A recurrence quantification analysis-based channel-frequency convolutional neural network for emotion recognition from EEG

Yu-Xuan Yang, Zhong-Ke Gao, Xin-Min Wang, Yan-Li Li, Jing-Wei Han, Norbert Marwan, and Jürgen Kurths

Citation: Chaos **28**, 085724 (2018); doi: 10.1063/1.5023857
View online: https://doi.org/10.1063/1.5023857
View Table of Contents: http://aip.scitation.org/toc/cha/28/8
Published by the American Institute of Physics

A recurrence quantification analysis-based channel-frequency convolutional neural network for emotion recognition from EEG

Yu-Xuan Yang,[1] Zhong-Ke Gao,[1,a)] Xin-Min Wang,[1] Yan-Li Li,[1] Jing-Wei Han,[1] Norbert Marwan,[2] and Jürgen Kurths[2,3]

[1]*School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China*
[2]*Potsdam Institute for Climate Impact Research, Telegraphenberg A31, 14473 Potsdam, Germany*
[3]*Department of Physics, Humboldt University Berlin, 12489 Berlin, Germany*

(Received 29 January 2018; accepted 25 June 2018; published online 31 August 2018)

Constructing a reliable and stable emotion recognition system is a critical but challenging issue for realizing an intelligent human-machine interaction. In this study, we contribute a novel channel-frequency convolutional neural network (CFCNN), combined with recurrence quantification analysis (RQA), for the robust recognition of electroencephalogram (EEG) signals collected from different emotion states. We employ movie clips as the stimuli to induce happiness, sadness, and fear emotions and simultaneously measure the corresponding EEG signals. Then the entropy measures, obtained from the RQA operation on EEG signals of different frequency bands, are fed into the novel CFCNN. The results indicate that our system can provide a high emotion recognition accuracy of 92.24% and a relatively excellent stability as well as a satisfactory Kappa value of 0.884, rendering our system particularly useful for the emotion recognition task. Meanwhile, we compare the performance of the entropy measures, extracted from each frequency band, in distinguishing the three emotion states. We mainly find that emotional features extracted from the gamma band present a considerably higher classification accuracy of 90.51% and a Kappa value of 0.858, proving the high relation between emotional process and gamma frequency band. *Published by AIP Publishing.* https://doi.org/10.1063/1.5023857

Incorporating emotion interaction into the human-machine interaction (HMI) process is of vital importance for building a much more intelligent HMI system. To realize intelligent emotion interaction, one of the most primary requisites is to design a reliable emotion recognition system with high accuracy and robustness. Electroencephalogram (EEG)-based emotion classification systems have been widely used on account of their high accuracy and objective evaluation. We here propose a novel emotion recognition system which combines recurrence quantification analysis (RQA) with channel-frequency convolutional neural network (CFCNN), based on the successful applications of RQA in analyzing nonlinear EEG signals and the remarkable identification ability of convolutional neural network (CNN). The results suggest that our system has a reliable and stable emotion identification capability. In addition, we feed the RQA-based features extracted from each of the five frequency bands into CFCNN to discuss their respective effectiveness and find the critical frequency band which is closely linked to emotional process. The result indicates that emotional arousal is in particular relevant to gamma frequency band activities.

I. INTRODUCTION

As a kind of human psychological and physical state, emotion is caused by a complex comprehensive action of feelings, thoughts, and behaviors and plays a critical role in our daily life especially in human-human interactions. Further, incorporating emotion interaction into the human-machine interaction (HMI) process is of vital importance for building a much more intelligent HMI system, which triggers the appearance of affective computing.[1] The field of affective computing focuses on endowing the machines with the capability of reading and responding to human emotions and further implementing a vivid communication between emotionally abundant and emotionally absent machines.

In order to realize an intelligent emotion interaction, one of the most primary requisites is to design a reliable emotion recognition system with high accuracy and robustness as well as fairly strong applicability. Various approaches based on physical or physiological features, extracted from humans, have been devoted to modeling human emotions including facial,[2–4] vocal,[5–7] body expressions,[8,9] touch behaviors,[10] and physiological changes;[11] of these, methods based on electroencephalogram (EEG) signals are widely used[12–17] on account of their objective evaluation[18] and high temporal resolution. However, the acquired EEG signals are unavoidably contaminated by much noise due to the low signal-to-noise ratio of EEG and high levels of cross-talk from different areas of the brain. Besides, the EEG is affected by the volume conductor effect, and signals from distant sources are distorted while emotions are emanated from the limbic system located far deep in the brain. These result in a challenge for researchers to seek advanced data analysis methods capable of detecting more effective and detailed emotion-related information from the polluted and mixed signals.

A variety of classification models, proposed for automated emotion recognition, employed linear time-domain or frequency-domain features from EEG signals to differentiate

a)Electronic mail: zhongkegao@tju.edu.cn

emotion states. Considering that EEG signals are complex, nonlinear, and dynamic, recently several nonlinear analysis methods have gained much concern as effective tools for detecting inherent dynamical properties of physiological phenomena.[19–24] In this paper, we introduce the nonlinear technique named recurrence plot (RP)[25,26] to study its effectiveness in distinguishing different emotional states. RP is a widely used tool for visualizing the recurrences of dynamical systems in phase space. A RP is a square with several black dots, one of them represents that a phase space vector is close to another one. A RP exhibits typical large-scale and small-scale structures resulted from dynamical behaviors including single drift, dots, as well as diagonal lines, vertical and horizontal lines.[25,27] In order to quantify a RP, Zbilut and Webber[28] provided a method called recurrence quantification analysis (RQA), which actually depicts the number and duration of the appearance of recurrences in a dynamical system and further measures the complexity and nonlinearity of a system.[25,26,28,29] Therefore, RQA is a quite decent tool for characterizing complex dynamics of EEG signals. RQA measure-based EEG analysis has gained a great popularity in diverse research areas, including epileptic identification,[30,31] sleep apnea syndrome analysis,[32] depression diagnosis,[33] memory retrieval research,[34] and anaesthesia monitoring.[35] Those successful applications inspire us to extract typical RQA measures from EEG signals, and feed them into classifiers to distinguish different emotion states and improve the accuracy of our emotion recognition system.

Traditionally, researchers mostly chose shallow models, e.g., support vector machine (SVM), as classifiers to recognize emotion states and have acquired certain achievements.[36,37] However, those shallow models have limitations in learning the inherent characteristics of training samples when encountering complicated classification problems. In 2006, Hinton and Salakhutdinov[38] overcame the explaining-away effects encountered in densely connected belief nets using a greedy algorithm based on complementary priors, which successfully ignites the enthusiasm of researchers to the study of deep learning and deep neural networks. Subsequently, a series of deep network structures with high accuracy and generalization ability have been presented, including convolutional neural network (CNN)[39] and long-short term memory (LSTM) network.[40,41] The great superiority of deep neural networks in learning the intrinsic characteristics, motivated the affective computing researchers to study the possibility of deep network structures settling emotion classification issues and several achievements have been fulfilled.[42–44]

The combination of emotion classification and deep neural network methods can be a promising research topic for improving the accuracy and practicability of emotion recognition systems. In this paper, we first design a movie-induced emotion experiment to obtain EEG signals corresponding to three basic emotion states, namely, happiness, sadness, and fear. Then a RP procedure is employed to each channel of pre-processed EEG signals under five frequency bands. A typical entropy measure is extracted by RQA from each RP, and further, a series of feature vectors is formed by the entropy measure obtained from different channels and frequency bands. Numerous traditional methods simply and directly concatenate those features from different channels and frequency bands to feed the classifier, ignoring the inherent characteristics among channels and leading to dimension disaster which needs to be solved by some complicated algorithms. Therefore, we here propose a novel channel-frequency CNN (CFCNN) to fully utilize the multi-dimensional information and further facilitate emotion recognition task. The results indicate that our RQA-based CFCNN emotion recognition system possesses excellent and stable emotion identification ability and its performance is superior to two traditional combination methods. In addition, we also find that entropy measure from high frequency band activities in EEG is highly indicative of emotional activations.

II. EXPERIMENTAL DESIGN AND DATA ACQUISITION

The experiments were conducted in the Laboratory of Complex Networks and Intelligent Systems at Tianjin University, China. The experimental process was approved by the ethics committee of General Hospital Affiliated to Tianjin Medical University in China. We focused on identifying three specified emotion states, namely, happiness, sadness, and fear. In this work, we chose movie clips to induce subjects' emotions owing to the fact that movies contain vivid visual and auditory stimuli, making them appear to be one of the most effective ways to arouse emotions.[36] We invited 20 volunteers (not the subjects) to assess and grade a series of movie clips, and then eight top-rated emotional movies clips were selected. The concrete names of the eight well-chosen movie clips and their corresponding emotion states and duration time are given in Table I. Five healthy and right-handed students from Tianjin University were chosen as the subjects and all of them have no vision and hearing problems. All the subjects were given written consent prior to the recording and information about the design and purpose of the experiment. We conducted the experiment in the morning, and only the examined subject and one researcher, being responsible for recording the performance of the subject and providing help for the subject, stayed in the quiet laboratory during the recording time. Figure 1 presents the experiment scene.

All the subjects were required to have a good sleep the night before the experiment. In the preparation stage of each experiment, each examined subject was explained the recording procedure and the points for attention such as averting body and facial movements as fully as possible while watching movie clips. The EEG recording device was equipped by

TABLE I. The sources of the selected movie clips and their corresponding emotional labels.

Number	Film clip source	Emotional label	Duration of the clip
1	Find miracle in cell No. 7	Sadness	3 min 25 s
2	Dearest	Sadness	4 min 30 s
3	A hero or not	Happiness	5 min 30 s
4	Mr. Bean	Happiness	7 min 20 s
5	Mr. Bean	Happiness	5 min 52 s
6	Dead silence	Fear	8 min 24 s
7	The Conjuring 2	Fear	4 min
8	Lights out	Fear	3 min 25 s

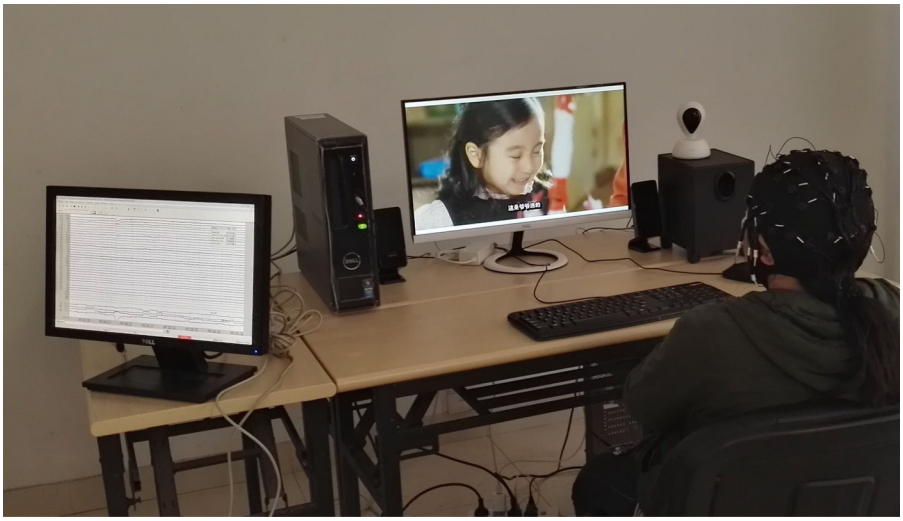

FIG. 1. The real experiment scene.

ESI NeuroScan System with 40 electrodes, arranged according to international 10–20 electrode placement system, and the sampling rate was 1000 Hz. Particularly, A1 and A2 linked at mastoids were chosen as the reference electrodes, and four electrodes (vertically and horizontally placed around eyes) were used to record electrooculogram (EOG) signals. Meanwhile, we applied a smart camera to record the facial expression of subjects during the whole recording stage.

The recording stage was proceeded by the help of the guiding words appearing in the middle of the screen. Each recording stage consisted of eight sections (movie clips). In each section, the guiding words first appeared on the screen for 5 s to remind the start of this section and then, the stage entered the movie screening process during which subjects were informed to concentrate on current movie clip which appeared randomly. After the screening, a series of questions would pop up to help the subject to report their true emotion states and rate the intensity of emotion in 10 point scale with regard to current movie clip. Finally, the subject was directed to have a 30-s break to prepare for the next section.

After the recording stage, the EEG data collected by the ESI NeuroScan System were first preprocessed by the EEGLAB toolbox on Matlab where signals were filtered with a bandpass of 1–50 Hz, and Independent Component Analysis (ICA) was performed for removing eye movement and blink artifacts. After eliminating EOG components, we obtained a series of signals in 30 channels. Then we down-sampled the preprocessed data from 1000 Hz to 200 Hz and divided the data into a series of same-length epochs of 5 s without overlapping. Eventually, there were 171, 222, and 188 clean epochs for the sadness, happiness, and fear category, respectively, and totally, 551 clean epochs for one subject can be obtained.

III. RQA-BASED CFCNN EMOTION RECOGNITION SYSTEM

In this work, we conduct RQA on RPs, obtained from movie-induced EEG signals, to extract measures for characterizing the complexity and characterize dynamics of EEG signals. First of all, a RP is obtained by the following steps. Considering a phase space trajectory x_i, RP is capable of visualizing the recurrence of x_i and can be obtained from calculating the following matrix:

$$R_{i,j} = \Theta\left(\varepsilon - ||x_i - x_j||\right),\ i,j = 1,2,\ldots,N, \tag{1}$$

where $\Theta(\cdot)$ is the Heaviside function, ε is a predefined threshold, $||\cdot||$ represents the maximum norm, and N is the number of vector points of the phase space trajectory. In contrast to many other works, we use the scalar time series instead of the higher-dimensional phase space vectors (embedded time series) in consideration of the possible spurious correlations caused by embedding and the great complexity of choosing appropriate embedding parameters for a set of extremely non-stationary EEG signals.[45–47] A fixed $\varepsilon = 0.3$ is selected according to the work and the discussion in Refs. 47–49, and each time series is normalized to zero mean and unit standard deviation. In order to characterize and quantify the obtained RPs, we perform RQA to extract different measures, including recurrence rate, determinism, entropy, averaged diagonal length, length of the longest diagonal line, laminarity, trapping time, length of longest vertical line, recurrence time of 1st type, and recurrence time of 2nd type.[26] Our test results indicate that the entropy measure provides a remarkable classification accuracy among these measures. So, the remaining parts of the work focus only on the entropy measure. Specifically, the entropy (ENTR) measure is proposed for quantifying the complexity of the RP in terms of diagonal lines[26] and can be defined by

$$\mathrm{ENTR} = -\sum_{l=l_{\min}}^{N} p(l)\ \ln p(l), \tag{2}$$

where

$$p(l) = P(l) / \sum_{l=l_{\min}}^{N} P(l),$$

where $l_{\min}$ is the length of the shortest diagonal line which is set as 2 in this work, $P(l)$ indicates the histogram of diagonal

lines of length l, and $p(l)$ refers to the probability of finding a diagonal line of exactly length l in the RP.

In this paper, channel-wise RQA is operated on each epoch. As specific frequency ranges are more prominent in certain states of mind,[42] we filter the signals into five specific frequency bands (delta: 1–3 Hz, theta: 4–7 Hz, alpha: 8–13 Hz, beta: 14–30 Hz, and gamma: 31–50 Hz) and discuss the characteristics of EEG signals in these five frequency bands. Thus, for each epoch, we obtain an entropy-based feature vector of $30 \times 5 = 150$ (number of channel $\times$ number of frequency bands), respectively. In order to avoid a class-imbalance issue, we randomly selected 170 epochs from each category and their corresponding feature matrixes can form a dataset with 510 samples.

The generated feature vectors corresponding to the three emotion categories are fed into the novel CFCNN, respectively, which has high computational efficiency and low model complexity and has been widely used in solving classification issues. Commonly, a CNN is stacked with several network layers, which generally include the convolution stage, detector stage, and pooling stage.[43] Specifically, the convolution stage focuses on employing different kinds of convolutional filters on original data to extract multiple features and then a series of feature maps with different weight vectors are formed. In this stage, two important architectural ideas, which make CNNs superior to traditional neural networks, are applied to help decrease the amount of weight parameters: (1) local receptive fields and (2) shared weights (among all the units in a feature map). Subsequently, in the detector stage, a nonlinear transformation (e.g., rectified linear activation function)[50] is executed on the feature maps acquired from the convolution stage, which is the key point for the generalization ability. The last pooling operation (e.g., max pooling, average pooling, or stochastic pooling) is chosen to reduce the resolution of the input to the next convolutional layer or the fully connected layer.

Considering the high dimensionality and high resolution of EEG signals, we design a novel CFCNN to refrain from dimension disaster and more importantly, to make full use of the information from multiple channels and different frequency bands to benefit the emotion recognition task. In this paper, the constructed and trained CFCNN model consists of four convolutional layers, two fully connected layers, and a softmax layer, implemented by the deep learning library Keras.[51] Figure 2 shows the detailed architecture of the novel CFCNN. The model inputs each sample in the form of a matrix $X \in R^{E \cdot F}$ formed by E electrodes and F frequency bands and outputs a probability of each category. Thus, the predicted label is the category which corresponds to the maximum probability. The number of filters in the convolutional layers are 16, 32, 64, and 64, respectively. The first three convolutional layers only move forward in channel dimension and the kernel sizes are set as 3×1 with default stride 1. The last convolutional layer is designed to fuse different frequency bands information by extracting high level information with kernel size 1×5. Additionally, each convolutional layer is followed by an activation block consisting of a rectified linear activation[50] and a batch normalization layer.[52] A flatten operation is subsequently conducted after four convolution operations to transform the feature maps into a one-dimensional vector. Then two fully connected layers with size 64 and 3, respectively, are connected to the model following the flatten operation and meanwhile, a softmax layer equipped with the cross-entropy objective function is applied to produce a distribution for each emotion category.

During the model training process, the Glorot_normal initializer[53] is applied to initialize the weights of the convolutional layers. The model optimization process is realized by the stochastic gradient descent (SGD)[54] optimizer with the learning rate of 0.001, decay of 10^{-6}, and momentum of 0.9. Besides, the number of learning iterations is 300 and the batch size is set as 32. A neuron at the position (m, n) of a feature map k in the layer l is denoted by $x_{l,k,(m,n)}$. Similarly, we define $\sigma_{l,k,(m,n)}$ as the scalar product between a set of input neurons and the neurons in the same maps share the same sets of weights

$$x_{l,k,(m,n)} = f(\sigma_{l,k,(m,n)}), \tag{3}$$

where f is a rectified linear function[50] for all the network layers in this paper. We denote the four convolutional layers as L_1, L_2, L_3, and L_4, respectively. For those layers, each neuron of one map shares the same sets of weights and is connected to a subset of neurons from the previous layer. The Appendix provides the detailed information transmission process of CFCNN. We tune up the network weights and biases through the traditional back propagation strategy[55] to achieve

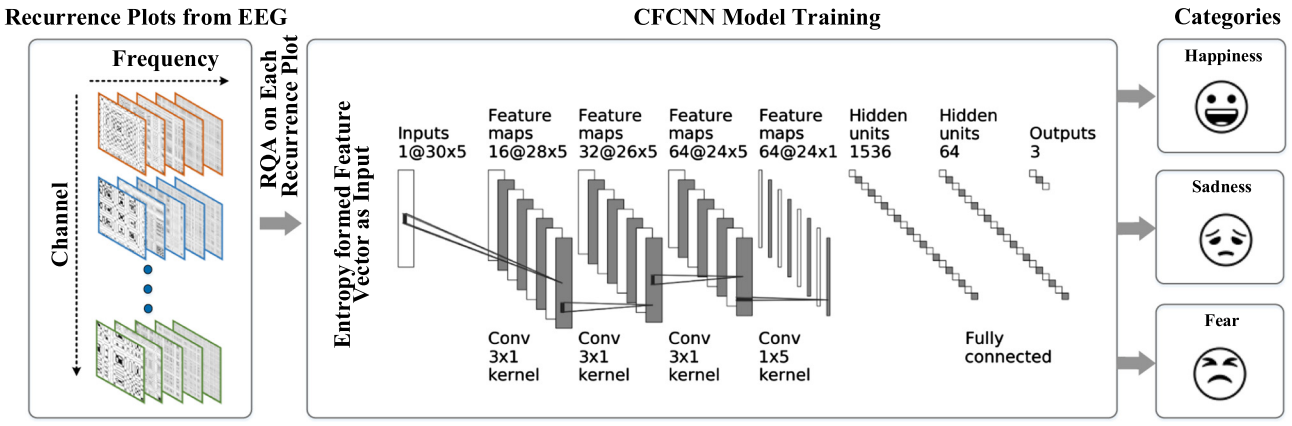

FIG. 2. The framework of the novel emotion recognition system including the architecture of CFCNN.

a better model accuracy for the validation set, and cross-entropy objective function is used as a loss function here to estimate the model performance.

Furthermore, in order to contrast the classification effect of our system, we also extract two traditional features, namely, power spectral density (PSD) and differential entropy (DE) in each frequency band, respectively, and combine them with the classifier SVM, respectively, as the baselines. PSD is defined by

$$P_x(f) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |s(t)e^{-j2\pi ft}dt|^2, \tag{4}$$

where $s(t)$ is the time series, T is a time interval, and f is the frequency. And according to the deduction provided by Zheng and Lu,[42] DE in a certain frequency band can be obtained by calculating the logarithm energy spectrum:

$$h(X) = \frac{1}{2} \log\, 2\pi e\sigma^2, \tag{5}$$

where X is the time series obeying the Gauss distribution $N(\mu, \sigma^2)$.

IV. RESULTS

We use 5-fold cross validation method to evaluate the performance of our system and two baselines. Specifically, for each subject, 80% of 510 samples are selected randomly as the training set and the remaining 20% as the test set. Then, we can obtain five confusion matrices for each subject. The average confusion matrix across five folds from the RQA + CFCNN system of each subject is shown in Fig. 3. The main diagonal entries represent the average number of correct classification samples, and the off-diagonal entries denote the average number of misclassification samples. As can be seen, the results vary among different subjects, resulting from many reasons, including subjects' individuality, body state, and concentration degree during the experiments. However, in general, the three emotions can be recognized with almost the same accuracies for most of the subjects.

In order to better present and compare the performance of the three systems, we calculate the accuracy and Kappa value[56,57] based on each confusion matrix, and the detailed results are shown in Tables II and III. For a confusion matrix A, the Kappa value can be obtained from

$$\text{Kappa} = \frac{p_o - p_e}{1 - p_e}, \tag{6}$$

where $p_o = \frac{\sum_{i=1}^{k} a_{ii}}{N}$ is the accuracy (N is the number of samples, k is the number of category, and a_{ii} is the value of row i column i on a confusion matrix) and $p_e = \frac{\sum_{i=1}^{k} \left(\sum_{j=1}^{k} a_{ij} \times \sum_{j=1}^{k} a_{ji}\right)}{N \times N}$ is the result of random classification (a_{ij} is the value of row i column j on a confusion matrix and a_{ji} is the value of row j column i).

From Table II, we can find that each of the results obtained by our system exceeds both the two baselines. Further, the average performance of our system is 92.24%, which is 19.65% and 6.83% higher than that of PSD + SVM and DE + SVM, respectively. In addition, the with-subject standard deviation for our system is 2.11% which is 0.35% and 1.33% lower than that for two baselines. Besides, the inter-subject standard deviation for our system is 3.71% which is also lower than that for the two baselines, which shows that our system is relatively more robust to subject dependent differences than the two baselines. As can be seen from Table III,

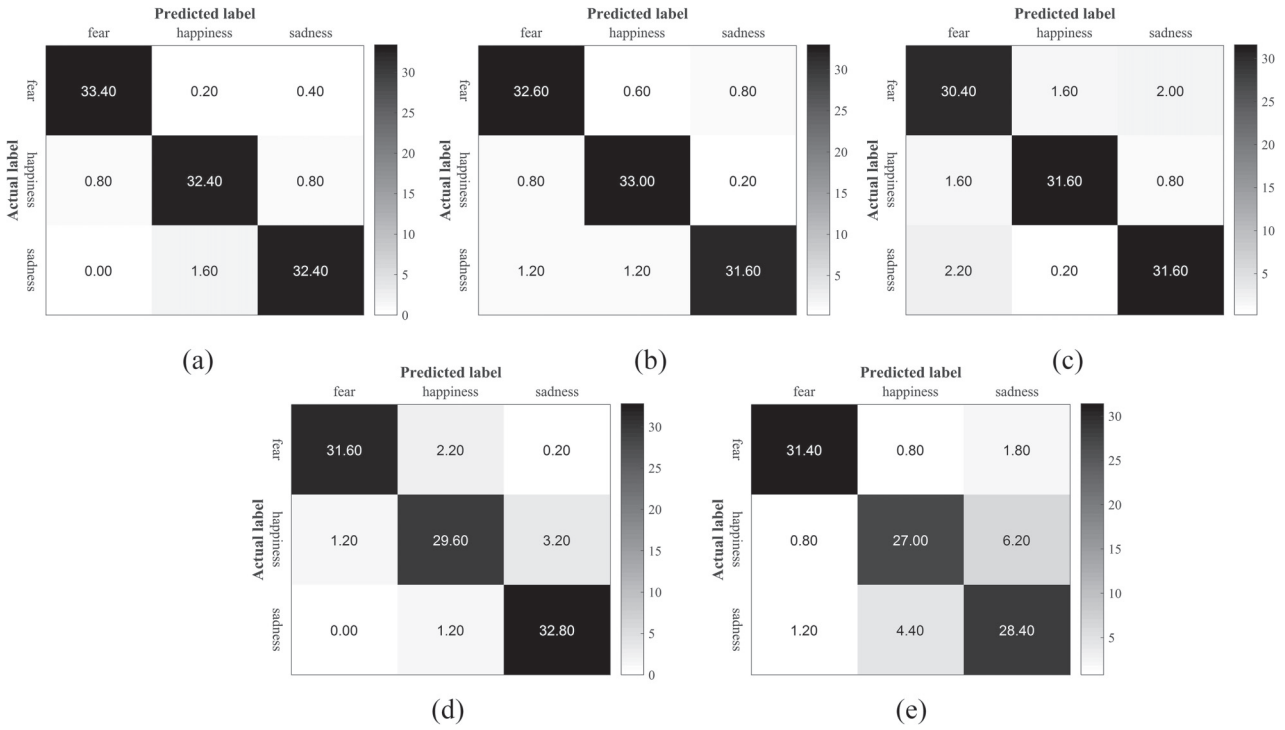

FIG. 3. The average confusion matrix across five folds from the RQA + CFCNN system: (a) subject 1, (b) subject 2, (c) subject 3, (d) subject 4, (e) subject 5.

TABLE II. Accuracy (%) and standard deviation (std.) results of RQA + CFCNN compared with DE + SVM and PSD + SVM systems.

	Accuracy (%) (mean ± std.)		
Subject	RQA + CFCNN	DE + SVM	PSD + SVM
1	96.27 ± 1.44	93.53 ± 3.07	94.90 ± 1.61
2	95.29 ± 2.27	92.94 ± 1.28	80.98 ± 1.49
3	91.76 ± 2.81	76.08 ± 3.64	68.63 ± 3.73
4	92.16 ± 1.75	81.76 ± 2.26	67.25 ± 3.83
5	85.69 ± 2.29	82.75 ± 2.03	51.18 ± 6.55
Average	92.24 ± 2.11	85.41 ± 2.46	72.59 ± 3.44
Inter-subject std.	3.71	6.78	14.64

we can find that all the subjects have a higher Kappa value than two baselines. On the average, RQA + CFCNN system obtains a Kappa value of 0.884, whereas the average Kappa values are 0.781 and 0.589 for PSD + SVM and DE + SVM, respectively. This shows that our method gives a remarkable increase of 13% with respect to DE + SVM and 50% improvement with respect to PSD + SVM in terms of average Kappa value. Correspondingly, our system can substantially facilitate the development of an outstanding HMI system which requires the machine possesses a strong identification capability and a satisfactory stability.

Besides, finding the critical frequency band which is highly associated with emotion process is also very useful for emotion recognition. Consequently, we remove the last convolutional layer of the CFCNN to compare the performance of each frequency band in distinguishing the three emotion states. We find that the entropy measure from gamma frequency band has a far more prominent emotion identification capability compared with the other four frequency bands, and its average accuracy of 90.51% is almost equivalent to the result obtained by total frequency bands (detailed results are shown in Table IV). From the perspective of Kappa value, average Kappa value from gamma frequency band quite approaches to that from total frequency bands, and the with-subject standard deviation is nearly the same for the two inputs. All these results indicate that emotional arousal is closely linked to gamma frequency band activities which is compatible with other studies reporting that high frequency band activities (higher than 30 Hz) are related to emotional process.[58–60]

TABLE III. Kappa value and standard deviation (std.) results of RQA + CFCNN system compared with DE + SVM and PSD + SVM systems.

	Kappa (mean ± std.)		
Subject	RQA + CFCNN	DE + SVM	PSD + SVM
1	0.944 ± 0.022	0.903 ± 0.041	0.924 ± 0.022
2	0.929 ± 0.034	0.894 ± 0.017	0.715 ± 0.020
3	0.876 ± 0.042	0.641 ± 0.049	0.529 ± 0.050
4	0.882 ± 0.026	0.726 ± 0.030	0.509 ± 0.051
5	0.785 ± 0.034	0.741 ± 0.027	0.268 ± 0.088
Average	0.884 ± 0.032	0.781 ± 0.033	0.589 ± 0.046
Inter-subject std.	0.056	0.102	0.220

TABLE IV. Accuracy (%) and Kappa value results [with standard deviation (std.)] of entropy feature from total frequency bands and gamma frequency band, respectively.

	Accuracy (%) (mean ± std.)		Kappa (mean ± std.)	
Subject	Total	Gamma	Total	Gamma
1	96.27 ± 1.44	95.69 ± 1.47	0.944 ± 0.022	0.935 ± 0.022
2	95.29 ± 2.27	95.10 ± 0.88	0.929 ± 0.034	0.926 ± 0.013
3	91.76 ± 2.81	88.04 ± 3.58	0.876 ± 0.042	0.821 ± 0.054
4	92.16 ± 1.75	90.20 ± 2.84	0.882 ± 0.026	0.853 ± 0.043
5	85.69 ± 2.29	83.53 ± 1.80	0.785 ± 0.034	0.753 ± 0.027
Average	92.24 ± 2.11	90.51 ± 2.11	0.884 ± 0.032	0.858 ± 0.032
Inter-subject std.	3.71	4.53	0.056	0.068

V. CONCLUSIONS

The design of a reliable and stable emotion classification system is one of the most primary requisites in the affective computing field which concentrates on implementing a vivid communication between humans and machines. We have designed an emotional arousal experiment to induce three specific emotions, namely, happiness, sadness, and fear and measure corresponding EEG signals of each subject. Then we offer a peculiar but vigorous RQA-based CFCNN recognition system for distinguishing the emotion states. Our results based on this application suggest that our method can get access to effectual emotion classification with high accuracy as well as good stability and clearly outperforms two traditional methods. In addition, the distinct performance of the gamma frequency band in classifying emotions shows a strong correlation between emotional process and gamma frequency band activities. As a further work, we will increase the number of subjects to build a larger dataset for studying the ability of our method in eliminating the inter-subjective differences. The excellent performance of the proposed RQA-based CFCNN recognition system gives itself a great prospect in the research of an affective HMI system and is quite likely to be generalized to the identification task of EEG signals in other areas.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61473203 and 61873181, and the Natural Science Foundation of Tianjin, China, under Grant No. 16JCYBJC18200.

APPENDIX: THE INFORMATION TRANSMISSION PROCESS OF CFCNN

The detailed information transmission process between two layers is as follows:

(1) For layer L_1

$$\sigma_{1,k,(m,n)} = \omega_{1,k,0} + \sum_{j=1}^{N_{ch}} I_{m,n+j-1} \cdot \omega_{1,k,j}, \tag{A1}$$

where $\omega_{1,k,0}$ is a bias and $\omega_{1,k,j}$ denotes a set of weights with $1 \le j \le N_{ch}$. Each map contains N_{ch} weights in layer L_1, and the convolution kernels are a series of spatial filters each of which has a size of $N_{ch} \times 1$. This layer concentrates on mining spatial information within each frequency band. Layers L_2 and L_3 follow the same rules except for the different numbers of filters and can be induced from L_1.

(2) For layer L_4

$$\sigma_{4,k,n} = \omega_{4,k,0} + \sum_{i=1}^{N_3} \sum_{j=1}^{N_f} x_{3,i,(j,n)} \omega_{4,(i,k)}, \tag{A2}$$

where $\omega_{4,k,0}$ is a bias, $\omega_{4,(i,k)}$ is a set of weights, and $1 \le j \le N_f$, $1 \le i \le N_3$ (N_3 is the number of filters in layer L_3). Each map contains $N_f \times N_3$ weights in layer L_4 and the size of filter is $1 \times N_f$. This layer is used to extract high level information underlying different frequency bands.

(3) For layer L_5

$$\sigma_{5,n} = \omega_{5,0,n} + \sum_{t=1}^{N_4} \sum_{l=1}^{N_{c4}} x_{4,t,l} \omega_{5,n}, \tag{A3}$$

where $\omega_{5,0,n}$ is a bias. L_4 consists of N_4 feature maps, each of which has N_{c4} neurons. L_5 is fully connected to each neuron of L_4.

(4) For layer L_6

$$\sigma_{6,n} = \omega_{6,0,n} + \sum_{q=1}^{N_5} x_{5,q} \omega_{6,n}, \tag{A4}$$

where $\omega_{6,0,n}$ is a bias. Each neuron of L_6 is connected to each neuron of L_5 which has N_5 neurons.

[1] R. Picard, *Affective Computing* (MIT Press, 2000).
[2] M. Pantic and L. J. M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1424–1445 (2000).
[3] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE Trans. Pattern Anal. Mach. Intell. **31**(1), 39–58 (2009).
[4] K. Mistry, L. Zhang, S. C. Neoh, C. P. Lim, and B. Fielding, "A micro-GA embedded PSO feature selection approach to intelligent facial emotion recognition," IEEE Trans. Cybern. **47**(6), 1496–1509 (2017).
[5] M. E. Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," Pattern Recognit. **44**(3), 572–587 (2011).
[6] B. Schuller, A. Batliner, S. Steidl, and D. Seppi, "Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge," Speech Commun. **53**(9–10), 1062–1087 (2011).
[7] H. Y. Meng and N. Bianchi-Berthouze, "Affective state level recognition in naturalistic facial and vocal expressions," IEEE Trans. Cybern. **44**(3), 315–328 (2014).
[8] H. Gunes and M. Piccardi, "Automatic temporal segment detection and affect recognition from face and body display," IEEE Trans. Syst. Man Cybern. B **39**(1), 64–84 (2009).
[9] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," IEEE Trans. Affect. Comput. **4**(1), 15–33 (2013).
[10] Y. Gao, N. Bianchi-Berthouze, and H. Meng, "What does touch tell us about emotions in touchscreen-based gameplay?," ACM Trans. Comput. Hum. Interact. **19**(4), 31 (2012).
[11] H. P. Martinez, Y. Bengio, and G. N. Yannakakis, "Learning deep physiological models of affect," IEEE Comput. Intell. Mag. **8**(2), 20–33 (2013).
[12] 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. **57**(7), 1798–1806 (2010).
[13] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis," IEEE Trans. Affect. Comput. **1**(2), 81–97 (2010).
[14] G. Chanel, C. Rebetez, M. Btrancourt, and T. Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," IEEE Trans. Syst. Man Cybern. A **41**(6), 1052–1063 (2011).
[15] H. Xu and K. N. Plataniotis, "Affect recognition using EEG signal," in *Proceedings of 14th IEEE International Workshop on Multimedia Signal Processing (MMSP)* (IEEE, 2012), pp. 299–304.
[16] S. K. Hadjidimitriou and L. J. Hadjileontiadis, "EEG-based classification of music appraisal responses using time-frequency analysis and familiarity ratings," IEEE Trans. Affect. Comput. **4**(2), 161–172 (2013).
[17] X. Chai, Q. S. Wang, Y. P. Zhao, Y. Q. Li, D. Liu, X. Liu, and O. Bai, "A fast, efficient domain adaptation technique for cross-domain electroencephalography (EEG)-based emotion recognition," Sensors **17**(5), 1014 (2017).
[18] G. L. Ahern and G. E. Schwartz, "Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis," Neuropsychologia **23**(6), 745–755 (1985).
[19] U. R. Acharya, O. Faust, N. Kannathal, T. J. Chua, and S. Laxminarayan, "Non-linear analysis of EEG signals at various sleep stages," Comput. Methods Programs Biomed. **80**(1), 37–45 (2005).
[20] K. C. Chua, V. Chandran, U. R. Acharya, and C. M. Lim, "Analysis of epileptic EEG signals using higher order spectra," J. Med. Eng. Technol. **33**(1), 42–50 (2009).
[21] Z. D. Mu, J. F. Hu, J. L. Min, and J. H. Yin, "Comparison of different entropies as features for person authentication based on EEG signals," IET Biom. **6**(6), 409–417 (2017).
[22] O. De Wel, M. Lavanga, A. C. Dorado, K. Jansen, A. Dereymaeker, G. Naulaers, and S. Van Huffel, "Complexity analysis of neonatal EEG using multiscale entropy: Applications in brain maturation and sleep stage classification," Entropy **19**(10), 516 (2017).
[23] Z. K. Gao, Q. Cai, Y. X. Yang, W. D. Dang, and S. S. Zhang, "Multiscale limited penetrable horizontal visibility graph for analyzing nonlinear time series," Sci. Rep. **6**, 35662 (2016).
[24] Z. K. Gao, Q. Cai, Y. X. Yang, N. Dong, and S. S. Zhang, "Visibility graph from adaptive optimal kernel time-frequency representation for classification of epileptic from EEG," Int. J. Neural Syst. **27**(4), 1750005 (2017).
[25] J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," Europhys. Lett. **4**(9), 973–977 (1987).
[26] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, "Recurrence plots for the analysis of complex systems," Phys. Rep. **438**(5–6), 237–329 (2007).
[27] L. L. Trulla, A. Giuliani, J. P. Zbilut, and C. L. Webber, "Recurrence quantification analysis of the logistic equation with transients," Phys. Lett. A **223**(4), 255–260 (1996).
[28] J. P. Zbilut and C. L. Webber, Jr., "Embeddings and delays as derived from quantification of recurrence plots," Phys. Lett. A **171**(3–4), 199–203 (1992).
[29] H. Yang, "Multiscale recurrence quantification analysis of spatial cardiac vector cardiogram signals," IEEE Trans. Biomed. Eng. **58**(2), 339–347 (2011).
[30] X. L. Li, G. X. Ouyang, X. Yao, and X. P. Guan, "Dynamical characteristics of pre-epileptic seizures in rats with recurrence quantification analysis," Phys. Lett. A **333**(1–2), 164–171 (2004).
[31] U. R. Acharya, V. S. Sree, S. Chattopadhyay, W. W. Yu, and A. P. C. Alvin, "Application of recurrence quantification analysis for the automated identification of epileptic EEG signals," Int. J. Neural Syst. **21**(3), 199–211 (2011).
[32] L. H. Song, D. S. Lee, and S. I. Kim, "Recurrence quantification analysis of sleep electroencephalogram in sleep apnea syndrome in humans," Neurosci. Lett. **366**(2), 148–153 (2004).
[33] U. R. Acharya, V. K. Sudarshan, H. Adeli, J. Santhosh, J. E. W. Koh, S. D. Puthankatti, and A. Adeli, "A novel depression diagnosis index using nonlinear features in EEG signals," Eur. Neurol. **74**(1–2), 79–83 (2015).

[34]N. Talebi, A. M. Nasrabadi, and T. Curran, "Investigation of changes in EEG complexity during memory retrieval: The effect of midazolam," Cogn. Neurodyn. **6**(6), 537–546 (2012).
[35]K. Becker, G. Schneider, M. Eder, A. Ranft, E. F. Kochs, W. Zieglgansberger, and H. U. Dodt, "Anaesthesia monitoring by recurrence quantification analysis of EEG data," PLoS One **5**(1), e8876 (2010).
[36]X. W. Wang, D. Nie, and B. L. Lu, "Emotional state classification from EEG data using machine learning approach," Neurocomputing **129**, 94–106 (2014).
[37]H. Shahabi and S. Moghimi, "Toward automatic detection of brain responses to emotional music through analysis of EEG effective connectivity," Comput. Hum. Behav. **58**, 231–239 (2016).
[38]G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science **313**(5786), 504–507 (2006).
[39]Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks* (MIT Press, 1995), pp. 255–258.
[40]S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput. **9**(8), 1735–1780 (1997).
[41]F. A. Gers, N. N. Schraudolph, and J. Schmidhuber, "Learning precise timing with LSTM recurrent networks," J. Mach. Learn. Res. **3**(1), 115–143 (2003).
[42]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. Ment. Dev. **7**(3), 162–175 (2015).
[43]X. Li, D. W. Song, P. Zhang, G. L. Yu, Y. X. Hou, and B. Hu, "Emotion recognition from multi-channel EEG data through convolutional recurrent neural network," in *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (IEEE COMPUTER SOC, 2016), pp. 352–359.
[44]G. L. Yu, X. Li, D. W. Song, X. Z. Zhao, P. Zhang, Y. X. Hou, and B. Hu, "Encoding physiological signals as images for affective state recognition using convolutional neural networks," in *Proceedings of 38th International Conference on IEEE Engineering Medicine and Biology Society (EMBC)* (IEEE, 2016), pp. 812–815.
[45]T. K. Marcha, S. C. Chapman, and R. O. Dendy, "Recurrence plot statistics and the effect of embedding," Physica D **200**, 171–184 (2005).
[46]B. Goswami, N. Marwan, G. Feulner, and J. Kurths, "How do global temperature drivers influence each other? A network perspective using recurrences," Eur. Phys. J. Spec. Top. **222**, 861–873 (2013).
[47]E. J. Ngamga, S. Bialonski, N. Marwan, J. Kurths, C. Geier, and K. Lehnertz, "Evaluation of selected recurrence measures in discriminating pre-ictal and inter-ictal periods from epileptic EEG data," Phys. Lett. A **380**(16), 1419–1425 (2016).
[48]N. Marwan, "How to avoid potential pitfalls in recurrence plot based data analysis," Int. J. Bifurcat. Chaos **21**(4), 1003–1017 (2011).
[49]S. Schinkel, O. Dimigen, and N. Marwan, "Selection of recurrence threshold for signal detection," Eur. Phys. J. Spec. Top. **164**, 45–53 (2008).
[50]V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of 27th International Conference on Machine Learning* (Omnipress, USA, 2010), pp. 807–814.
[51]F. Chollet, see https://github.com/fchollet/keras for "Keras: Deep Learning Library for Theano and TensorFlow" (2015).
[52]S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Proceedings of Machine Learning Research (PMLR) **37**, 448–456 (2015).
[53]X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," Proceedings of Machine Learning Research (PMLR) **9**, 249–256 (2010).
[54]L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of 19th International Conference on Computational Statistics* (Physica-Verlag, Springer, 2010), pp. 177–186.
[55]D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature **323**(6088), 533–536 (1986).
[56]J. Cohen, "A coefficient of agreement for nominal scales," Educ. Psychol. Meas. **20**(1), 37–46 (1960).
[57]Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," J. Neural Eng. **14**, 016003 (2017).
[58]S. Aydin, S. Demirtas, K. Ates, and M. A. Tunga, "Emotion recognition with eigen features of frequency band activities embedded in induced brain oscillations mediated by affective pictures," Int. J. Neural Syst. **26**(3), 1650013 (2016).
[59]M. M. Muller, A. Keil, T. Gruber, and T. Elbert, "Processing of affective pictures modulates right hemispheric gamma band EEG activity," Clin. Neurophysiol. **110**(11), 1913–1920 (1999).
[60]M. A. Kisley and Z. M. Cornwell, "Gamma and beta neural activity evoked during a sensory gating paradigm: Effects of auditory, somatosensory and cross-modal stimulation," Clin. Neurophysiol. **117**(11), 2549–2563 (2006).