IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 2, 2019, accepted September 23, 2019, date of publication September 27, 2019, date of current version October 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944273

Channel Selection Method for EEG Emotion Recognition Using Normalized Mutual Information

ZHONG-MIN WANG[1,2], SHU-YUAN HU[1], AND HUI SONG[1,2]
[1]School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an 710121, China
[2]Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an 710121, China

Corresponding author: Shu-Yuan Hu (shuyuanhu01@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61373116, and in part by the Special Scientific Research Program of Education Department of Shaanxi Province, China, under Grant 18JK0698.

ABSTRACT Electroencephalography (EEG) signals can reflect activities of the human brain and represent different emotional states. However, recognizing emotions based on full-channel EEG signals will lead to redundant data and hardware complexity, thus it is not suitable for designing wearable devices for daily-life emotion recognition. This paper proposes a channel selection method to select an optimal subset of EEG channels by using normalized mutual information (NMI). Compared with other methods, the proposed method solves the problem of obtaining a higher recognition rate while reducing EEG channels sharply. First, EEG signals are sliced into fixed-length pieces with a sliding window, and short-time Fourier transform is adopted to capture EEG spectrogram. Then inter-channel connection matrix is calculated based on NMI, and channel reduction is conducted by using thresholding and connection matrix analysis. The experiments are based on the widely-used emotion recognition database DEAP. It can be derived from the experimental results that the proposed method can select optimal EEG channel subsets to a certain number while maintaining high accuracy of 74.41% for valence and 73.64% for arousal with support vector machines. Further analysis also reveals that the distribution of the selected channels is consistent with cortical areas for general emotion tasks.

INDEX TERMS Channel selection, electroencephalography, emotion recognition, normalized mutual information, support vector machine.

I. INTRODUCTION

Emotion is the most important component of human, and plays a significant role in people's daily communication [1]. Basically, emotion can be expressed verbally like speech or non-verbally like the facial expression and physiological signals [2]. However, voice and facial expression are not reliable indicators of emotion because they are usually subjective and indirectly reflect the brain activities. Compared with the nonverbally or verbal behaviors, physiological signals are bioelectrical signals that are controlled by the autonomic nervous system of humans and they are not affected by humans' subjective factors, and thus they can objectively and truly reflect the emotional states of people stimulated by the external environment. Generally, physiological signals include Electroencephalography (EEG), Electrocardiogram (ECG), Electromyography (EMG), Galvanic Skin Response (GSR), Blood Volume Pressure, and Skin Temperature [3] etc. EEG signals are directly recorded from human's brain cortex and hence they could be more reliable and objectively in reflecting the inner physiological states of the brain, and they have been widely applied into some fields such as emotion recognition [4], [5], brain-computer interface [6], [7], and epilepsy detection [8], [9]. Therefore, it is feasible to use EEG signals for emotion recognition.

In general, the obtained EEG signals are basically of multi-channel nature. To classify these signals, for example, we have two strategies: one is to work on a subset of channels selected based on certain standards, and the other is to work on all channels [10]. Using full-channel signals of EEG not only initiates many complex features, but also introduces interference information from irrelevant channels, and reduces the system robustness. Therefore, efficient channel selection algorithm is needed to reduce the computation

The associate editor coordinating the review of this manuscript and approving it for publication was Ramana Vinjamuri.

complexity and reduce the over-fitting problem that may be caused by the irrelevant channels, to improve the performance of the system [11]. In addition, there are certain areas in the brain that are concerned with emotions.

Mutual information is an information entropy based on information theory, which can describe the interaction of different brain regions from the perspective of information transmission. Over the decades, mutual information has been widely applied for image registration based on the marginal and joint entropies [12]. Moreover, its implementation has been applied to the field of data mining and grouping similar data recorders [13]. Recently, normalized mutual information (NMI) has been proposed for feature selection with the advantage of reduced complexity of features, and it has achieved excellent results [14]. The implementation of NMI algorithm in EEG signal processing provide us with a method to measure the relationship between features in EEG channels and their corresponding emotions [15].

In this paper, a new channel selection method to classify valence and arousal emotions using normalized mutual information to select the optimal subset of the EEG channels is presented. Specifically, the EEG signals expressed in time-frequency domain are more meaningful than time domain [16]. In our model, we first slice the EEG signals into fixed-length piece with a sliding window, and adopt short-time Fourier transform (STFT) as the preprocessing step to capture the time-frequency information, denoted as EEG spectrogram. Then inter-channel connection matrix is calculated based on NMI, and channel reduction are conducted by using thresholding and connection matrix analysis. The effect of channel selection on the classification results with support vector machine (SVM) algorithm is also discussed together with the channel location related to valence and arousal emotions. In summary, our contributions are as follows:

1) We develop the EEG spectrogram representation model as preprocessing, which contains the emotion information between time and frequency domain. The generated 2-D time-frequency-based fragments are suitable for latent feature extraction.
2) We propose the EEG channel selection procedure which use an NMI-based method to determine the critical channels. It can guide wearable devices configuration and improve data processing for daily-life EEG emotion recognition.
3) We construct the connection matrix by NMI, and finally the optimal channels are obtained. The results show that the proposed model can reduce the EEG channels sharply and maintain a relatively high accuracy compared with other methods.

The remainder of this paper is organized as follows. In section II, we give a brief overview for the related work of EEG channel selection methods. In section III, we propose our method for EEG channel selection and emotion recognition. Experimental results are shown in Section IV. Finally, the whole paper is concluded in Section V.

II. RELATED WORK

EEG channel selection can be treated as a feature selection problem. However, unlike usual feature selection, it is essential to treat all features coming from a channel together [17], because each EEG channel may contain more than one feature. Some related methods for evaluating channels such as the wrapper, filter, embedded and hybrid are introduced in the published literature [11]. In wrapper approaches, using a classification algorithm to assess the candidate channel subsets generated by a search algorithm [11]. For instance, feature selection combined with classification algorithms such as SVM classifiers [18]. Hoverer, it also tends to be more computationally expensive and leads to the overfitting problem. An independent evaluation standard is used to the filtering techniques, and the evaluation standard includes distance measure, information measure, dependency measure, and consistency measure etc. [11]. Filtering techniques have some advantages of the high speed, independence from the classifier, and scalability [19], but they suffer from low accuracy, because it takes into account the fusion of different channels [11]. In the embedded techniques, the selection process is included in the construction of the classifier, and the criteria is used to select the channel [20] in the classifier learning process. Embedded techniques acquire an interaction between the channel selection and the classification. A hybrid approach is fusion of a filtering and a wrapper attempting to take advantage of the two models to avoid the pre-specification of a stopping criterion [11]. In general, independent measure and mining algorithm are used to the hybrid approach, which can evaluate the available channel subsets [19].

In EEG emotion recognition, EEG channel selection has been proved to be the main factor affecting the performance of emotion recognition with the deepening of research [21]. The main task of EEG channel selection is to select a part of electrodes from all electrodes to reduce the computational cost and improve the accuracy rate of emotion recognition [22]. In the past few years, some EEG channel selection algorithms have been presented by many researchers. For instance, Rizon et al [23] proposed an asymmetric ratio (AR) based channel selection method for human emotion recognition from EEG signals. The results show their method can reduce channels and classify the emotions effectively. Lin et al [24] adopted the F-score index based on the ratio of between-class and within-class variances to find a set of optimal EEG channels for EEG-based emotion recognition. He et al [25] presented a Rayleigh coefficient (RC) maximization-based genetic algorithm (GA) for channel selection in motor-imagery BCI system, and achieve the optimal subset of all channels. J Zhang et al [26] adopted the ReliefF-based method to obtain the EEG channel that has the closest relationship with the emotion, and sharply reduce the number of EEG channels without sacrificing the recognition rate. Zheng et al [22] proposed a method based on deep neural networks to learn the average absolute weight distribution to select the optimal EEG channels, and obtain

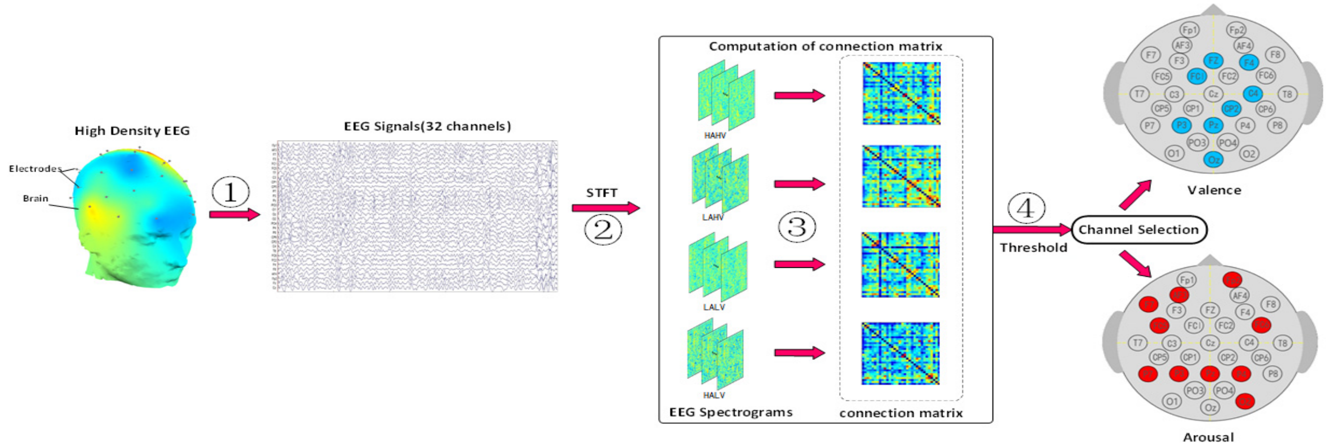


FIGURE 1. **Flowchart of the overall approach.**

preferable experimental results. In recent years, new methods are proposed. Gupta et al [27] proposed a flexible analytic wavelet transform (FAWT) based on six known channels for emotion recognition. The results have shown better performance for emotion classification as compared to the existing method. Bajaj et al [28] proposed multiwavelets decomposition based features for EEG emotion classification, and the result is performed well. Bajaj et al [29] proposed a new method for emotion recognition using multiwavelet transform with multiclass least squares support vector machine (MC-LS-SVM), and it provided classification accuracy of 84.79% for emotions. The results show the effectiveness of the proposed method for EEG emotion recognition.

In short, in the previously presented work, researchers proposed various channel selection methods to select the optimal channel subsets. Different types of classifiers have been applied to recognize emotion, and improve the accuracy of the classification.

III. CHANNEL SELECTION METHOD USING NORMALIZED MUTUAL INFORMATION

A. METHOD OVERVIEW

the overall approach pipeline of proposed channel selection with NMI is illustrated in Fig. 1. The proposed process consist of 4 main steps: ① High density EEG data were collected. ② The EEG spectrogram were generated by STFT. ③NMI connection matrix is computed with spectrogram. ④ Channel selection using thresholding.

B. EEG SPECTROGRAM REPRESENTATION

In our model, we use STFT to convert EEG signals into EEG spectrogram representations in time-frequency domain. Given EEG fragment $x(t)$ of one single-channel, the STFT of the signal in a continuous form is defined as [30]:

$$STFT_x(\tau, \omega) = \int_{-\infty}^{+\infty} x(t)\, \omega(t-\tau)\, e^{-j\omega t} dt, \tag{1}$$

where $\omega(t)$ is the window function centered around zero, and τ is the time index to obtain time localization by taking Fourier transform of the windowed signal. The spectrogram can be further calculated as the magnitude squared of the STFT to transform the complex values.

$$spectrogram_x(\tau, \omega) = |STFT_x(\tau, \omega)|^2, \tag{2}$$

The generated spectrogram is a matrix that reflect the energy distribution of the signal at different frequencies. In this way, by generating the EEG spectrogram representation, the frequency content of EEG signals can be described over time, which can be further learned as time-frequency images by the proposed model.

C. CHANNEL SELECTION WITH NMI CONNECTION MATRIX

Mutual information (MI) is the measurement of mutual dependence (amount of information) between two variables. Given a discrete random variable X has $\{x_1, x_2, \ldots, x_n\}$ random states with probability distribution of $\{p_x(x_1), p_x(x_2), \ldots, p_x(x_n)\}$, according to concept of Shannon's entropy, the information entropy of random X can be defined as:

$$H(X) = -\sum_{i=1}^{n} p_x(x_i) \log_2 p_x(x_i), \tag{3}$$

Similarly, the information entropy of random Y can be defined as:

$$H(Y) = -\sum_{i=1}^{n} p_y(y_i) \log_2 p_y(y_i), \tag{4}$$

where $p_x(x_i)$ represents the probability distribution function of i^{th} random state of X, and $p_y(y_i)$ represents the probability distribution of i^{th} random state of Y. the joint Shannon's entropy of X and Y can be presented as:

$$H(X,Y) = -\sum_{i=1}^{n} p_{xy}\left(x_i, y_i\right) \log_2 p_{xy}(x_i, y_i), \tag{5}$$

where $p_{xy}(x_i y_i)$ is the joint probability function of X and Y. Mutual information of two channel signals can be formulated as [12]:

$$\mathrm{MI}(\mathrm{X}, \mathrm{Y}) = H(X) + H(Y) - H(X, Y), \tag{6}$$

The NMI is normalized by MI to obtain a value between 0 (independence) and 1 (strong dependence) with the equation being expressed as:

$$NMI(\mathrm{X}, \mathrm{Y}) = \frac{MI(X, Y)}{H(X) + H(Y)}, \tag{7}$$

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

The MI has two main properties that distinguish it from other dependency measure: first, the capacity of measuring any kind of relationship between variables; second, its invariance under space transformations [31]. If the mutual information between two random variables in large, it means two variables are closely related. Indeed, MI is zero if an only if two random variables are strictly independent.

Generation of NMI connection matrices can be represented as:

$$G_n = \int_{i=1}^{N} NMI_i(X, Y), \tag{8}$$

where NMI_i is the connection matrix of the i^{th} the sample, N is the number of samples with $i = 1,2,\ldots,\mathrm{N}$, G_n is the connection matrix sum of all the samples.

Setting an appropriate threshold, channel selection is performed for G_n, and select the optimal channel sets according to the performance of each channel for the accuracy of emotion recognition.

D. SVM CLASSIFICATION ALGORITHM

SVM [32] is a simple and efficient computation of machine learning algorithm which is usually applied to classification and regression, and its aim is to find a classification hyperplane to distinguish different types of label. The classification hyperplane can be defined as:

$$w^T x + b = 0, \tag{9}$$

where ω is the normal vector and b is the displacement. And the distance between two different types of support vector and the hyperplane is defined to be γ, which can be expressed as:

$$\gamma = \frac{2}{(\|w\|)}, \tag{10}$$

To realize maximal γ, ω and b should be limited to satisfy the condition of

$$\min_{w,b} \frac{1}{2}\|\omega\|^2$$
$$s.t.\ y_i\left(w^{\mathrm{T}} x_i + b\right) \geq 1, \quad i = 1, 2, \ldots, m. \tag{11}$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In our work, NMI is used to construct connection matrix, and thresholding is employed for EEG channel selection to improve the performance of emotion classification. The proposed method is tested based on the DEAP dataset. In the experiment, we use full channels and the selected channels for emotion recognition, respectively, and the relationship between the emotion recognition rate and the selected channels for emotion recognition, respectively, and the relationship between the emotion recognition rate and the selected channels is analyzed.

A. EEG DATASET

All the EEG data we used for the experiment are from the DEAP database [33], which is especially for emotion analysis using physiological signals. It consists of original data whose sampling frequency is 512 Hz and preprocessed data with sampling frequency of 128 Hz collected from 32 healthy participants, including 16 males and 16 females. Each participant watched a one-minute long music video. After each trial/video, each participant performs self-assessment of their level of arousal, valence, like/dislike, and dominance. Therefore, there are 40 trials for each subject and each trial lasts for 63s, with 3s pretrial included. The DEAP dataset has 40 channels containing 32 EEG channels and 8 other peripheral channels, and we just only use 32 EEG channels. In this paper, we only pay attention to the two dimensions of emotional valence and arousal, and EEG signals are sliced into 60s pieces with a sliding window. Each participant's file contains two arrays as described in Table 1.

TABLE 1. DEAP arrays of each participant.

Array Name	Array shape	Array contents
Data	40× 40 × 8064	video/trail×channel×data
Labels	40× 4	video/trail× label

B. DEFINITION OF EMOTION STATES

In the experiment, emotion states rated ranging from 1 to 9 on the level of arousal-valence, and four conditions of emotional states are defined on the arousal-valence plane (AV plane). High-Arousal High-Valence (HAHV), High-Arousal Low-Valence (HALV), Low-Arousal Low Valence (LALV), and Low-Arousal High-Valence (LAHV). It is shown in Fig. 2. If the score on the dimension is larger than 5, the emotional stated is defined as High, otherwise it is defined as Low.

C. EXPERIMENTAL SETUP

1) NMI connection matrix: In this work, we calculate the NMI connection matrix of the four emotional states of HAHV, LAHV, LALV, and HALV as shown in Table 2. Since the value of NMI is between 0 and 1, 0 means that the two electrode signals are completely uncorrelated while 1 means

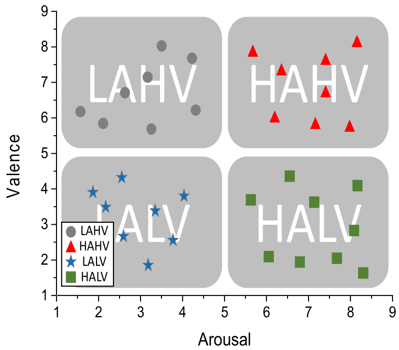
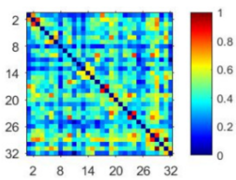
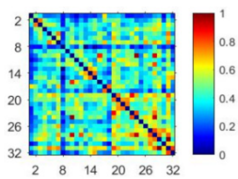
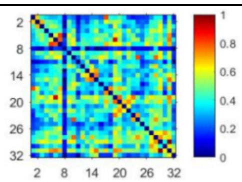
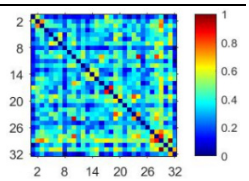


FIGURE 2. The distribution on the arousal-valence plane for the four conditions.

TABLE 2. NMI connection matrix.

HAHV	HALV
	
LALV	LAHV
	

completely correlated. In the work, in order to facilitate the calculation, we have de-diagonalized the NMI connection matrix.

By analyzing the NMI connection matrix of these four labels of emotional state, compared with LAHV state, it can be found that the number of electrodes with NMI values greater than 0.6 is the least, indicating a low correlation between channels in LAHV state. Therefore, on the whole, in HALV state, the range of activated brain regions is wider, forming a higher degree of intracerebral in formation correlation. In addition, it can be seen from the distribution of color patches that HAHV and HALV, LALV and LAHV are different in the arousal dimension. It can be estimated that arousal is the main factor affecting the correlation between channels. Similarly, the difference of HAHV and LAHV, HALV and LALV are significantly different in the valence dimension, and it can be supposed that the valence dominates the valence-arousal model, and its emotion classification rate should be more accurate than the arousal.

2) Implementation Details: We sum four types of NMI connection matrix and then average the value, the purpose of which is to highlight the most obvious features and channels of the four matrices. Moreover, we combine the four emotional states of HAHV, HALV, LALV and LAHV into two emotional states, namely, valence and arousal. Specially, one group is (HAHV, HALV) and (LALV, LAHV), and it represents the classification of arousal, and another group is (HAHV, LAHV) and (HALV, LALV), which represents the classification of valence. We then use the connection matrix to make emotional recognition of the full channels and the selected channels for arousal and valence, respectively. In the experiment, 80% samples of the dataset are used for training and 20% samples of dataset are used for testing.

D. RESULTS OF CHANNEL SELECTION

We get the mean classification accuracy of valence and arousal using different number of channels respectively, as shown in Fig. 3 As shown in Figure 3, when using all channels (32 channels), that is point B, the average recognition accuracy for valence reaches the maximum of 75.16% while 8 channels are enough to obtain 74.41% (point A) of classification accuracy. Meanwhile, for the arousal dimension, when using all channels (32 channels), that is point B', the average recognition accuracy reaches the maximum of 74.41% while 10 channels can also obtain 73.64% (point A') of classification accuracy. As the number of channels increase, the average recognition accuracy of arousal and valence changes little. The result show that we can select 8 channels for valence and 10 channels for arousal. It means we can significantly reduce the number of channels used without a significant decrease in the classification accuracy for recognizing emotions.

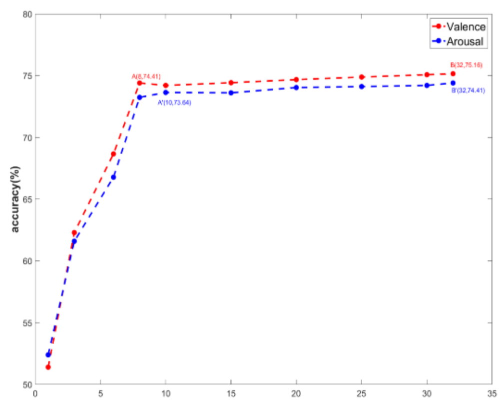


FIGURE 3. SVM-based channel number and accuracy rate.

Since we have determined the optimal channel number for the arousal and valence emotion, we then select the optimal channels of arousal and valence under different thresholds from the experiment, and obtain the best recognition rate under the optimal channels. As is shown in Table 3.

Table 3 summarizes the results of channel selection for arousal and valence with different thresholds. It is observed that with different thresholds, different channel set were obtained. SVM classifier is used to separately obtain the recognition rate of arousal and valence to evaluate the results of channel selection procedure. It is observed that the best

TABLE 3. Channels selected by NMI connection matrix.

	Thresholds	Selected Channels	Accuracy
Arousal	0.65	AF3,F3,FC5,Oz,O2,P4,Fp2,Fz,T8,Cz	71.63%
	0.70	AF3,F7,FC5,P7,Pz,O2,P4,Fp2,FC6,P3	73.64%
	0.75	AF3,FC5,C3,P3,O1,O2,P4,Fp2,C4,CP1	72.67%
	0.80	AF3,F7,FC5,C3,P3,Oz,P4,T8,FC2,PO3	71.46%
	0.85	AF3,F7,PO3,O2,P4,C4,T8,F4,O1,Fz	70.36%
Valence	0.65	F3,FC5,O1,P4,CP6,F4,AF4,Fz	72.31%
	0.70	AF3,F7,Pz,Oz,CP2,F4,Fz,Cz	73.47%
	0.75	FC1,P3,Pz,Oz,CP2,C4,F4,Fz	74.41%
	0.80	F7,P3,Oz,P4,C4,T8,Fz,Cz	73.62%
	0.85	CP5,Pz,O1,P4,C4,T8,F8,Fz	71.84%

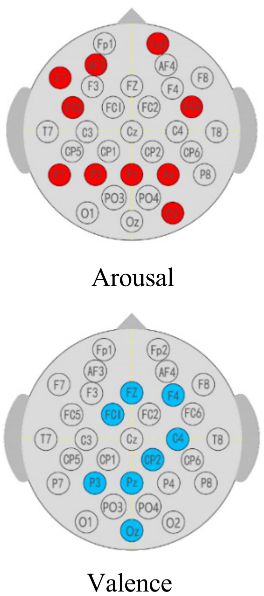


FIGURE 4. Location of EEG channels for valence and arousal emotion recognition after channel selection.

result obtained with threshold = 0.7 for arousal and threshold = 0.75 for valence. Finally, we select the optimal channels for valence and arousal. Valence emotion applies: FC1, P3, Pz, Oz, CP2, C4, F4, Fz channels. On the other hand, arousal emotion uses: AF3, F7, FC5, P7, Pz, O2, P4, Fp2, FC6, P3 channels.

The location of EEG channels for valence and arousal emotions after channel selection using the proposed method is illustrated in Fig. 4. It reveals that recognition of valence and arousal emotion involves a different combination of EEG channels. In addition, the location of these channels in the brain region is in Table 4.

According to Fig. 4 and Table 4, we find that EEG pathways related to emotions are mostly distributed in the front of the frontal lobe, the central lobe, the parietal lobe and the posterior occipital lobe, these regions are consistent with the physiological principles of emotional production. Especially, for valence emotion, which is related to the middle left and right hemisphere. On the other hand, for arousal emotion, which coincides with frontal and parietal lobes of the brain.

TABLE 4. Selected channels with brain region.

Brain region	Channel name
Frontal	AF3,F7,FC1,FC5,FC6,F4,Fp2
Parietal	P7,P3,Pz,P4
Occipital	Oz,O2
Central	CP2,C4

E. COMPARISONS AND DISCUSSIONS

Based on the same database DEAP, we compare the proposed channel selection method with the existing method listed in [34], ReliefF, mRMR (min-Redundancy-Max-Relevance) and the DE method. The comparison results are shown in Table 5 and Table 6.

As shown in Table 5 and Table 6, ReliefF, mRMR and DE require a large number of channels to participate in the emotion recognition. However, by using the proposed channel selection method, the channel number falls down to 1/3-1/2. The final 8 or 10 channels can be utilized in daily life scene for emotion states monitoring. At the same time, our method obtains a relatively high accuracy of 74.41% for valence and of 73.64% for arousal with selected channels, respectively. It is higher than the related work.

Furthermore, comparison results with related work using EEG signals in DEAP dataset are displayed in Table 7.

TABLE 5. Valence: Comparison with existing methods.

Method	Channel Num	Channel Subset	Accuracy
Relief	23	Fp1,AF3,F7,F3,FC1,FC5,T7,CP5,P3,Pz,O2, P4,P8,CP6,CP2,C4,T8, FC6,FC2,F4,F8,Fp2,Fz	64.21%
mRMR	16	Fp1,AF3,Oz,O2,P4,P8, CP2,C4,FC6,FC2,F4,F8, AF4,Fp2,Fz,Cz	62.89%
DE	23	AF3,F3,FC1,FC5,T7,C3,CP1,CP5,P7,Pz,PO3,Oz, PO4,CP6,CP2,C4,T8,FC6, FC2,F4,AF4,Fp2,Fz	72.63%
Our method	8	FC1,P3,Pz,Oz,CP2,C4,F4,Fz	74.41%

TABLE 6. Arousal: Comparison with existing methods.

Method	Channel Num	Channel Subset	Accuracy
Relief	19	Fp1,AF3,FC1,C3,CP1,CP5,P7,P3,Pz,PO3,O1,Oz, CP6,T8,FC6,F8,AF4,Fz,Cz	57.37%
mRMR	15	Fp1,PO3,Oz,O2,P4,P8, CP6,CP2,C4,FC2,F4,F8, AF4,Fp2,Fz	62.89%
DE	19	Fp1,AF3,F7,F3,FC5,C3,CP5,P7,P3,Pz,O1,PO4, P4,P8,CP2,C4,T8,FC2,Fz	73.94%
Our method	10	AF3,F7,FC5,P3,P7,Pz,O2,P4,FC6,Fp2	73.64%

TABLE 7. Comparison with related work using EEG signals in DEAP dataset.

Authors	Approach	Accuracy (%)
Koelstra et al. [33]	Power spectral features	57.6(valence) 62.0(arousal)
Chung et al. [35]	Power spectral analysis with Bayes classifier	66.6(valence) 66.4(arousal)
Yoon et al. [36]	FFT enhanced feature extraction and classification	70.9(valence) 70.1(arousal)
Zhuang et al. [37]	Spectral Power	70.9(valence) 67.1(arousal)
Candra et al. [38]	Three band wavelet entropy	65.1(valence) 64.8(arousal)
Verma et al. [39]	Mltireslolution analysis and multilayer perceptron	63.5(valence) 69.6(arousal)
Our method	NMI-based channel selection	74.41(valence) 73.64(arousal)

Our research selects 8 channels for valence and 10 channels for arousal to obtain an accuracy which is higher than other 2-class valence-arousal recognition listed in Table 7. The experiment results indicate that the proposed method can select the key channels for daily-life EEG emotion recognition and ensure a relatively high accuracy.

Computational complexity is also analyzed. From a qualitative point of view, each channel has a large amount of data. After we conducted the channel selection experiment, the number of channels reduced sharply, and the overall data is also greatly reduced. Therefore, the computational complexity is certainly reduced. On the other hand, from the perspective of complex network theory, the connection between electrodes can constitute a brain network. Before the channel selection, the amount of data is large, and 32 full channels are connected to each other to form a fully connected brain network, which has a complicated structure. But after channel selection, only a few channels related to emotions were retained, and the amount of data is greatly reduced. Then the connection relationship between channels is relatively simple and clear, indicating that the proposed method is indeed efficient and reasonable, and the computation complexity is also significantly reduced.

Analysis of the classification results of the all 32 channels and the selected channels is shown in Fig. 5 The figure shows the valence has higher accuracy (75.16% to 74.41%) than arousal (74.41% to 73.64%). By using channel selection, the average accuracy of arousal and valence are slightly lower than the recognition rate of all channels. However, we don't sacrifice too much recognition accuracy, but channel number falls down to 1/4 of the full channels, and it reduces the computational complexity. Furthermore, the proposed method can be utilized in the daily-life emotion recognition to reduce the complexity of wearable devices structure and improve the recognition performance.

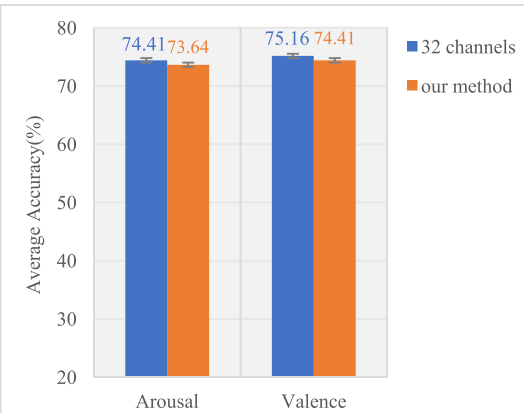


FIGURE 5. Comparison of classification valence and arousal emotion between 32 channels and selected channels.

In other fields, such as epilepsy detection, channel selection is also very important. Bhattacharyya et al [40] proposed a method to use mutual information to select 5 channels for EEG seizure detection. Compared with our method, although both methods use mutual information, the focus of the two methods is quite different. In [40], the channel selection step is to use the least standard deviation (SD) to determine the first channel, and then to calculate the mutual information between the remaining channels and the first selected channel. Take four channels with high mutual information values. Finally, five channels have been selected. In this process, the key is to select the first channel by the least standard deviation, and the mutual information is only used to quantitively calculate the similarity or interdependency between the remaining channels and the first selected channel. In our method, we compute the mutual information between channels and normalized it to get the normalized mutual information (NMI) connection matrix, and then select the channels related to emotion states by the thresholds, and the computation of mutual information is the key step to experiment.

V. CONCLUSION

For daily-life emotion recognition, traditional approaches based on full-channel EEG signals will lead to redundant data and hardware complexity. This paper presents a new channel selection method using NMI to select optimal channels for EEG emotion recognition. SVM classifier is used to classify emotion state. The number of EEG channels can be reduced from 32 to 8 for valence and 32 to 10 for arousal by using the NMI. In addition, we compared our method with other state-of-the-art approaches. The results show the proposed method effectively improves the rate of emotion recognition while reduces the channels sharply. The thresholds are predefined. Finally, we also explore brain regions that are related to emotions, which help us to study the relationship between specific brain regions and emotions. In a real application, these values like thresholds are hard to determine in advance and inappropriate values have a negative effect on the performance of the model. Thus, future directions include the application of self-adaptive strategy which is able to adjust the thresholds for multimodal emotion analysis.

REFERENCES

[1] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Inf. Fusion*, vol. 37, pp. 98–125, Sep. 2017.

[2] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of EEG signals and facial expressions for continuous emotion detection," *IEEE Trans. Affect. Comput.*, vol. 7, no. 1, pp. 17–28, Jan./Mar. 2016.

[3] G. K. Verma and U. S. Tiwary, "Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals," *NeuroImage*, vol. 102, pp. 162–172, Nov. 2014.

[4] W. Zheng, "Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis," *IEEE Trans. Cogn. Devel. Syst.*, vol. 9, no. 3, pp. 281–290, Sep. 2017.

[5] Y. Yang, Q. M. J. Wu, W.-L. Zheng, and B.-L. Lu, "Eeg-based emotion recognition using hierarchical network with subnetwork nodes," *IEEE Trans. Cogn. Devel. Syst.*, vol. 10, no. 2, pp. 408–419, Jun. 2017.

[6] S. Zhao, Z. Li, R. Cui, Y. Kang, F. Sun, and R. Song, "Brain-machine interfacing-based teleoperation of multiple coordinated mobile robots," *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 5161–5170, Jun. 2017.

[7] Z. Li, S. Zhao, J. Duan, C.-Y. Su, C. Yang, and X. Zhao, "Human cooperative wheelchair with brain-machine interaction based on shared control strategy," *IEEE/ASME Trans. Mechatronics*, vol. 22, no. 1, pp. 185–195, Feb. 2017.

[8] M. Wu, T. Wan, X. Wan, Y. Du, and J. She, "Fast, accurate localization of epileptic seizure onset zones based on detection of high-frequency oscillations using improved wavelet transform and matching pursuit methods," *Neural Comput.*, vol. 29, no. 1, pp. 194–219, 2017.

[9] N. Rommens, E. Geertsema, L. J. Holleboom, F. Cox, and G. Visser, "Improving staff response to seizures on the epilepsy monitoring unit with online EEG seizure detection algorithms," *Epilepsy Behav.*, vol. 84, pp. 99–104, Jul. 2018.

[10] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 141–144, Jun. 2003.

[11] T. Alotaiby, F. E. Abd El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing," *EURASIP J. Adv. Signal Process.*, vol. 2015, no. 1, p. 66, 2015.

[12] N. D. Cahill, "Normalized measures of mutual information with general definitions of entropy for multimodal image registration," in *Proc. Int. Workshop Biomed. Image Registration*. Springer, 2010, pp. 258–268.

[13] Z. He, X. Xu, and S. Deng, "k-ANMI: A mutual information based clustering algorithm for categorical data," *Inf. Fusion*, vol. 9, no. 2, pp. 223–233, Apr. 2008.

[14] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Trans. Neural Netw.*, vol. 20, no. 2, pp. 189–201, Feb. 2009.

[15] H. Candra, M. Yuwono, A. Handojoseno, R. Chai, S. Su, and H. T. Nguyen, "Recognizing emotions from EEG subbands using wavelet analysis," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2015, pp. 6030–6033.

[16] Y. Yuan, G. Xun, Q. Suo, K. Jia, and A. Zhang, "Wave2Vec: Learning deep representations for biosignals," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 1159–1164.

[17] 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.*, Apr./May 2011, pp. 667–670.

[18] T. N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Scholkopf, "Support vector channel selection in BCI," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1003–1010, Jun. 2004.

[19] H. Liu and L. Yu, "Toward integrating feature selection algorithms for classification and clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 491–502, Apr. 2005.

[20] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Jan. 2003.

[21] K. Ansari-Asl, G. Chanel, and T. Pun, "A channel selection method for EEG classification in emotion assessment based on synchronization likelihood," in *Proc. 15th Eur. Signal Process. Conf.*, Sep. 2007, pp. 1241–1245.

[22] 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.

[23] M. Rizon, M. Murugappan, and S. Yaacob, "Asymmetric ratio and FCM based salient channel selection for human emotion detection using EEG," Tech. Rep, 2008.

[24] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, Jul. 2010.

[25] L. He, Y. Hu, Y. Li, and D. Li, "Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG," *Neurocomputing*, vol. 121, pp. 423–433, Dec. 2013.

[26] J. Zhang, M. Chen, S. Zhao, S. Hu, Z. Shi, and Y. Cao, "ReliefF-based EEG sensor selection methods for emotion recognition," *Sensors*, vol. 16, no. 10, p. 1558, 2016.

[27] 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, May 2019.

[28] V. Bajaj and R.B. Pachori, "Human emotion classification from EEG signals using multiwavelet transform," in *Proc. Int. Conf. Med. Biometrics*, Shenzhen, China, May/Jun. 2014, pp. 125–130.

[29] V. Bajaj and R. B. Pachori, "Detection of human emotions using features based on the multiwavelet transform of EEG signals," in *Brain-Computer Interfaces*. Cham, Switzerland: Springer, 2015, pp. 215–240.

[30] L. Cohen, "Time-frequency distributions-a review," *Proc. IEEE*, vol. 77, no. 7, pp. 941–981, Jul. 1989.

[31] S. Kullback, *Information Theory and Statistics*. Chelmsford, MA, USA: Courier Corporation, 1997.

[32] V. Vapnik, *The Nature of Statistical Learning Theory*, 1995.

[33] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; Using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan./Mar. 2012.

[34] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 327–339, Jul. 2014.

[35] S. Y. Chung and H. J. Yoon, "Affective classification using Bayesian classifier and supervised learning," in *Proc. 12th Int. Conf. Control, Automat. Syst.*, Oct. 2012, pp. 1768–1771.

[36] H. J. Yoon and S. Y. Chung, "EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm," *Comput. Biol. Med.*, vol. 43, no. 12, pp. 2230–2237, 2013.

[37] X. Zhuang, V. Rozgić, and M. Crystal, "Compact unsupervised EEG response representation for emotion recognition," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Inform.*, Jun. 2014, pp. 736–739.

[38] H. Candra, M. Yuwono, R. Chai, A. Handojoseno, I. Elamvazuthi, H. T. Nguyen, and S. Su, "Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2015, pp. 7250–7253.

[39] G. K. Verma and U. S. Tiwary, "Affect representation and recognition in 3D continuous valence–arousal–dominance space," *Multimedia Tools Appl.*, vol. 76, no. 2, pp. 2159–2183, Jan. 2017.

[40] A. Bhattacharyya and R. B. Pachori, "A multivariate approach for patient-specific EEG seizure detection using empirical wavelet transform," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2003–2015, Sep. 2017.

ZHONG-MIN WANG received the M.S. degree in mechatronic engineering, in 1993, and the Ph.D. degree in mechanical manufacture and automation from the Beijing Institute of Technology, Beijing, China, in 2000. He was with Xidian University, from April 1993 to August 1997. From February 2004 to February 2005, he was a Visiting Scholarship with the Robotics Laboratory of The Australian National University, Canberra, Australia. Since 2000, he has been a Full Professor with the School of Computer Science and Technology, Xi'an University of Posts and Telecommunications. His current research interests include embedded intelligent perception, intelligent information processing, machine learning, brain–computer interface, and affective computing.

SHU-YUAN HU received the B.Eng. degree in communication engineering from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2017, where he is currently pursuing the master's degree in computer technology. His research interests include EEG-based brain–computer interaction, physiological signals analysis for emotion recognition, and affective computing.

HUI SONG received the M.Sc. degree in information and computing systems from the University of Greenwich. He is currently with the Software Engineering Department, School of Computer Science, Xi'an University of Posts and Telecommunications. His interested fields include artificial intelligence, mobile computing, and software modeling.