parameter: complexity), and No. 7 (maximum power spectral frequency) in the beta rhythm led to the best mean recognition accuracy. Only the non-linear features No. 12 (correlation dimension), No. 13 (Kolmogorov entropy), and No. 17 (Shannon entropy) can lead to a mean accuracy over 60%. **Figure 5E** presents the performance comparison between the linear and non-linear features in different frequency bands. The results show that using linear features outperformed the use of non-linear features in each frequency band when linear SVM and random forest (RF) classifiers were applied. Hence, considering the high computational overhead of extracting the non-linear features, solely adopting linear features seems an effective choice for constructing a real-time emotion recognition system. Nevertheless, we should also clarify that the values of the non-linear features calculated in this work may be not optimal. The performance of the non-linear features are influenced by many factors, e.g., the parameter settings and the data volume

limitations for the search space. The optimal values of those non-linear features are worth further exploration.

3.3. Correlation Analysis

As we can see in **Figure 5D**, the performances of some feature type (e.g., features No. 2, No. 3, No. 4, and No. 9) are seemingly identical. This result likely indicates that some features can be highly correlated in identifying a certain emotion class. Hence, for examining those highly correlated features, we calculated the Pearson correlation coefficients for those 18 different feature types. For example, as presented in **Figure 6**, the linear features No. 2, No. 3, No. 4, and No. 9 are absolutely positively correlated in each rhythm, which explains why the performances of these features are approximately identical. Linear features No. 1 and No. 8 are highly positively correlated with all other linear features except for the Hjorth parameters. For the Hjorth parameters, feature No. 5 is highly positively correlated with feature No. 7 in each rhythm, and is highly negatively correlated with feature No. 6 in the beta rhythm. For the non-linear features, we can see that feature No. 12 is highly and positively correlated with feature No. 13 and No. 17 in the higher-frequency bands, and that feature No. 16 is highly and negatively correlated with feature No. 18 in each rhythm.

Moreover, through analyzing the correlation of the channels based on those features, we attempted to investigate the
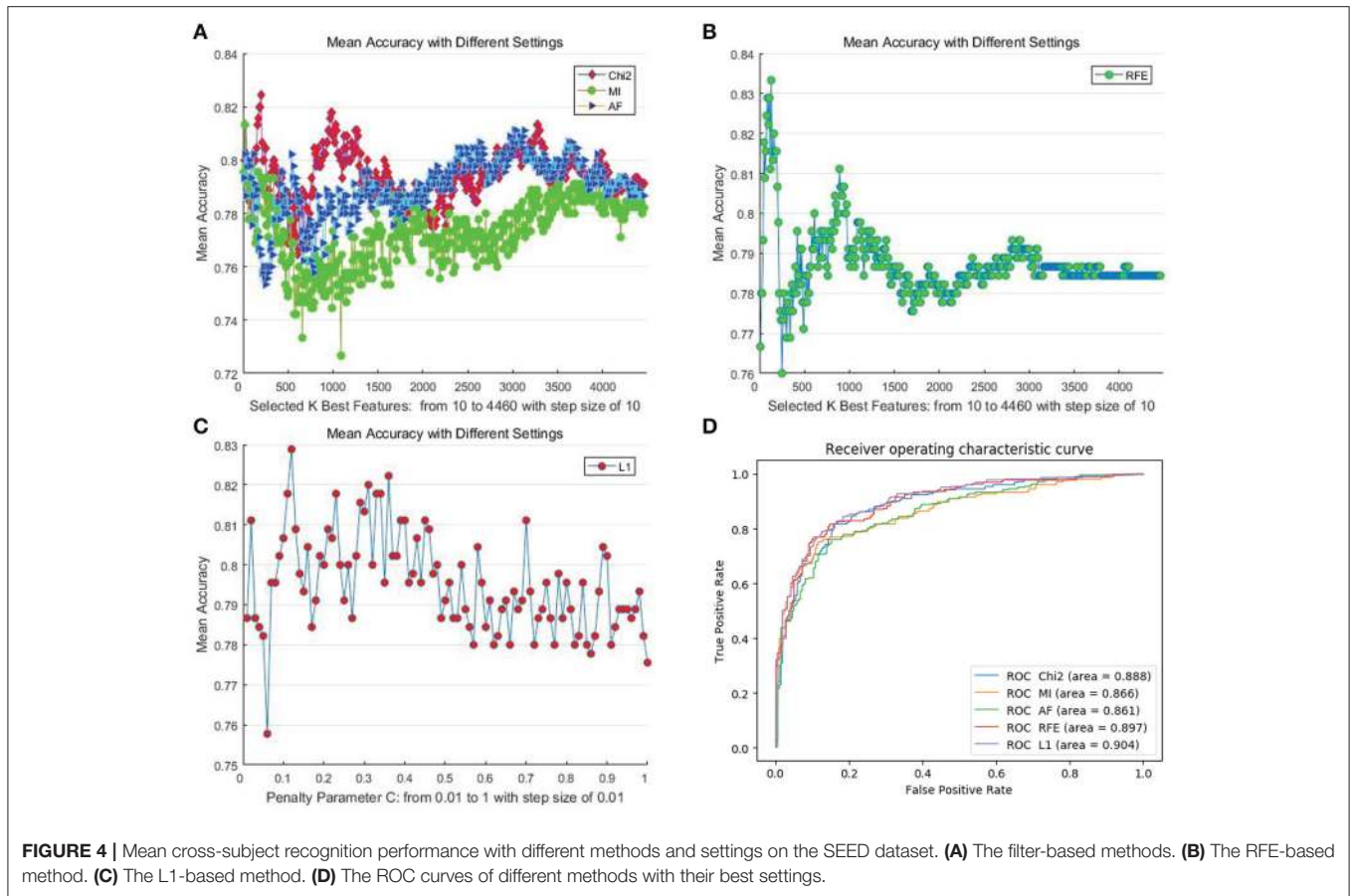
FIGURE 4 | Mean cross-subject recognition performance with different methods and settings on the SEED dataset. **(A)** The filter-based methods. **(B)** The RFE-based method. **(C)** The L1-based method. **(D)** The ROC curves of different methods with their best settings.

TABLE 2 | The performance upper bound of the proposed features using different automatic feature selection methods.

DEAP	χ^2	MI	AF	RFE
	Step No.: 56	Step No.: 29	Step No.: 22	Step No.: 181
	Mean: 0.5773	Mean: 0.5617	Mean: 0.5789	Mean: 0.5594
	St.Dev.: 0.0841	St.Dev.: 0.0914	St.Dev.: 0.1004	St.Dev.: 0.0818
L1				
Step No.: 8	$p = 0.5364$	$p = 0.1992$	$p = 0.6192$	$p = 0.1432$
Mean: **0.5906**				
St.Dev.: 0.0868				
SEED	χ^2	MI	AF	L1
	Step No.: 20	Step No.: 2	Step No.: 2	Step No.: 13
	Mean: 0.8244	Mean: 0.8133	Mean: 0.8111	Mean: 0.8289
	St.Dev.: 0.1151	St.Dev.: 0.1227	St.Dev.: 0.1389	St.Dev.: 0.0899
RFE				
Step No.: 12	$p = 0.8242$	$p = 0.6305$	$p = 0.4895$	$p = 0.8999$
Mean: **0.8333**				
St.Dev.: 0.1016				

Meanwhile, the p-values calculated through one-way ANOVA between the method with best performance and other methods are also exhibited. The highest mean recognition accuracy is shown in bold type.

underlying mechanisms of those features that allow for differentiating cross-subject emotions. Specifically, for each subject and for each specific feature, we constructed correlation matrices of the 62 channels for subjects' negative trials and positive trials. After all of the correlation matrices had been constructed, we averaged the correlation matrices in the negative group and positive group. The mean correlation matrices for specific features are presented in **Figure 7**. We also conducted statistical analyses to compare the differences in channel correlations between the negative group and the positive group. The t test results are illustrated in **Figure 8**. The results in both **Figures 7**, **8** indicate that for almost every feature, the mean correlations in the negative group are higher than those in the positive group.

The connection network of the 62 channels is represented in the form of a binary matrix, which was constructed based on the obtained correlation matrices. We first needed to determine the threshold of the correlation coefficients, based on which the connection between two channels could be established. To be more specific, the value in the binary matrix was set to 1 when the corresponding value in the correlation matrix was greater than the threshold. Otherwise, the value in the binary matrix was set to 0. The value of 1 in the binary matrix indicates that there is a connection between the two corresponding nodes. Based on the obtained binary matrix, the connection network of the channel nodes was constructed.

For measuring the coherence of different channel locations in different emotional states, we calculated the clustering coefficients of each node in the connection network. The clustering coefficient was a reflection of the degree of aggregation
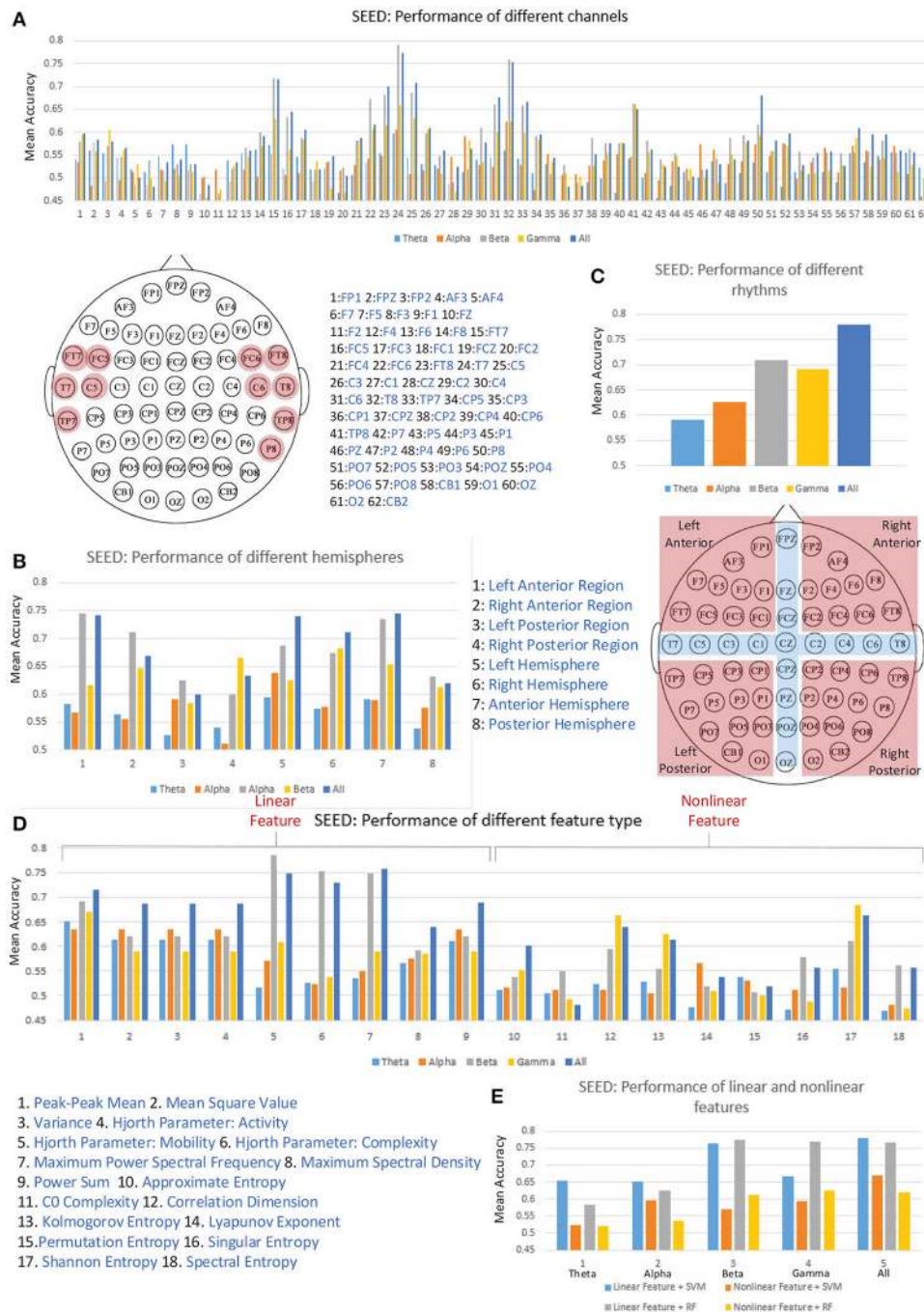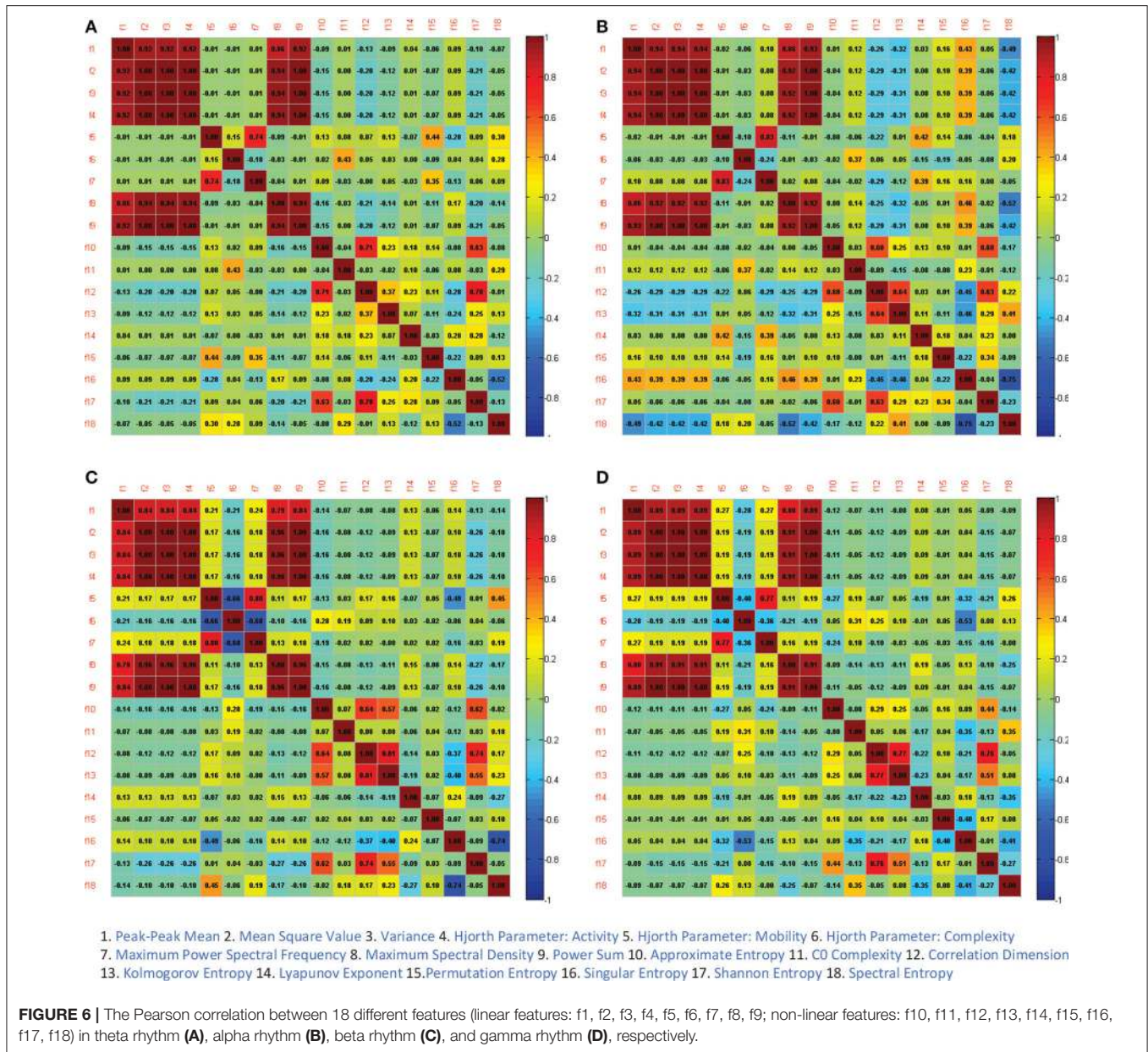
FIGURE 5 | The cross-subject recognition performance based on features from different channels **(A)**, different regions **(B)**, different rhythms **(C)**, different features **(D)**, and different feature types **(E)**.

of different channel locations. The thresholds at which the global clustering coefficients were significantly different between the positive group and negative group are illustrated in **Table 3**. As mentioned above in **Figure 5D**, feature No. 5 (Hjorth parameter: mobility) in beta rhythm achieved the best recognition performance. Thus, in this paper, we use this feature as an example to illustrate the topographic plot of the clustering

coefficients of the groups with negative and positive emotions. As shown in **Figure 9A**, at each threshold, the clustering coefficients in the left anterior regions of the negative groups are consistently higher than those of the positive groups. This dynamic may account for the results obtained in **Figure 5B** indicating that the left anterior regions yields the best recognition performance when beta rhythm information is utilized. Nevertheless, different

FIGURE 6 | The Pearson correlation between 18 different features (linear features: f1, f2, f3, f4, f5, f6, f7, f8, f9; non-linear features: f10, f11, f12, f13, f14, f15, f16, f17, f18) in theta rhythm **(A)**, alpha rhythm **(B)**, beta rhythm **(C)**, and gamma rhythm **(D)**, respectively.

1. Peak-Peak Mean 2. Mean Square Value 3. Variance 4. Hjorth Parameter: Activity 5. Hjorth Parameter: Mobility 6. Hjorth Parameter: Complexity
7. Maximum Power Spectral Frequency 8. Maximum Spectral Density 9. Power Sum 10. Approximate Entropy 11. C0 Complexity 12. Correlation Dimension
13. Kolmogorov Entropy 14. Lyapunov Exponent 15.Permutation Entropy 16. Singular Entropy 17. Shannon Entropy 18. Spectral Entropy

features and thresholds could have different topographic plots in which the clustering coefficients may be quite different from those of feature No. 5 in the beta rhythm. For example, in addition to feature No. 5 in beta rhythm, feature No. 6 in beta rhythm also led to a high performance. We have also presented the clustering coefficients in **Figure 9B**. However, the topographic plot was different from that in **Figure 9A**, and the left anterior region was no longer significantly different between the two groups. Hence, the important locations for emotion recognition cannot be determined simply by analyzing only one or two features.

As described above, we should point out that such correlation analysis may not be adequate to fully interpret the mechanism of the features. Moreover, for example, as shown in **Figure 8**, the feature No. 1 in the gamma rhythm cannot significantly

differentiate the correlation coefficients of the two groups. However, as shown in **Figure 5D**, this feature can still lead to a better performance than most of the non-linear features.

4. DISCUSSIONS

In this work, we verified the effectiveness of 18 kinds of EEG features in cross-subject emotion recognition, including 9 kinds of time-frequency domain features and 9 kinds of dynamical system features. We adopted a "leave-one-subject-out" method to verify the performance of the proposed features. After automatic feature selection, the highest mean recognition accuracies of 59.06% (AUC = 0.605) on the DEAP dataset and 83.33% (AUC = 0.904) on the SEED dataset were reached.
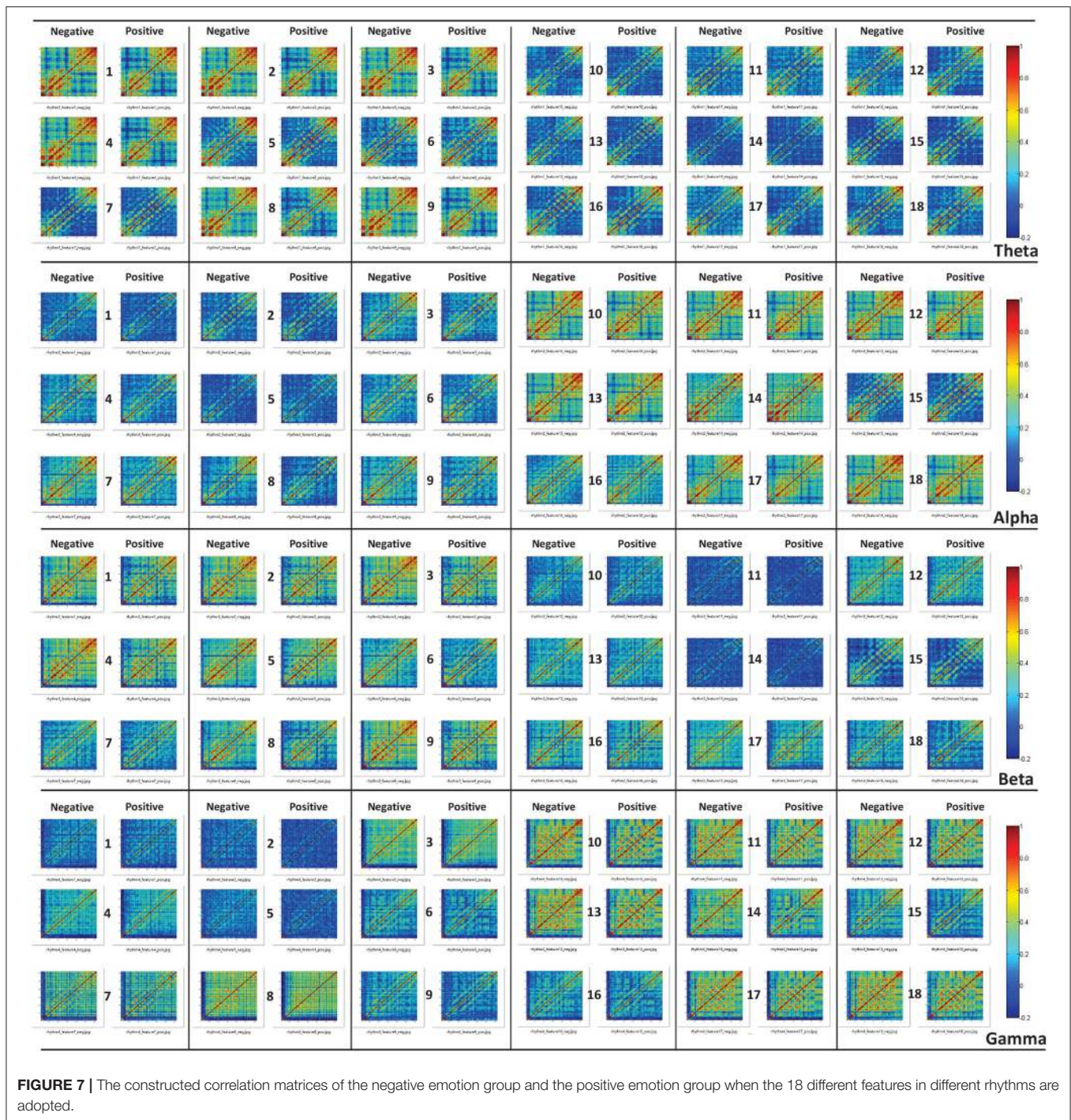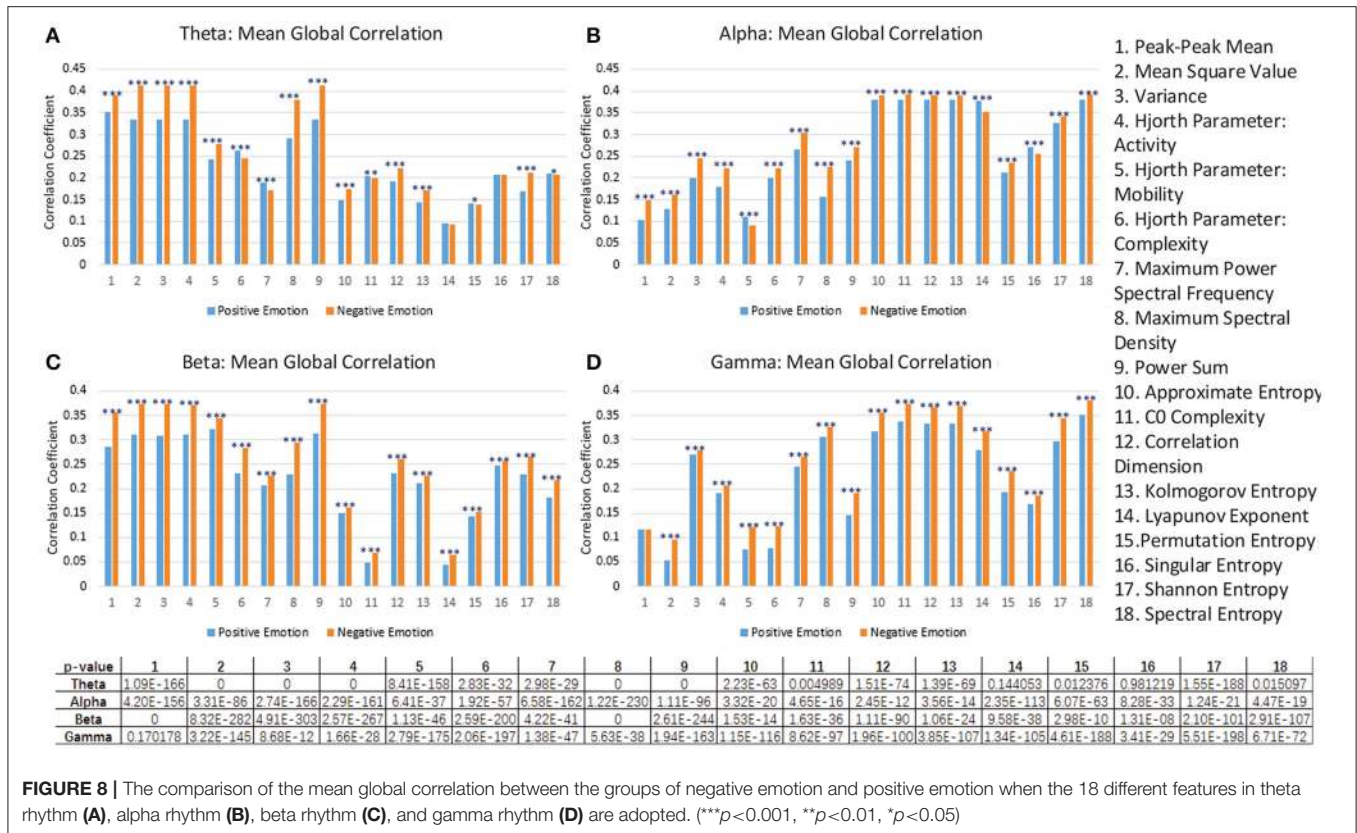
FIGURE 7 | The constructed correlation matrices of the negative emotion group and the positive emotion group when the 18 different features in different rhythms are adopted.

The performance on DEAP was not as good as that on SEED, which could be due to the low quality of the data in the emotional ratings of the trials. The noise in the emotional ratings degraded the ability of the model to differentiate between different classes. Through drawing the ROC curves, we found that the L1-norm penalty-based feature selection method exhibited robust performance on both two datasets. Considering its lower computational overhead, this method is the

best strategy to adopt when analyzing large numbers of candidate features.

We also evaluated the cross-subject recognition performance from different perspectives, including different EEG channels, different regions, different rhythms, different features, and different feature types. We chose to conduct analyses on the SEED dataset because of its better performance. Specifically, through evaluation over individual channels, we found that

p-value	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Theta	1.09E-166	0	0	0	8.41E-158	2.83E-32	2.98E-29	0	0	2.23E-63	0.004989	1.51E-74	1.39E-69	0.144053	0.012376	0.981219	1.55E-188	0.015097
Alpha	4.20E-156	3.31E-86	2.74E-166	2.29E-161	6.41E-37	1.92E-57	6.58E-162	1.22E-230	1.11E-96	3.32E-20	4.65E-16	2.45E-12	3.56E-14	2.35E-113	6.07E-63	8.28E-33	1.24E-21	4.47E-19
Beta	0	8.32E-282	4.91E-303	2.57E-267	1.13E-46	2.59E-200	4.22E-41	0	2.61E-244	1.53E-14	1.63E-36	1.11E-90	1.06E-24	9.58E-38	2.98E-10	1.31E-08	2.10E-101	2.91E-107
Gamma	0.170178	3.22E-145	8.68E-12	1.66E-28	2.79E-175	2.06E-197	1.38E-47	5.63E-38	1.94E-163	1.15E-116	8.62E-97	1.96E-100	3.85E-107	1.34E-105	4.61E-188	3.41E-29	5.51E-198	6.71E-72

FIGURE 8 | The comparison of the mean global correlation between the groups of negative emotion and positive emotion when the 18 different features in theta rhythm **(A)**, alpha rhythm **(B)**, beta rhythm **(C)**, and gamma rhythm **(D)** are adopted. (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$)

the channels with the best performances were mainly located in the bilateral temporal regions, which was consistent with the finding in Soleymani et al. (2012), Zheng et al. (2016). We partitioned the channels into different groups according to the different regions and evaluated the performances of the different groups. We found that the left anterior region achieved a better performance compared to the other sub-regions, especially when the information in the beta band was utilized. The left hemisphere performed better than the right hemisphere except for in the gamma band. Furthermore, the anterior hemisphere exhibited an improved recognition performance compared to the posterior hemisphere, especially when data from all rhythms were utilized. The relationship between emotion recognition and frontal regions was illustrated in the studies of Schmidt and Trainor (2001), Lin et al. (2010). Schmidt and Trainor (2001) found the relatively higher left frontal EEG activity under exposure to happy musical excerpts and relatively higher right frontal EEG activity under exposure to sad musical excerpts, and the overall frontal EEG activity could distinguish the intensity of the emotions. Lin et al. (2010) found that the frontal and parietal electrode pairs were the most informative on emotional states.

The evaluation of different rhythms indicated that the information in higher-frequency bands contributed more to cross-subject emotion recognition compared to lower-frequency bands. The effectiveness of the beta and gamma rhythms in promoting emotion recognition was also presented in

Lin et al. (2010), Soleymani et al. (2012), Zheng et al. (2016). Moreover, some neuroscience studies have found that emotion-related neural information mainly resides in higher-frequency bands (Müller et al., 1999; Kortelainen et al., 2015). By evaluating the performances of individual features, we found that linear features No. 5 (Hjorth parameter: mobility), No. 6 (Hjorth parameter: complexity), and No. 7 (maximum power spectral frequency) in the beta rhythm led to the best mean recognition accuracy. However, the Hjorth parameters have not been widely adopted in EEG-based emotion recognition. We also found that the combination of the linear features greatly outperformed the combination of non-linear features in each frequency band. Considering the high computational overhead in extracting the non-linear features, adopting linear features in designing real-time emotion recognition systems is recommended. Nevertheless, the non-linear features calculated here may be not the optimal ones, given that emotional information is very likely processed in a non-linear way. The optimal values of the non-linear features for reflecting emotional processes are certainly worth further exploration.

For a better understanding of the mechanisms of those features that allow for differentiating between emotions, we further conducted a correlation analysis for 62 channels for each feature. We calculated and constructed correlation matrices using different features. We found that for nearly every feature, the negative group has a higher mean correlation coefficient
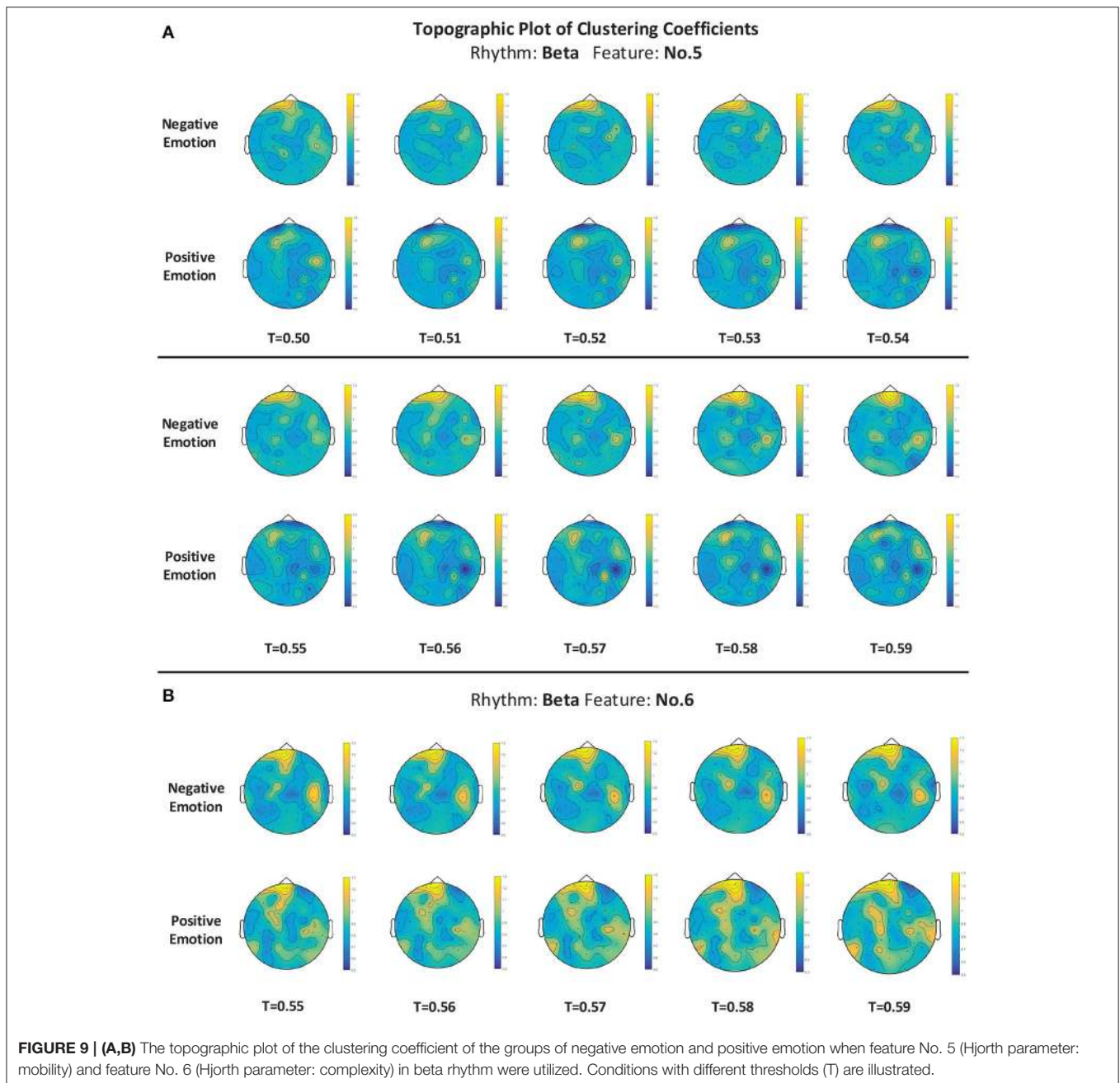
TABLE 3 | The threshold scope that can significantly differentiate the clustering coefficients in groups of positive emotion and negative emotion ($p < 0.05$).

Rhythm	Feature	Threshold scope	Feature	Threshold scope
Theta	No.1	0.34~0.65, 0.92~0.99	No.10	0.01~0.36, 0.62~0.63, 0.69~0.71, 0.89~0.92
Theta	No.2	0.01~0.66, 0.92~0.99	No.11	0.14~0.47, 0.71~0.74, 0.84~0.87, 0.97~0.98
Theta	No.3	0.01~0.66, 0.92~0.99	No.12	0.01~0.35, 0.59~0.62, 0.77~0.78, 0.94~0.96
Theta	No.4	0.01~0.66, 0.92~0.99	No.13	0.01~0.20, 0.49~0.50
Theta	No.5	0.01~0.39	No.14	0.21~0.31
Theta	No.6	0.01~0.24	No.15	0.07~0.32, 0.40~0.49, 0.70~0.72, 0.95~0.96
Theta	No.7	0.01~0.15	No.16	0.01~0.34, 0.54~0.58
Theta	No.8	0.01~0.65, 0.93~0.99	No.17	0.01~0.42
Theta	No.9	0.01~0.66, 0.93~0.99	No.18	0.49~0.50, 0.90~0.98
Alpha	No.1	0.01~0.58, 0.66~0.95	No.10	0.01~0.08, 0.53~0.56, 0.67~0.72, 0.96~0.99
Alpha	No.2	0.01~0.27, 0.49~0.60, 0.75~0.91	No.11	0.01~0.32, 0.64~0.76, 0.95~0.99
Alpha	No.3	0.01~0.41, 0.56~0.66, 0.83~0.95	No.12	0.01~0.08, 0.23~0.32, 0.67~0.76, 0.95~0.99
Alpha	No.4	0.01~0.40, 0.47~0.59, 0.90~0.93	No.13	0.01~0.05, 0.28~0.32, 0.67~0.78, 0.95~0.99
Alpha	No.5	0.01~0.29, 0.50~0.52	No.14	0.01~0.25
Alpha	No.6	0.01~0.25, 0.48~0.73	No.15	0.01~0.17, 0.24~0.28, 0.38~0.40, 0.89~0.92
Alpha	No.7	0.01~0.35, 0.46~0.48, 0.95~0.97	No.16	0.01~0.14, 0.36~0.39
Alpha	No.8	0.01~0.36, 0.49~0.70, 0.79~0.99	No.17	0.01~0.05, 0.64~0.66, 0.90~0.99
Alpha	No.9	0.01~0.35, 0.82~0.85, 0.95~0.98	No.18	0.01~0.06, 0.65~0.78, 0.95~0.99
Beta	No.1	0.01~0.44, 0.53~0.58, 0.90~0.92	No.10	0.01~0.13, 0.68~0.69
Beta	No.2	0.01~0.47	No.11	0.01~0.32, 0.42~0.43, 0.46~0.54, 0.67~0.80
Beta	No.3	0.01~0.49, 0.85~0.86	No.12	0.01~0.26, 0.91~0.98
Beta	No.4	0.01~0.49, 0.53~0.56, 0.85~0.86	No.13	0.01~0.34, 0.48~0.54, 0.87~0.92
Beta	No.5	0.01~0.16, 0.20~0.23, 0.48~0.61	No.14	0.01~0.26, 0.38~0.43, 0.62~0.74, 0.81~0.87
Beta	No.6	0.01~0.41, 0.55~0.64, 0.85~0.91	No.15	0.01~0.21
Beta	No.7	0.01~0.31, 0.67~0.71, 0.91~0.99	No.16	0.01~0.21, 0.57~0.62, 0.81~0.83, 0.86~0.88
Beta	No.8	0.01~0.40, 0.86~0.89, 0.96~0.99	No.17	0.01~0.12, 0.67~0.69, 0.86~0.88, 0.95~0.99
Beta	No.9	0.01~0.47	No.18	0.01~0.29, 0.39~0.46, 0.74~0.88
Gamma	No.1	0.14~0.27, 0.66~0.70, 0.76~0.81	No.10	0.15~0.33, 0.40~0.64, 0.83~0.88
Gamma	No.2	0.01~0.33, 0.85~0.98	No.11	0.17~0.22, 0.28~0.29
Gamma	No.3	0.01~0.17, 0.50~0.51	No.12	0.26~0.49
Gamma	No.4	0.01~0.47, 0.97~0.98	No.13	0.23~0.54
Gamma	No.5	0.01~0.31, 0.73~0.98	No.14	0.01~0.35, 0.97~0.99
Gamma	No.6	0.01~0.30, 0.80~0.99	No.15	0.01~0.22, 0.42~0.47, 0.72~0.74, 0.83~0.88
Gamma	No.7	0.01~0.35, 0.93~0.99	No.16	0.01~0.05, 0.81~0.84
Gamma	No.8	0.48~0.61, 0.98~0.99	No.17	0.01~0.49, 0.97~0.99
Gamma	No.9	0.01~0.32, 0.90~0.91, 0.95~0.99	No.18	0.01~0.17, 0.21~0.32, 0.40~0.42

than the positive group. Based on the constructed correlation matrices, we further calculated the clustering coefficients at different thresholds. We listed the thresholds at which the clustering coefficients were significantly different between the two groups, and we presented the clustering coefficients in detail with a topographic plot for features No. 5 and No. 6 in the beta rhythm. The preliminary analysis implied that the features' ability to reflect the channel correlation may contribute to the recognition of emotions. Nevertheless, considering the differences in the clustering coefficients of the different features, we should note that the correlation analysis is not sufficient to fully explain the mechanism of those features or to determine

the important locations. Additional analyses from different perspectives using different approaches are still needed in future work.

In the future, we should also further study the oscillatory and temporal process of emotion perception based on these features and verify the effectiveness of the proposed features on other datasets. In addition to correlation analysis, we need an in-depth study of the mechanisms of the features that allow for differentiating between positive and negative emotions. Another potential line of research is to further verify the ability of those features to identify emotion-related mental disorders, e.g., depression, as well as the

FIGURE 9 | (A,B) The topographic plot of the clustering coefficient of the groups of negative emotion and positive emotion when feature No. 5 (Hjorth parameter: mobility) and feature No. 6 (Hjorth parameter: complexity) in beta rhythm were utilized. Conditions with different thresholds (T) are illustrated.

effectiveness of those features in studying other cognitive processes.

AUTHOR CONTRIBUTIONS

XL proposed the idea, conducted the experiments, and wrote the manuscript. DS, PZ, YZ, and YH provided advice on the research approaches, guided the experiments, and checked and revised the manuscript. BH offered important help on EEG processing and analysis methods.

REFERENCES

Aftanas, L. I., Lotova, N. V., Koshkarov, V. I., Pokrovskaja, V. L., Popov, S. A., and Makhnev, V. P. (1997). Non-linear analysis of emotion EEG: calculation of Kolmogorov entropy and the principal Lyapunov exponent. *Neurosci. Lett.* 226, 13–16. doi: 10.1016/S0304-3940(97)00232-2

Chanel, G., Rebetez, C., Bétrancourt, M., and Pun, T. (2011). Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* 41, 1052–1063. doi: 10.1109/TSMCA.2011.2116000

Coan, J. A., and Allen, J. J. B. (2004). Frontal EEG asymmetry as a moderator and mediator of emotion. *Biol. Psychol.* 67, 7–50. doi: 10.1016/j.biopsycho.2004.03.002

Gross, J. J., and Muñoz, R. F. (1995). Emotion regulation and mental health. *Clin. Psychol. Sci. Pract.* 2, 151–164. doi: 10.1111/j.1468-2850.1995.tb00036.x

Guyon, I., and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182.

Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Mach. Learn.* 46, 389–422. doi: 10.1023/A:1012487302797

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J. D., Blankertz, B., et al. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87, 96–110. doi: 10.1016/j.neuroimage.2013.10.067

Hjorth, B. (1970). EEG analysis based on time domain properties. *Electroencephalogr. Clin. Neurophysiol.* 29, 306–310.

Huang, J., Cai, Y., and Xu, X. (2006). "A filter approach to feature selection based on mutual information," in *Proceedings of 5th IEEE International Conference on Cognitive Informatics*, Vol. 1 (New York, NY: IEEE Press), 84–89.

Khalili, Z., and Moradi, M. H. (2009). "Emotion recognition system using brain and peripheral signals: using correlation dimension to improve the results of EEG," in *Proceedings of the 2009 International Joint Conference on Neural Networks* (New York, NY: IEEE Press), 1571–1575.

Kim, J. (2007). "Chapter 15: Bimodal emotion recognition using speech and physiological changes," in *Robust Speech Recognition and Understanding*, eds M. Grimm and K. Kroschel (Rijeka: InTech Press), 265–280.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., et al. (2012). DEAP: A database for emotion analysis using physiological signals. *IEEE Trans. Affect. Comput.* 3, 18–31. doi: 10.1109/T-AFFC.2011.15

Kortelainen, J., and Seppänen, T. (2013). "EEG-based recognition of video-induced emotions: selecting subject-independent feature set," in *Proceedings of the 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (New York, NY: IEEE Press), 4287–4290.

Kortelainen, J., Väyrynen, E., and Seppänen, T. (2015). High-frequency electroencephalographic activity in left temporal area is associated with pleasant emotion induced by video clips. *Comput. Intellig. Neurosci.* 2015:31. doi: 10.1155/2015/762769

Li, M., and Lu, B.-L. (2009). "Emotion classification based on gamma-band EEG," in *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (New York, NY: IEEE Press), 1223–1226.

Li, X., Ouyang, G., and Richards, D. (2007). Predictability analysis of absence seizures with permutation entropy. *Epilepsy Res.* 77, 70–74. doi: 10.1016/j.eplepsyres.2007.08.002

Lin, Y. P., Wang, C. H., Jung, T. P., Wu, T. L., Jeng, S. K., Duann, J. R., et al. (2010). EEG-based emotion recognition in music listening. *IEEE Trans. Biomed. Eng.* 57, 1798–1806. doi: 10.1109/TBME.2010.2048568

Lu, Y., Jiang, D., Jia, X., Qiu, Y., Zhu, Y., Thakor, N., et al. (2008). "Predict the neurological recovery under hypothermia after cardiac arrest using C0 complexity measure of EEG signals," in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (New York, NY: IEEE Press), 2133–2136.

Maldonado, S., and Weber, R. (2008). A wrapper method for feature selection using support vector machines. *Inform. Sci.* 179, 2208–2217. doi: 10.1016/j.ins.2009.02.014

Mao, X., and Li, Z. (2009). "Implementing emotion-based user-aware e-learning," in *Extended Abstracts Proceedings of the 2009 Conference on Human Factors in Computing Systems* (New York, NY: ACM Press), 3787–3792.

Moshfeghi, Y. (2012). *Role of Emotion in Information Retrieval*, PhD thesis, University of Glasgow, Glasgow.

Müller, M. M., Keil, A., Gruber, T., and Elbert, T. (1999). Processing of affective pictures modulates right-hemispheric gamma band EEG activity. *Clin. Neurophysiol.* 110, 1913–1920. doi: 10.1016/S1388-2457(99)00151-0

Ng, A. Y. (2004). "Feature selection, L1 vs. L2 regularization, and rotational invariance," in *Proceedings of the 21st International Conference on Machine Learning* (New York, NY: ACM Press), 78–85.

O'Leary, A. (1990). Stress, emotion, and human immune function. *Psychol. Bull.* 108, 363–382.

Pincus, S. M., Gladstone, I. M., and Ehrenkranz, R. A. (1991). A regularity statistic for medical data analysis. *J. Clin. Monit. Comput.* 7, 335–345.

Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178.

Sanei, S., and Chambers, J. A. (2013). *EEG Signal Processing*. Chichester: John Wiley & Sons Ltd.

Schmidt, L. A., and Trainor, L. J. (2001). Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cogn. Emot.* 15, 487–500. doi: 10.1080/02699930126048

Soleymani, M., Pantic, M., and Pun, T. (2012). Multimodal emotion recognition in response to videos. *IEEE Trans. Affect. Comput.* 3, 211–223. doi: 10.1109/T-AFFC.2011.37

Stam, C. J. (2005). Nonlinear dynamical analysis of EEG and MEG: review of an emerging field. *Clin. Neurophysiol.* 116, 2266–2301. doi: 10.1016/j.clinph.2005.06.011

Übeyli, E. D. (2010). Lyapunov exponents/probabilistic neural networks for analysis of EEG signals. *Expert Syst. Appl.* 37, 985–992. doi: 10.1016/j.eswa.2009.05.078

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cogn. Sci.* 7, 553–559. doi: 10.1016/j.tics.2003.10.012

Zhang, A., Yang, B., and Huang, L. (2008). "Feature extraction of EEG signals using power spectral entropy," in *Proceedings of the 2008 International Conference on Biomedical Engineering and Informatics* (New York, NY: IEEE Press), 435–439.

Zhang, X. P., Fan, Y. L., and Yang, Y. (2009). On the classification of consciousness tasks based on the EEG singular spectrum entropy. *Comput. Eng. Sci.* 31, 117–120. doi: 10.3969/j.issn.1007-130X.2009.12.035

Zheng, W. L., Zhu, J. Y., and Lu, B. L. (2016). Identifying stable patterns over time for emotion recognition from EEG. *arXiv preprint arXiv:1601.02197.*

Zhu, J., Rosset, S., Tibshirani, R., and Hastie, T. J. (2004). "1-norm support vector machines," in *Proceedings of the 16th International Conference on Neural Information Processing Systems* (Cambridge: MIT Press), 49–56.

Zhu, J. Y., Zheng, W. L., and Lu, B. L. (2015). "Cross-subject and cross-gender emotion classification from EEG," in *Proceedings of the World Congress on Medical Physics and Biomedical Engineering* (Berlin: Springer Science and Business Media), 1188–1191.