



## Locally robust EEG feature selection for individual-independent emotion recognition

Zhong Yin <sup>a,b,\*</sup>, Lei Liu <sup>c</sup>, Jianing Chen <sup>b</sup>, Boxi Zhao <sup>b</sup>, Yongxiong Wang <sup>a,b,\*</sup>

<sup>a</sup> Engineering Research Center of Optical Instrument and System, Ministry of Education, Shanghai Key Lab of Modern Optical System, University of Shanghai for Science and Technology, Jungong Road 516, Yangpu District, Shanghai 200093, PR China

<sup>b</sup> School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Jungong Road 516, Yangpu District, Shanghai 200093, PR China

<sup>c</sup> School of Management, University of Shanghai for Science and Technology, Jungong Road 516, Yangpu District, Shanghai 200093, PR China



### ARTICLE INFO

#### Article history:

Received 17 September 2019

Revised 30 May 2020

Accepted 14 July 2020

Available online 25 July 2020

#### Keywords:

Emotion recognition

Affective computing

Physiological signals

Machine learning

Feature selection

### ABSTRACT

Brain computer interface (BCI) systems can decode brain affective activities into interpretable features and facilitate emotional human-computer interaction. However, individual differences of neurophysiological responses from BCI subjects constitute a stumbling block in individual-independent emotion recognition. In this study, we propose a new locally-robust feature selection (LRFS) method to determine generalizable features of electroencephalography (EEG) within several subsets of accessible subjects. In the LRFS framework, extracted EEG features are first modeled with probability densities. By evaluating the similarity of all density functions between each two subjects, inter-individual consistency of the EEG features is described. The derived consistency determines locally-robust EEG features, wherein importance of each feature is examined according to margin loss between emotions. To fuse selected features from multiple subsets of subjects, we employ ensemble learning principle and build an emotion classifier committee. Based on public DEAP and MAHNOB-HCI databases, individual-independent classification accuracy of the LRFS-based classifier is achieved by 0.65–0.68 (DEAP) and 0.67–0.70 (MAHNOB-HCI) for arousal and valence domains, respectively. Competitiveness of the LRFS has been validated when compared with several existing feature selection methods and emotion recognition systems.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

Human beings communicate with each other by using different forms of speeches, gestures, and expressions. These natural ways of information transferring convey both messages and verbal or non-verbal emotional cues of humans. Understanding of emotional intelligence can facilitate agents of human-computer interaction (HCI) to execute proper actions in a human-centered manner (Calvo & D'Mello, 2010). Emotion recognition methods of HCI systems based on speech signals (Yogesh et al., 2017), facial expressions (Jain, Shamsolmoali, & Sehdev, 2019), texts (Quan & Ren, 2016) and neurophysiological signals (Hassan et al., 2019) are documented and applied in various human centered working

environment (Lopez-de-Ipiña et al., 2013; Mano et al., 2019; Torres, Wei, Hua, & Chen, 2019; Yin et al., 2019).

From a perspective of cognitive neuroscience, emotions are known as psycho-physiological processes generated by neural mechanisms that regulate perceptions of an environment and detection of needs and goals of organisms (Panksepp, Lane, Solms, & Smith, 2017). Brain-computer interfaces (BCIs) made it possible by building cortical communicating channels that are independent of pathways from central nervous system to muscles (Alazraia, Alwanni, & Daoud, 2019). A common neuroimaging modality in BCI is electroencephalography (EEG). Cortical, electrical activity is recorded via the EEG with an electrode array positioned on scalp.

An EEG-BCI system for recognizing emotions usually comprises three components: EEG filtering/preprocessing, neurophysiological feature extraction/selection, and emotion classification. The second component is of particular importance since finding informative EEG features can improve inter-emotion discriminating capability with a limited dimensionality of feature space. Previous works demonstrated that power spectral densities, higher order crossing spectrums, and temporal or temporal-frequential parameters of

\* Corresponding authors at: Engineering Research Center of Optical Instrument and System, Ministry of Education, Shanghai Key Lab of Modern Optical System, University of Shanghai for Science and Technology, Jungong Road 516, Yangpu District, Shanghai 200093, PR China.

E-mail addresses: [yinzhong@usst.edu.cn](mailto:yinzhong@usst.edu.cn) (Z. Yin), [liulei@usst.edu.cn](mailto:liulei@usst.edu.cn) (L. Liu), [1712120109@st.usst.edu.cn](mailto:1712120109@st.usst.edu.cn) (J. Chen), [1712120129@st.usst.edu.cn](mailto:1712120129@st.usst.edu.cn) (B. Zhao), [wyxiong@usst.edu.cn](mailto:wyxiong@usst.edu.cn) (Y. Wang).

the EEG signals are sensitive to emotional variations (Liang, Shigeyuki, & Shin, 2019; Taran & Varun, 2019). Feature selection methods such as dependence guided unsupervised feature selection (Guo & Zhu, 2018), infinite latent feature selection (Roffo, Melzi, & Cristani, 2015) and recursive feature elimination (Raza et al., 2019) can be applied to discover useful EEG features.

A critical obstacle to determining salient EEG features is that both affective cognition processes and EEG statistics are individually dependent. Individuals may differ in their behavioral and neurophysiological responses about controllability of emotions (King & dela Rosa, 2019; Gurera & Isaacowitz, 2019). The EEG power features in the same frequency band may also show a varied distribution for different BCI users involved in the same cognitive environment (Richhariya, Tanveer, & Rashid, 2020). These facts lead to difficulty of discovering generic EEG features with a robust capability to distinguish emotions.

Motivated by the previous works, in this study we present a novel locally-robust feature selection (LRFS) method for promoting EEG-based emotion recognition systems with capabilities of determining generalizable EEG indicators among a subset of BCI users. The contribution of the LRFS lies in three folds: (1) The density estimation and recursive feature elimination (RFE) are combined to quantitatively evaluate the average individual difference of each EEG feature. (2) The locally-robust features are defined based on the highest similarity among a “local scope” of individuals evoked by the same emotional state. We developed three efficient algorithms that ensure the correctness of such feature selection procedure. (3) To better leverage the heterogeneous feature rankings from multiple local scopes, ensemble learning is employed to build a classifier committee to elicit emotion predictions.

The basic idea of the LRFS-based emotion recognition system consists of four stages. In Stage 1, kernel smoothing density estimation is used to discover probability density functions (PDFs) hidden in the EEG features of each individual. Inter-individual similarity of the PDFs is measured to generate a ranking matrix for each EEG feature in Stage 2. In Stage 3, we implement RFE feature ranking to select locally-robust features from subsets of subjects based on the ranking matrices. Finally, in Stage 4 emotion classifiers trained from the selected EEG features of multiple subsets are properly integrated to form a classifier ensemble and to estimate temporal emotional states.

The paper is structured as follows. Previous works are reviewed in Section 2. In Section 3, we describe materials and methods used for EEG feature selection and emotion recognition. Classification results are presented in Section 4. In Section 5, we discuss the findings in detail. Contribution and limitation of the present study is concluded in Section 6.

## 2. Related works

Human brain reflects emotional states in a multifaceted nature associated with multiple cortical and subcortical structures (LeDoux, 2012). For example, ventromedial prefrontal cortex may trigger affective reactions in response to cognitive changes (Critchley et al., 2003). Dorsomedial prefrontal cortex appears to be relevant to emotion generation (Kober et al., 2008). These observations suggest that emotions can be analyzed by acquiring useful information directly from the brain besides facial expressions and gestures. The cortical activities can be interpreted by EEG signals. However, frequencies of the EEG vary with subject functional states of workload, fatigue, vigilance, and situation awareness (Borghini et al., 2014). An EEG-based BCI system has a capability to decode brain affective activities into interpretable features and provides an estimation of human emotional states (Atkinson & Daniel, 2016). In particular, features of EEG signals reflect inner

cognitive states and cannot be influenced intentionally. By requiring a portable, wireless electrode device, emotions can be continuously predicted at a high temporal resolution (Iacoviello, Petracca, Spezialetti, & Placidi, 2015).

The DEAP and MAHNOB-HCI databases are built based on the EEG-based BCI system and were leveraged to validate the proposed emotion classifier. The DEAP database was originally built by (Koelstra et al., 2012) with multimodal signals of human users including facial expressions, EEG, and physiological data from the peripheral nervous system. In (Koelstra et al., 2012; Liu & Sourina, 2012; Yoon & Chung, 2013; Wang & Shang, 2013; Chen et al., 2015; Atkinson & Daniel, 2016), various emotion recognition systems on arousal, valence, and dominance dimensions based on the EEG modality of DEAP were reported with accuracies from 0.51 to 0.76. Two to four emotional states have been classified based on naive Bayesian (NB) classifier, support vector machine (SVM), modified NB, deep belief network (DBN), and C4.5 decision tree. Feature selection methods founded on correlation analysis, gain ratio, and minimum redundancy maximum relevance (mRMR) were applied. Specifically, these emotion classifiers were trained and tested on the EEG data from each individual. In (Nakisa et al., 2018) and our previous work (Yin, Liu, Liu, Zhang, & Wang, 2017), individual-generic scheme was employed, wherein evolutionary computation based probabilistic neural network (PNN) and dynamical RFE (D-RFE) with least square support vector machine (LSSVM) (Suykens & Vandewalle, 1999) were developed with accuracy of 0.71–0.79 (two emotion classes) and 0.67 (four emotion classes), respectively. It is important to note that the training and testing datasets of the individual-generic scheme also shares the EEG features from the same individual.

Multimodal database MAHNOB-HCI was originally built by Soleymani, Lichtenauer, Pun, and Pantic (2012). Peripheral physiological signals, EEG, facial and body videos, eye gaze and audio data were collected. In Soleymani, Lichtenauer, Pun, and Pantic (2012), Huang et al. (2016), EEG-based emotion recognition accuracy of 0.52 and 0.61 was reported with an individual-independent scheme for two dimensions. Training and testing datasets were constructed from different subjects and a leave-one-subject-out paradigm was used to validate classifier performance. In these two works, EEG features were selected by ANOVA test or fused by kernel locality preserving projection (KLPP) technique before transferring to SVM classifiers. In Zhu, Wang, and Ji (2014), an individual-generic classification scheme was enacted with canonical correlation analysis (CCA) for feature fusion. In particular, (Nakisa et al., 2018) achieves emotion classification accuracy of 0.97 with the same training scheme. In Pereira, Gomes, Veloso, and Mota (2018), individual-specific accuracy was 0.59 and 0.76 based on the SVM classifier for arousal and valence dimensions, respectively.

From the literature review, we found that temporal/frequency statistics, power on different bands, fractal dimension, and high-order crossings have been extracted as EEG features to indicate emotion variations. Manifold learning (e.g., CCA and KLPP), statistical learning (e.g., D-RFE and SVM), and deep learning (e.g., DBN) approaches were employed for feature selection, fusion and classification. Specifically, schemes for training emotion classifiers can be summarized to individual-specific, individual-generic, and individual-independent paradigms. Three modeling schemes based on individual-specific, individual-generic, and individual-independent paradigms possess advantages under different application environments. The individual-specific paradigm could train emotion classifiers with perfect performance when physiological data from a single subject is adequate. The reason is that cortical feature patterns from the same person are stable to fit statistical learning principles. The individual-generic paradigm fuses the data from multiple subjects to build classifiers. Distinct patterns of

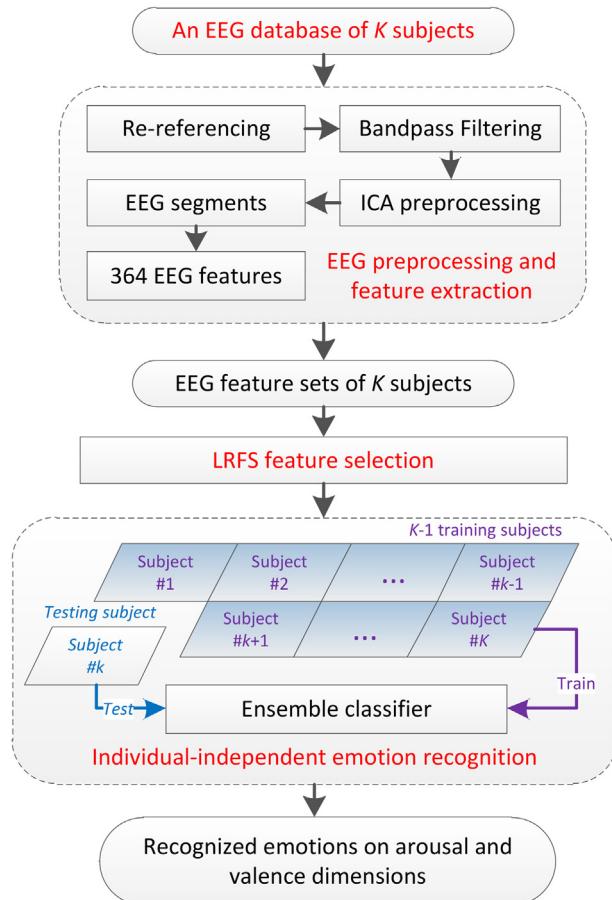
feature distribution are filtered when the size of the subject pool is large enough. On the other hand, the individual-independent paradigm could transfer knowledge of cortical activations between several training subjects and a specific testing subject. Its advantage is the case that training and testing data for machine learning can be from completely different BCI users. According to our previous works on the issues of individual-specific emotion recognition (Yin, Zhao, Wang, Yang, & Zhang, 2017) and EEG feature selection (Yin & Zhang, 2014), robust EEG features that reflect cognitive states may exist among a small group of carefully-selected BCI users. To this end, we develop the LRFS for EEG feature selection and attempt to tackle concept drifts across training and testing neurophysiological data from different individuals based on the last scheme.

### 3. Materials and method

In this section, employed EEG databases were first introduced. Then, we provide procedures for data preprocessing and feature extraction. It also describes the details of the LRFS method for EEG feature selection and emotion recognition under an individual-independent paradigm. The pattern classification framework is illustrated in Fig. 1.

#### 3.1. Descriptions of emotional databases

The DEAP database was aimed to investigate the spontaneous emotions of human users when they were watching musical videos. Forty videos for inducing subject's affective reactions with



**Fig. 1.** Emotion classification framework based on the LRFS method. The full name of ICA shown in the figure is independent component analysis.

the clearest emotional responses have been identified by an online subjective assessment experiment. In total, 32 subjects (50% female) participated and each subject accomplished 40 trials of experiments, wherein each trial corresponds to a selected video clip with a duration of a minute. Physiological signals and face videos have been simultaneously acquired for all trials. At the end of a trial, arousal, valence, dominance, liking and familiarity scales were rated by the subject with a range of 1–9.

The goal of the MAHNOB-HCI database is for promoting emotion recognition and implicit tagging studies. It was built by a similar experimental setting with the DEAP. Twenty movie clips were selected from online annotated video sets. Twenty-seven subjects including 11 males and 16 females participated the data acquisition experiment. Each subject performed 20 trials of the experimental task. The EEG data of 24 subjects was available since the data of three subjects was incomplete. Time duration of a trial was approximately two and half minutes. During the experiments, physiological data, audio, eye gaze, face and body videos were simultaneously recorded. At the end of each trial, emotional keywords, arousal, valence, dominance and predictability scales were rated by the subjects within a range of 1–9. In the rest of the study, we only adopted the EEG modality of the DEAP and MAHNOB-HCI databases to validate our method.

#### 3.2. EEG data preprocessing and feature extraction

Detail information of each step of EEG preprocessing is shown in Table 1. For two databases, the same 32 channels with a sampling frequency of 256 Hz were downsampled to 128 Hz. The average amplitude of the EEG signals recorded in the MAHNOB-HCI database is approximately  $-1.2 \times 10^6$  uV while that of the DEAP is in normal condition. To compute correct power features, re-referencing step is only applied to the MAHNOB-HCI. By analyzing the ratio between different frequency components and fullband power, we found eye movements (0–4 Hz) and the power within 45–60 Hz are severely distributed in the EEG of the DEAP. A band-pass filter (4–45 Hz) is combined with the independent component analysis (ICA) to eliminate such ocular artifacts and high-frequency noise. For the MAHNOB-HCI, potential artifacts of respiration and body movements within 0–3 Hz among most of the subjects are observed by examining the power spectral plots. Thus, only a high-pass filer ( $\geq 3$  Hz) is implemented. Therefore, different types of preprocessing steps are adapted for the two databases based on the quality of the EEG signals. The length of the preprocessed EEG

**Table 1**  
Summary of EEG signal preprocessing steps.

Preprocessing steps	Annotations	Applied databases
EEG channel settings	Thirty two channels: Fp1, AF3, F3, F7, FC5, FC1, C3, T7, CP5, CP1, P3, P7, PO3, O1, Oz, Pz, Fp2, AF4, Fz, F4, F8, FC6, FC2, Cz, C4, T8, CP6, CP2, P4, P8, PO4, and O2 according to the 10–20 system.	Both
Downsampling	Sampling frequency of 128 Hz.	Both
Re-referencing	Subtract the average amplitude of all 32 channels.	MAHNOB-HCI
Bandpass filtering	Third-order Butterworth filter with cutoff frequency of 4–45 Hz.	DEAP
Highpass filtering	Seventh-order Butterworth filter with cutoff frequency of 3 Hz.	MAHNOB-HCI
ICA artifact removal	The independent component corresponding to the highest ratio between the energy of 30–45 Hz and the full band was set to zero in the remix stage of the FastICA algorithm.	DEAP
Number of EEG segments	Three non-overlapped segments of a trial.	Both

signal of a trial is 60 s and approximately two and a half minutes for the DEAP and the MAHNOB-HCI databases, respectively. Finally, we evenly divided a trial EEG data into three nonoverlaped segments.

The reason that only the EEG modality was picked is in three folds. (1) The proposed feature selection procedure did not leverage previous studies related to individual differences on facial expression, eye-gaze, and audio data. The current module of the LRFS for evaluating feature similarity could be ineffective for these data modalities. (2) Electrode locations of 32 EEG channels in the two databases are exactly the same. It facilitates a relative fair validation for the stability of the feature ranking procedure. (3) To improve training efficiency of machine learning classifiers, data augmentation is implemented by creating three nonoverlaped segments. However, such procedure may generate incorrect features for low-frequent peripheral signals.

For each EEG segment, we extracted 364 features and summarize details of each feature subset in Table 2. Power features of classical EEG bands and power differences between right and left scalps were computed based on fast Fourier transform. Power ratios of specific channels and frequency bands were also calculated since these features are sensitive to variation of stress and negative feelings (Menon et al., 2000). Alpha suppression in central and parietal regions is potentially related to high alertness (Christensen, Estep, Wilson, & Russell, 2012). An increase of the frontal theta power may be caused by a higher stress level (Gundel, & Wilson, 1992). Sufficient temporal statistics and complexity variables were also incorporated. In total, 3840 (120 for each subject) and 1440 (60 for each subject) data points were extracted from the DEAP and MAHNOB-HCI, respectively. All EEG features were standardized to zero mean and unit variance.

In Fig. 2, we compute Pearson correlation coefficients between the time course of all EEG features and the subjective ratings on the arousal scale. The results of three representative subjects in the DEAP database are shown. The feature names are labeled along with their indices. The ranges of #1–#128, #129–#184, #185–#204, #205–#300, and #301–#364 represent the power spectrums, power difference, power ratios between different channels, temporal statistics, and complex indicators, respectively. From the figure, the coefficients corresponding to the power spectral and temporal statistics of subjects #1 and #3 are quite similar.

**Table 2**  
Descriptions of the extracted EEG features.

Feature name	Feature description	Dimensionality of features
Theta, alpha, beta, and gamma power	The average power of the frequency band in 4–8 Hz, 9–12 Hz, 13–30 Hz, and 31–45 Hz of each EEG channel, respectively.	128
Power difference	Frontal scalp: $p_{Fp2} - p_{Fp1}$ , $p_{AF4} - p_{AF3}$ , $p_{F4} - p_{F3}$ , $p_{F8} - p_{F7}$ , $p_{FC6} - p_{FC5}$ , and $p_{FC2} - p_{FC1}$ ; Central scalp: $p_{C4} - p_{C3}$ , $p_{T8} - p_{T7}$ , $p_{CP6} - p_{CP5}$ , and $p_{CP2} - p_{CP1}$ ; Parietal scalp: $p_{P4} - p_{P3}$ , $p_{P8} - p_{P7}$ , $p_{P04} - p_{P03}$ , and $p_{O2} - p_{O1}$ . $p_{F2}/(p_{AF3} + p_{AF4})$ , $p_{Cz}/p_{Fz}$ , $p_{Pz}/p_{Cz}$ , $p_{Oz}/p_{Pz}$ , $p_{AFz.theta}/p_{Pz.alpha}$ , $p_{Afz.theta}/p_{Cz.alpha}$ , $p_{Cz.theta}/p_{Pz.alpha}$ , and $p_{Cz.theta}/p_{Oz.alpha}$	56
Power ratio	Variance, kurtosis, and skewness of each channel	20
Temporal statistics	Zero crossing rate and Shannon entropy of each channel	96
Complexity indicators		64

Note: The term  $p_A$  denotes a four-dimensional feature subset of channel A of theta, alpha, beta and gamma bands. The term  $p_{A,B}$  represents an EEG power feature of channel A on frequency band B.

However, such values are completely different from that of subject #2. Specifically, power spectrums on four bands are all positively correlated with the subjective ratings for subject #2 while negative correlations are observed for subjects #1 and #3. The observation suggests a difficulty pertaining to emotion recognition across subjects #2 and #3 or across #2 and #1. Class labels of EEG data points are set as  $-1$  when the subjective rating scales are within the range of [1, 5]. It indicates that low level of arousal or valence states are stimulated. When rating scales are within [5, 9], the class labels are set as 1 showing high arousal or valence states.

### 3.3. Kernel density estimation and RFE

In this section, we briefly review kernel density estimation (KDE) (Gramacki, 2019) and the RFE approaches. These two techniques are the basis for developing the proposed method.

We utilize nonparametric KDE to reveal probability density behind EEG feature samples. For a given EEG feature variable  $x$  of class  $C$ , the class-conditional probability density  $\tilde{p}(x|C)$  can be estimated via its histogram,

$$\tilde{p}(x|C) = \frac{1}{2d} \lim_{d \rightarrow 0} \frac{N_{C,x_i \in [x-d,x+d]}}{N_C}. \quad (1)$$

In Eq. (1),  $x_i$  and  $N_C$  denote a feature sample with an index  $i$  and number of all samples of class  $C$ , respectively. The term  $N_{C,x_i \in [x-d,x+d]}$  represents amount of samples falling in a closed interval  $[x-d, x+d]$ .

Given a specific value of  $d$ , we get discontinuous  $\tilde{p}(x|C)$  which can be smoothed by a kernel function  $K(\cdot)$ . A PDF is then estimated by rewritten Eq. (1) using  $K(\cdot)$ ,

$$\tilde{p}(x|C) = \frac{1}{N_C \cdot d} \sum_{i=1}^{N_C} K\left(\frac{|x - x_i|}{d}\right). \quad (2)$$

The optimal window parameter  $d^*$  is determined by minimizing a mean integrated squared error,

$$d^* = \underset{d}{\operatorname{argmin}} E\left(\int (\tilde{p}(x|C) - p(x|C))^2 dx\right). \quad (3)$$

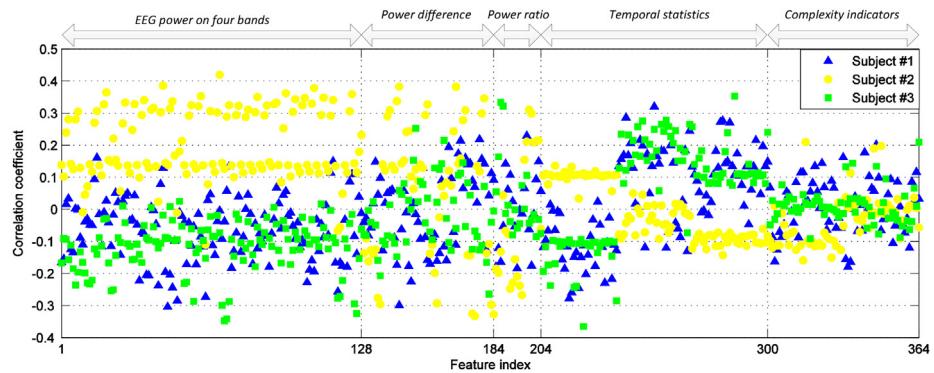
The term  $p(x|C)$  denotes true distribution of an EEG feature of emotion class  $C$ . For practical applications, we could assume  $p(x|C)$  as a Gaussian probability density function, i.e.  $p(x|C) = 1/\sqrt{2\pi\sigma_C} \cdot \exp(-\|x - u_C\|_2^2/\sigma_C^2)$ . The parameters,  $u_C$  and  $\sigma_C^2$ , are computed based on sample mean and variance of feature values belong to class  $C$ , respectively. By minimizing the error between the estimated density  $\tilde{p}(x|C)$  and the predefined  $p(x|C)$ , the proper window width is determined based on Eq. (3).

The RFE technique defines a margin  $\phi$  by a SVM classifier,

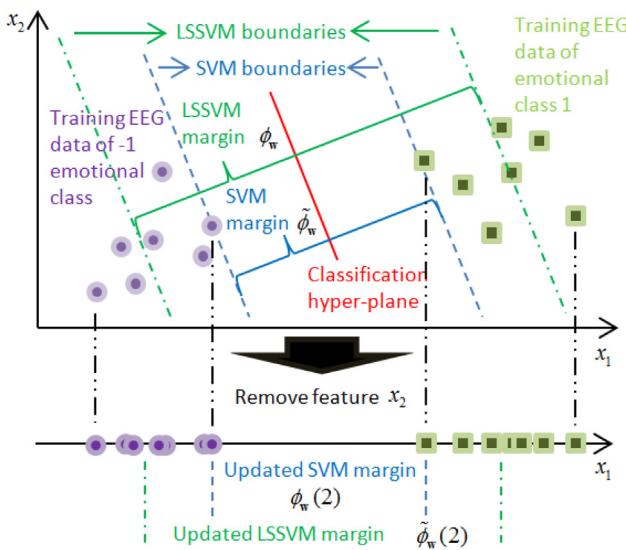
$$\begin{aligned} \min_{\mathbf{w}} \psi(\mathbf{w}) &= \frac{1}{\phi(\mathbf{w})} = \frac{1}{2} \mathbf{w}^T \mathbf{w} \\ \text{s.t. } C_i \cdot (\mathbf{w} \cdot \mathbf{x}_i + b) &\geq 1, \quad i = 1, 2, \dots, N. \end{aligned} \quad (4)$$

In Eq. (4),  $\mathbf{x}_i \in R^D$  represents an EEG data point in a  $D$ -dimensional space with emotion class of  $C_i \in \{-1, +1\}$ ,  $N$  denotes number of training instances,  $\mathbf{w}$  is the norm vector of classification hyperplane  $\mathbf{w} \cdot \mathbf{x} + b = 0$  and  $\psi$  denotes the reciprocal of the margin  $\phi$ . Based on Eq. (4), an emotion classifier  $\mathbf{w} \cdot \mathbf{x} + b = 0$  with the largest  $\phi$  is trained, wherein  $\phi$  is the distance between two class boundaries of  $\mathbf{w} \cdot \mathbf{x} + b = 1$  and  $\mathbf{w} \cdot \mathbf{x} + b = -1$ .

We employ LSSVM, a variant of SVM with fast training speed, to generate the margin. Its optimization constraints has changed to  $C_i \cdot (\mathbf{w} \cdot \mathbf{x}_i + b) = 1 - \xi_i$  with  $\xi_i$  denoting a slack variable to allow nonzero training error (Suykens, & Vandewalle, 1999). We illustrate differences between LSSVM and SVM in Fig. 3. When removing an EEG feature from original data space, LSSVM classifier is



**Fig. 2.** Pearson correlation coefficient between time courses of each EEG feature and subjective rating scales. Three subjects' coefficients (#1, #2, and #3) of the DEAP database are visualized.



**Fig. 3.** Relationship between margin loss and feature importance of the SVM and LSSVM.

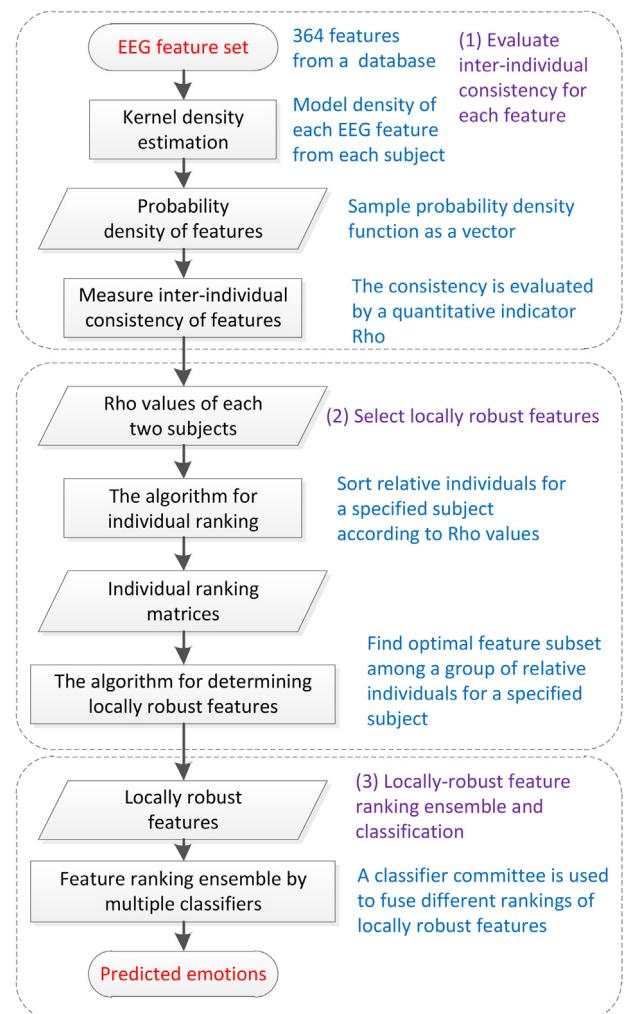
updated in a  $D-1$  dimensional feature subspace. Importance of this removed feature is quantified by loss of the margin between the original and updated hyper-planes,

$$\Delta\phi_w(j) = |\phi_w - \phi_w(j)| \quad (5)$$

As shown in Fig. 3, the value of  $\Delta\phi_w(2) = |\phi_w - \phi_w(2)|$  is quite small with  $x_2$  removed. It implies that  $x_1$  makes a greater contribution to distinguishing two emotional classes than  $x_2$  does. Since higher feature importance corresponds to larger margin loss, feature importance from high to low can be computed depending on Eq. (5) from  $q = 1$  to  $q = D$ . Based on recursive computation of  $\Delta\phi_w$  by retraining  $w$  with the highest-ranked feature removed in each iteration, the ranking of the feature is determined by the removal order.

### 3.4. Locally robust feature selection

In Fig. 4, we illustrate three main steps of the LRFS method. The extracted EEG features were used to estimate the PDFs via the KDE approach. According to the PDFs, the inter-individual consistency is evaluated to generate individual ranking matrices of all EEG features. The derived consistency determines the locally-robust EEG features, wherein the importance of each feature is examined



**Fig. 4.** Locally robust feature selection based emotion recognition framework.

according to the margin loss between emotions. Finally, a classifier ensemble was used to fuse the predicted emotions from multiple feature rankings. The details of these steps are described as follows.

#### 3.4.1. Inter-individual consistency evaluation

Given the  $j^{\text{th}}$  EEG feature  $x^{(j)}$ , its PDF of each emotional class  $C$  is estimated and sampled to a vector,

$$\hat{\mathbf{p}}_{(k,j,C)}[s] = \frac{1}{N_{C,k}} \sum_{i=1}^{N_{C,k}} K_G(sT - x_{i,k}^{(j)}), \quad j = 1, 2, \dots, D. \quad (6)$$

In Eq. (6), the term  $T$  and  $N_{C,k}$  denote the sampling period of the PDF and the number of instances belonging to class  $C$  for subject  $\#k$ . The term  $K_G(\cdot)$  represents the Gaussian kernel function,

$$K_G(x^{(j)} - x_{i,k}^{(j)}) = \exp\left(-\|x^{(j)} - x_{i,k}^{(j)}\|_2^2 / (2d^2)\right). \quad (7)$$

The window parameter  $d$  is optimized according to Eq. (3).

To evaluate inter-individual consistency of the sampled PDF, we proposed three quantitative indicators as follows,

$$\rho_1(\hat{\mathbf{p}}_{(k_1,j,C)}, \hat{\mathbf{p}}_{(k_2,j,C)}) = \frac{1}{l} \sum_l \frac{\sum_{s=1}^{N_s} \prod_{k=k_1, k_2} (\hat{p}_{(k,j,C)}[s] - \bar{p}_{(k,j,C)})}{\prod_{k=k_1, k_2} \sqrt{\sum_{s=1}^{N_s} (\hat{p}_{(k,j,C)}[s] - \bar{p}_{(k,j,C)})^2}}. \quad (8)$$

$$\rho_2 = \frac{\rho_1(\hat{\mathbf{p}}_{(k_1,j,C)}, \hat{\mathbf{p}}_{(k_2,j,C)})}{\rho_1(\hat{\mathbf{p}}_{(k_1,j,C=-1)}, \hat{\mathbf{p}}_{(k_2,j,C=-1)})}, \quad \rho_3 = \frac{\rho_1(\hat{\mathbf{p}}_{(k_1,j,C=-1)}, \hat{\mathbf{p}}_{(k_2,j,C=-1)})}{\rho_1(\hat{\mathbf{p}}_{(k_1,j,C=-1)}, \hat{\mathbf{p}}_{(k_2,j,C=1)})}. \quad (9)$$

In Eq. (8), average correlation between the sampled PDFs of subject  $k_1$  and  $k_2$  for all emotional classes was measured by  $\rho_1$ . The term  $N_s$  denotes number of the points sampled from the PDF. In Eq. (9), the consistency between PDFs across  $C = -1$  and  $C = 1$  classes is also considered in  $\rho_2$ . For indicator  $\rho_3$ , only the inter-individual consistency of a single class is computed to reduce computational cost. Larger values of  $\rho_1$ ,  $\rho_2$ , and  $\rho_3$  correspond to higher inter-individual consistency and lower interclass consistency of the EEG feature  $x^{(j)}$ .

In Fig. 5, we present an algorithm for evaluating the inter-individual consistency of all  $D$  features of  $K$  subjects. In lines 1 to 7, the weight matrices  $\mathbf{W}_j$  for quantifying the consistency are computed according to Eqs. (8) and (9). Note that three consistency indicators,  $\rho_1$ ,  $\rho_2$ , and  $\rho_3$ , are presented by  $\rho$  for simplicity in line

	<b>Function:</b>	<b>Eval_interind</b> ( $\mathbf{X}, C, K, T$ )
	<b>Input:</b>	EEG features of all subjects $\mathbf{X}$ Emotional classes $C$ Number of subjects $K$ Sampling period of the PDF $T$
1	<b>Codes:</b>	<b>for</b> $j = 1:D$
2		<b>for</b> $k = 1:K$
3		<b>for</b> $s = 1:N_s$
4		$\hat{\mathbf{p}}_{(k,j,C)}[s] \leftarrow 1/N_{C,k} \cdot \sum_{i=1}^{N_{C,k}} K_G(sT - x_{i,k}^{(j)})$
5		<b>for</b> $k' = 1:K$
6		<b>for</b> $k'' = 1:K$
7		$\mathbf{W}_j(k,k') \leftarrow \rho(\hat{\mathbf{p}}_{(k,j,C)}, \hat{\mathbf{p}}_{(k'',j,C)})$
8		Normalize $\mathbf{W}_j$
9		<b>for</b> $j = 1:D$
10		$\mathbf{R}_j \leftarrow \text{Sort}(\mathbf{W}_j)$
11		<b>for</b> $k = 1:K$
12		$\mathbf{R}_j(k,:) \leftarrow \mathbf{R}_j(k,:) \setminus \mathbf{R}_j(\mathbf{R}_j(k,:)) == k$
13		$\mathbf{R}_j(k,:) \leftarrow [k, \mathbf{R}_j(k,:)]$
14		<b>Return</b> $\mathbf{R}_j, \mathbf{W}_j$
	<b>Output:</b>	Individual ranking matrices $\mathbf{R}_j$ Weight of inter-individual consistency $\mathbf{W}_j$

Fig. 5. Pseudo codes for evaluating the inter-individual consistency of EEG features.

8. The range of entries in each  $\mathbf{W}_j$  is mapped to the interval  $[0, 1]$  in line 8. Then, each row of these matrices is sorted in a descending order to generate individual ranking matrices  $\mathbf{R}_j$  in lines 9 to 11. It is noted that a subject with high interclass similarity can be ranked in the first place based on Eq. (9). Therefore, lines 12 to 14 ensure that the highest consistency only exists between two identical EEG feature sets.

### 3.4.2. Determination of locally robust features

The subject ranking matrix  $\mathbf{R}_j$  with a size of  $K \times K$  are used to discover locally robust EEG features that share similar PDFs among a subset of all  $K$  subjects. We define that each subset consists of  $P - 1$  neighbor subjects and a target subject. Note that determination of subject-specific (or subject-generic) features is based on  $P = 1$  (or  $P = K$ ) while in this study we set  $1 < P < K$ . For a given database with  $K$  individuals, the case of  $P = 1$  specifies that the target and neighbor subjects are identity. Thus, inter-individual knowledge transferring is unnecessary because of the same feature distribution is employed. On the other hand,  $P = K$  indicates that a target subject is always an element in the set of neighbor subjects. It should be avoided considering the same reason for  $P = 1$ . Such case also fixes all the  $K - 1$  subjects, i.e. except the target, as neighbors. However, knowledge transferring might fail when low feature similarity arose between a target and an improper neighbor.

In Fig. 6, an algorithm is designed for determining the locally robust features. In line 4,  $P - 1$  neighbor subjects are selected around the target subject  $\#k$  based on the subject ranking matrices  $\mathbf{R}_j$ . Since the neighbor subject indices  $\mathbf{s}_p(j)$  corresponding to the same target subject  $\#k$  vary across different EEG features, we take all addresses of entries in  $\mathbf{s}_p$  with  $\mathbf{s}_p(i) == \hat{k}$  as indices of the locally

	<b>Function:</b>	<b>Deter_local</b> ( $\mathbf{R}_j, K, P, \mathbf{X}$ )
	<b>Input:</b>	Subject ranking matrices $\mathbf{R}_j$ Number of subjects $K$ Number of neighbor subjects $P$ EEG features of all subjects $\mathbf{X}$
1	<b>Codes:</b>	<b>for</b> $k = 1:K$
2		<b>for</b> $j = 1:D$
3		<b>for</b> $p = 2:P$
4		$\mathbf{s}_p(j) \leftarrow \mathbf{R}_j(k, p)$
5		<b>for</b> $\hat{k} = 1:K$
6		<b>for</b> $p = 2:P$
7		<b>if</b> $\mathbf{s}_p(i) == \hat{k}$
8		$F_p \leftarrow i$
9		$F(k, \hat{k}) \leftarrow \text{unique}(F_1, F_2, \dots, F_p)$
10		<b>if</b> $k == \hat{k}$
11		$F(k, \hat{k}) \leftarrow \emptyset$
12		$q = 1, \beta(k, \hat{k}) = F(k, \hat{k})$
13		<b>while</b> $\beta(k, \hat{k}) \neq \emptyset$
14		<b>for</b> $j = 1: \beta(k, \hat{k}) $
15		Compute $\Delta\phi_{w^{(i,j)}}(j)$ via Eqn. (11)
16		$\mathbf{R}_j^{(k,\hat{k})}(q) \leftarrow \arg \min \Delta\phi(j)$
17		$\beta(k, \hat{k}) \setminus F_{\arg \min \Delta\phi(j)}$
18		$q = q + 1$
19		<b>Return</b> $F(k, \hat{k}), \mathbf{R}_j^{(k,\hat{k})}$
	<b>Output:</b>	Locally robust feature indices $F(k, \hat{k})$ Locally robust feature rankings $\mathbf{R}_j^{(k,\hat{k})}$

Fig. 6. Pseudo codes for determining the locally robust features.

robust features between subjects  $\#k$  and  $\#\hat{k}$  (as shown in lines 5 to 8). Line 9 ensures that each of the selected feature index in  $F(k, \hat{k})$  is unique.

Given an EEG data point  $\mathbf{x}_i^{(k,\hat{k})} \in R^{|F(k,\hat{k})|}$  from feature subset  $F(k, \hat{k})$  of each pair of subjects, the  $\psi$  shown in Eq. (4) can be rewritten as,

$$\psi(\mathbf{w}^{(k,\hat{k})}) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j C^{(\hat{k})}(i) C^{(\hat{k})}(j) K_L(\mathbf{x}_i^{(k,\hat{k})}, \mathbf{x}_j^{(k,\hat{k})}). \quad (10)$$

In Eq. (10),  $C^{(\hat{k})} \in \{-1, +1\}$  indicates the emotional class of low or high levels on arousal or valence dimension for the EEG data of subject  $\#\hat{k}$ . The norm vector of the classification hyperplane is denoted by  $\mathbf{w}^{(k,\hat{k})}$ . The  $N$  is the number of the training instances. The term  $K_L$  represents a kernel function for computing Lagrange multiplier  $\alpha_i$  in LSSVM.

According to Eqn. (5), the loss of the margin with an EEG feature removed is rewritten as,

$$\Delta \phi_{\mathbf{w}^{(k,\hat{k})}}(j) = \left| \frac{\psi(\mathbf{w}^{(k,\hat{k})}(j)) - \psi(\mathbf{w}^{(k,\hat{k})})}{\psi(\mathbf{w}^{(k,\hat{k})}) \psi(\mathbf{w}^{(k,\hat{k})}(j))} \right|, \quad (11)$$

wherein  $\mathbf{w}^{(k,\hat{k})}(j)$  is the norm vector of the classification hyperplane with the  $j^{\text{th}}$  EEG feature eliminated for training  $\mathbf{w}^{(k,\hat{k})}$ . For lines 13 to 18 in Fig. 6, the selected feature indices  $F(k, \hat{k})$  are ranked as  $R_f^{(k,\hat{k})}$  from high to low importance. Line 16 determines the index of the EEG feature corresponding to the smallest margin loss. Line 17 represents that the optimal EEG feature is removed from the current  $F(k, \hat{k})$ .

#### 3.4.3. Ensemble of feature rankings

The selected locally robust features across  $\#k$  and  $\#\hat{k}$  are recorded in  $F(k, \hat{k})$ . However, when multiple  $\hat{k}$  values are considered, entries in  $F(k, \hat{k})$  could be significantly varied. That is, different features are selected in different training subset. A possible scheme to tackle such feature inconsistency is to merge different  $F(k, \hat{k})$ . Therefore, implementing ensemble classifiers learned based on different training subsets and corresponding  $F(k, \hat{k})$  representations could be a straightforward way to ensure the consistency of the selected feature indices.

Let us denote a testing data point from a target subject  $\#k$  as  $\mathbf{x}_i^{(k,\hat{k})}$  according to feature ranking  $\mathbf{R}_f^{(k,\hat{k})}$ . The estimated emotional class is computed based on an ensemble classifier,

$$\tilde{C}^{(k)}(i) = \sum_{\hat{k}=1}^{K, \hat{k} \neq k} v_{(k,\hat{k})} g_{(k,\hat{k})}(\mathbf{x}_i^{(k,\hat{k})}), \quad (12)$$

where  $g_{(k,\hat{k})}(\cdot)$  denotes a base learner required to be trained on the EEG feature set of subject  $\#\hat{k}$ . Output weight of  $g_{(k,\hat{k})}(\cdot)$  is  $v_{(k,\hat{k})}$  and represents degree of confidence of each learner. Predicted emotional class of  $\mathbf{x}_i^{(k,\hat{k})}$  is defined by  $\tilde{C}^{(k)}(i)$ . When LSSVM is used as the base learner, Eq. (12) can be rewritten as,

$$\tilde{C}^{(k)}(i) = \sum_{\hat{k}=1}^{K, \hat{k} \neq k} v_{(k,\hat{k})} \left( \sum_{j=1}^{N_k} \alpha_j C^{(\hat{k})}(j) K_L(\mathbf{x}_j^{(k,\hat{k})}, \mathbf{x}_i^{(k,\hat{k})}) + b^{(\hat{k},k)} \right). \quad (13)$$

In the equation,  $\mathbf{x}_j^{(k,\hat{k})}$  denotes a training instance from subject  $\#\hat{k}$  according to  $\mathbf{R}_f^{(k,\hat{k})}$ . Bias of the classification hyperplane is denoted by  $b^{(\hat{k},k)}$ . Other notations are consistent with those defined in Eqs. (10) and (12).

<b>Function:</b>	<b>LRFS_emo</b> ( $K, P, X, T, C, \beta; \sigma$ )
<b>Input:</b>	Number of subjects $K$ Number of neighbor subjects $P$ EEG features of all subject $X$ Sampling period of the PDF $T$ Emotional classes $C$ Number of the selected features $\beta$ Width of the Gaussian function $\sigma$
1 <b>Codes:</b>	$\mathbf{R}_j, \mathbf{W}_j \leftarrow \text{Eval_interind}(\mathbf{X}, C, K, T)$
2	$F(k, \hat{k}), \mathbf{R}_f^{(k,\hat{k})} \leftarrow \text{Deter_local}(\mathbf{R}_j, K, P, \mathbf{X})$
3	<b>for</b> $k = 1 : K$
4	<b>for</b> $\hat{k} = 1 : K, \hat{k} \neq k$
5	Select $\mathbf{X}^{(\hat{k},k)}$ from $\mathbf{X}$ via $F(k, \hat{k}), R_f^{(k,\hat{k})}$ , and $\beta$
6	Get $C^{(\hat{k})}$ from $C$
7	Train $g_{(k,\hat{k})}(\cdot)$ based on $\mathbf{X}^{(\hat{k},k)}, C^{(\hat{k})}$
8	Select $\mathbf{X}^{(k,\hat{k})}$ from $\mathbf{X}$ via $F(k, \hat{k}), R_f^{(k,\hat{k})}$ , and $\beta$
9	$\bar{\mathbf{W}} = 1/D \cdot \sum_{j=1}^D \mathbf{W}_j$
10	$\bar{\mathbf{R}} \leftarrow \text{sort}(\bar{\mathbf{W}})$
11	Normalize $\bar{\mathbf{R}}$
12	Generate $\mathbf{V}_{(k,\hat{k})}$ by Eqn. (14) with $\sigma$
13	$\mathcal{E}^{(k)}(i) \leftarrow \sum_{\hat{k}=1}^{K, \hat{k} \neq k} v_{(k,\hat{k})} g_{(k,\hat{k})}(\mathbf{x}_i^{(k,\hat{k})})$
14	<b>Return</b> $\mathcal{E}^{(k)}(i)$
15	<b>Output:</b> Predicted emotional class $\mathcal{E}^{(k)}(i)$

Fig. 7. Pseudo codes for the LRFS-based emotion recognition under an individual-independent paradigm.

In Fig. 7, we list the pseudo codes for feature ranking ensemble and emotion recognition under an individual-independent paradigm. The  $\hat{k}$  and  $k$  denote indices of the training and testing subjects, respectively. In line 7, a base learner is trained by a supervised machine learning algorithm. Lines 10 to 13 show that an output weight is generated based on a Gaussian function,

$$\mathbf{V}_{(k,\hat{k})} = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{\bar{\mathbf{R}}(k, \hat{k})^2}{2\sigma^2}\right), k \neq \hat{k}. \quad (14)$$

The predetermined constants  $\sigma$  is utilized to control the width of the distribution. Then, a larger (or smaller)  $v_{(k,\hat{k})}$  can be assigned to a base learner that is trained by the subject with higher (or lower) consistency. Average subject ranking matrix  $\bar{\mathbf{R}}(k, \hat{k})$  is computed based on average weights  $\mathbf{W}$  of inter-individual consistency across  $D$  EEG features,  $\bar{\mathbf{W}} = 1/D \cdot \sum_{j=1}^D \mathbf{W}_j$ .

## 4. Results

### 4.1. Inter-individual consistency of EEG features

For the DEAP database, the average feature distribution of subjects #27 show low consistency with the other subjects. Similar observation is also found for subject #3 in the MAHNOB-HCI. It indicates that predicting emotions for the two subjects can be incorrect without personalized EEG training data. There also exists

a lot of cases with the consistency values lower than 0.5 in Fig. 8(b). The observation shows the difficulty for recognizing the valence dimension in the DEAP. On the other hand, most of the consistency values in Fig. 8(c) and (d) are larger than 0.5. It suggests that EEG distribution of multiple individuals possesses higher consistency in the MAHNOB-HCI.

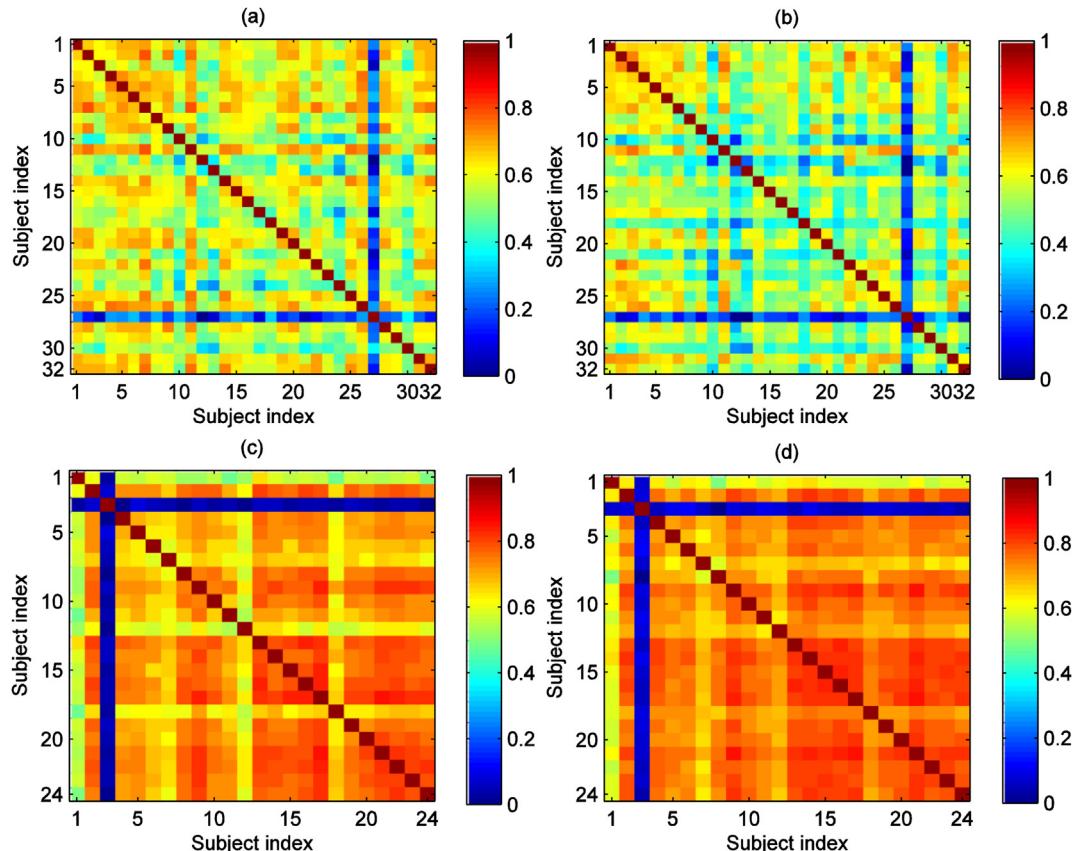
In Fig. 9, PDFs of an EEG feature (average theta power in Fp1 channel) extracted from 12 representative subjects from two databases are illustrated. Subfigures (a), (b), (d), and (e) show that the PDFs of low valence class in the DEAP possess a lower average for subjects #13, #14, #16, and #19. However, a higher average of the low valence level is found for subjects #15 and #20 in subfigures (c) and (f). For the MAHNOB-HCI, the averages of the PDFs for subjects #1, #2, #12, and #24 of the two classes are quite close to each other in subfigures (i) and (k). A larger average of the low valence class is found in subjects #4 and #22 in subfigures (j) and (l) while contradictory distribution exists in subfigures (b) and (e). These observations imply the difficulty of the individual-independent emotion recognition.

#### 4.2. Classification performance on each consistency indicator

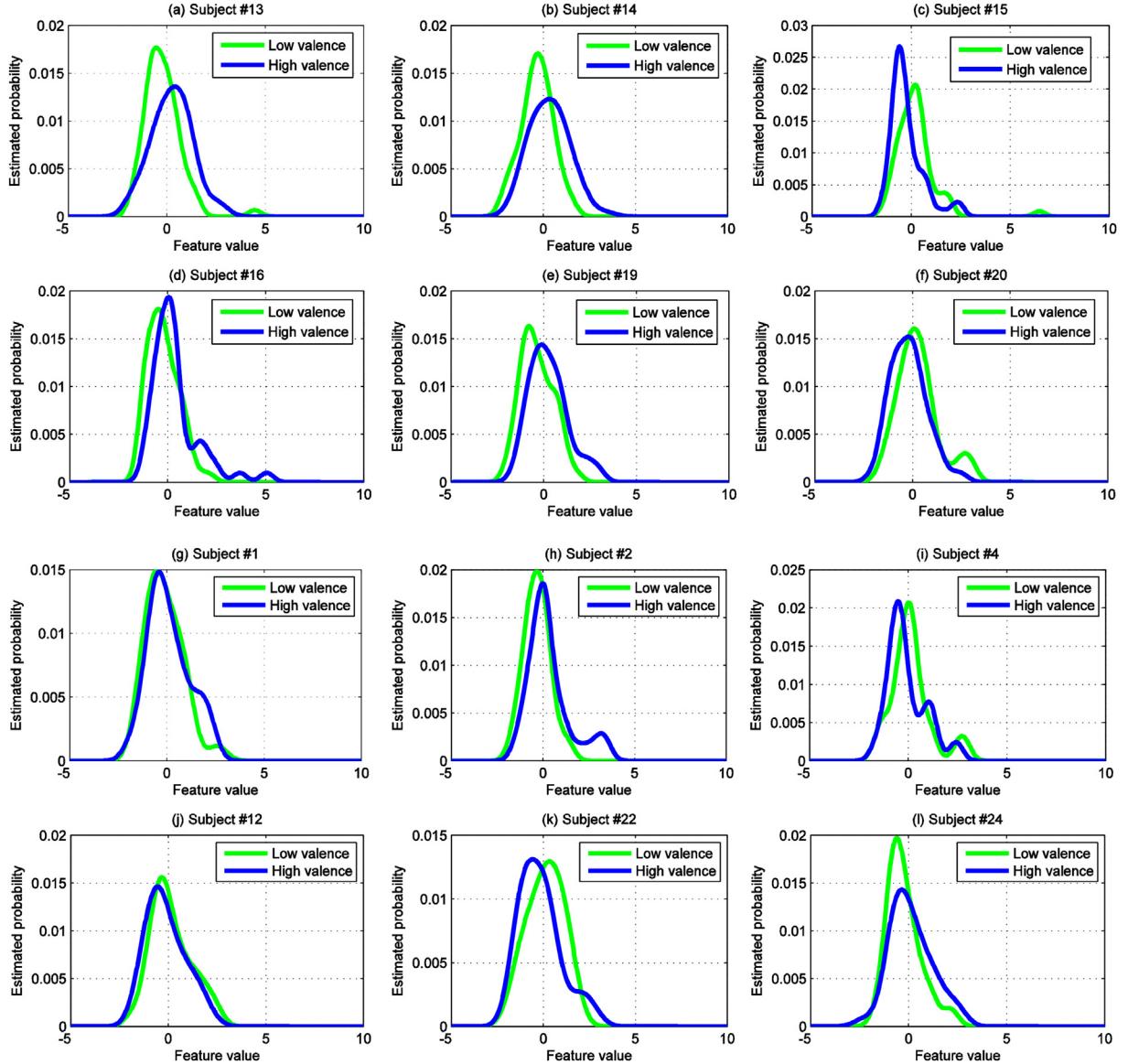
To optimize LRFS's performance, classification accuracy and F1 score of each indicator, i.e.,  $\rho_1$ ,  $\rho_2$ , and  $\rho_3$ , for evaluating the inter-individual consistency is examined in Fig. 10. A leave-one-subject-out paradigm was employed. In a testing iteration, 31 (or 23) subjects' data of the DEAP (or MAHNOB-HCI) database are used for training while the remaining data are used for testing the trained classifier. Number of the iteration is equal to the size of the subject pool of each database. We first determine the optimal

regularization parameter of a linear LSSVM within a candidate set  $\{10^{-1}, 1, 10, 10^2, 10^3\}$  according to the highest leave-one-subject-out accuracy. We also initiate a RBF-LSSVM that inherit the selected regularization parameter and then choose the kernel width among  $\{10^{-1}, 1, 10, 10^2, 10^3\}$  in the same manner. Then, these two model structures are compared based on the optimal accuracy to train the final LSSVM for the emotion recognition. In the end, the LSSVM with the RBF kernel function is used as the base learner. The regularization parameter and the kernel width are predetermined at 100 and 10, respectively. The LRFS's parameters are selected as follows. The sampling period  $T$  of the PDF and the number of neighbor subjects are set to 0.0375 and 3, respectively. Number of selected features is determined based on the 75% of the locally robust features of each base learner according to the feature ranking. The width of the Gaussian function used in feature ranking ensemble is set to 6.

From the figure, it is seen that average accuracy and F1 score generated by  $\rho_1$  outperform those of  $\rho_2$  and  $\rho_3$ . In Fig. 10(a) and (b),  $\rho_1$  achieves the highest accuracy and F1 score on the arousal dimension of the DEAP database by 0.6510 and 0.6299, respectively. For  $\rho_2$  and  $\rho_3$ , the accuracy falls to 0.5974 and 0.5744 and the F1 score falls to 0.6137 and 0.6029, respectively. Similar observations are shown in Fig. 10(c) and (d) on the valence dimension, wherein the highest accuracy and F1 score of  $\rho_1$  are 0.6797 and 0.6689. Results of the MAHNOB-HCI database are consistent with that of the DEAP. In Fig. 10(e)–(g), the highest average accuracy and F1 score elicited by  $\rho_1$  are 0.6743, 0.6858, 0.6993, and 0.7122, respectively. Therefore, the indicator of the inter-individual consistency computed by  $\rho_1$  achieves the optimal performance and is employed for further investigation of the LRFS.



**Fig. 8.** Average inter-individual consistency of all EEG features. Results of (a) arousal and (b) valence dimensions of the DEAP database and (c) arousal and (d) valence dimensions of the MAHNOB-HCI database.



**Fig. 9.** Estimated probability density functions of average theta power in Fp1 channel for high and low valence classes. Function plots from 12 representative subjects in the (a)–(f) DEAP and (g)–(l) MAHNOB-HCI databases are shown.

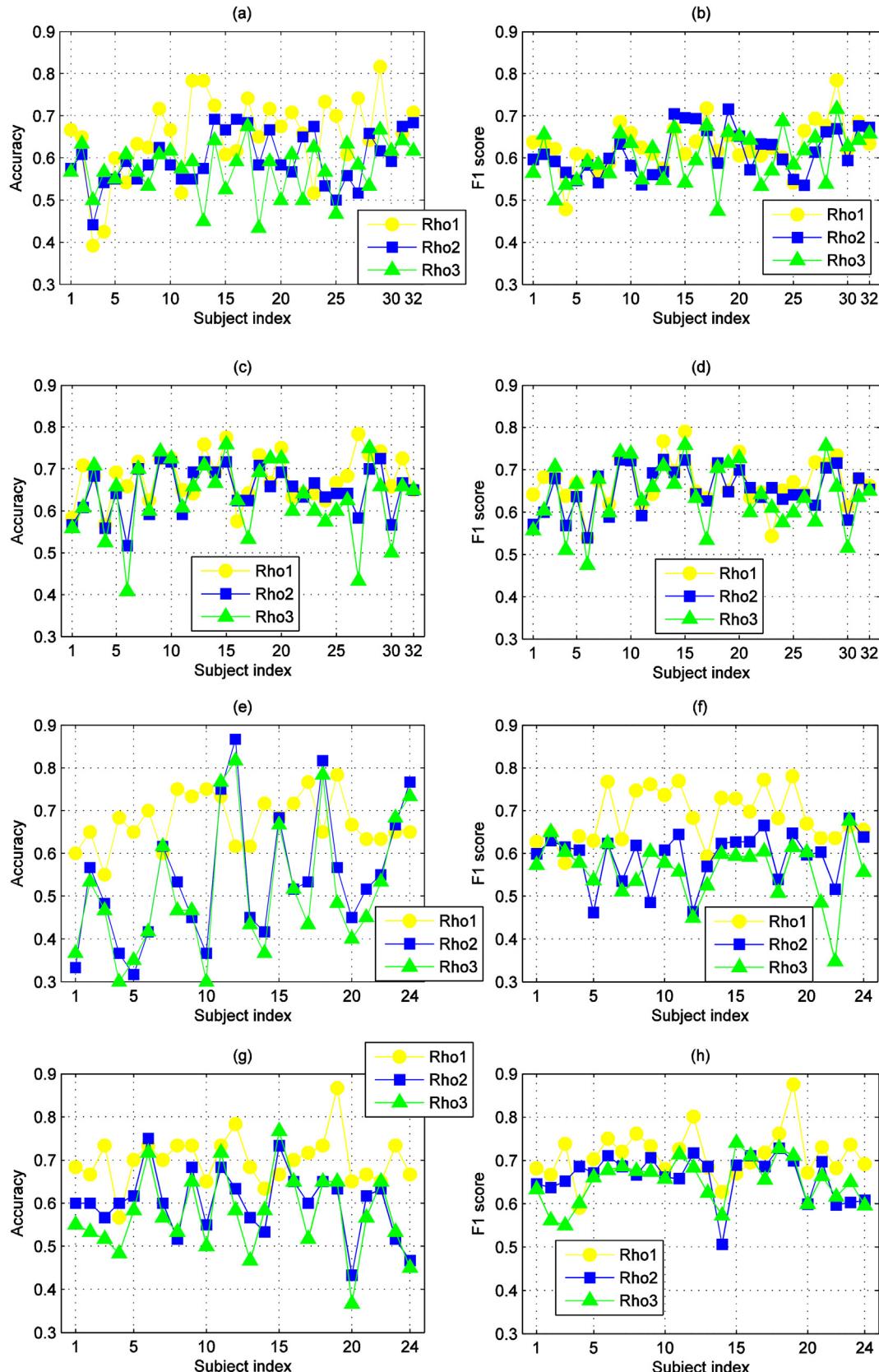
#### 4.3. Classification performance comparison

In this section, we compare LRFS's performance with several existing feature selection approaches combined with seven machine learning classifiers in [Tables 3 and 4](#). Feature rankings of all methods were first elicited based on the whole dataset. Then, the leave-one-subject-out paradigm was employed again to derive the classification metrics.

The feature selection was separately carried out with baseline condition (all features were adopted), correlation coefficient (CC) method, dependence guided unsupervised feature selection (DGU-FS) ([Guo & Zhu, 2018](#)), infinite latent feature selection (INF-FS) ([Roffo, Melzi, & Cristani, 2015](#)), local learning-based clustering (LLC) ([Wu, Wang, Bu, & Chen, 2016](#)), classical RFE, and the proposed LRFS. It is noted that the CC, INF-FS, RFE, and LRFS are supervised by target emotional classes while DGU-FS and LLC are unsupervised approaches. To facilitate a fair comparison, we have examined three cases with the first 25%, 50%, or 75% features adopted according to their rankings among all combinations of feature selection and classification methods, wherein the

highest accuracy of the three cases is used for comparison. The parameters of the LRFS method are consistent with those set in [Section 4.2](#).

The employed classifiers include naive Bayesian model (NB), logistic regression (LR), K-nearest neighbor (KNN), extreme learning machine (ELM) ([Huang, Zhu, & Siew, 2006](#)), artificial neural network with single hidden layer (ANN), deep denoising stacked autoencoder with three hidden layers (SAE), and LSSVM. Number of neurons in each hidden layer of ELM, ANN, and SAE is separately investigated from a candidate set, {50, 100, 150, 200, 250}. The optimal parameter corresponding to the highest accuracy is selected. Number of nearest neighbors of KNN was examined based on a set {5, 10, 15, 20, 25}. The way of selecting the optimal regularization parameter and kernel width are consistent with those described in [Section 4.2](#). For the LR and NB, none of hyperparameters is required to be predetermined. A modern deep learning framework developed in our previous work ([Yang et al., 2019](#)) has been also implemented, i.e., ensemble stacked denoising autoencoder with local information preservation (denoted as EL-SDAE).



**Fig. 10.** Classification accuracy and F1 score elicited by the LRFS and LSSVM classifier on each subject: (a)–(b) Arousal and (c)–(d) valence classification results of the DEAP database; (e)–(f) Arousal and (g)–(h) valence classification results of the MAHNOB-HCI database.

**Table 3**

Classification performance comparison across different combinations of feature selection methods and machine learning classifiers on the DEAP database.

Performance	Classifier	Feature selection method						
		Baseline	CC	DGU-FS	INF-FS	LLC	RFE	LRFS
Average arousal accuracy	NB	0.4951*	0.4844*	0.5091*	0.5065*	0.5133*	0.5635*	<b>0.6313</b>
	LR	0.5469*	0.5802	0.5451*	0.5646*	0.5544*	0.5398*	<b>0.6065</b>
	KNN	0.5599*	0.5659*	0.5594*	0.5612*	0.5695*	0.5641*	<b>0.6005</b>
	ELM	0.5586*	0.5727*	0.5646*	0.5661*	0.5685*	0.5643*	<b>0.6219</b>
	ANN	0.5888*	0.5646*	0.5737*	0.5747*	0.5641*	0.5815*	<b>0.6185</b>
	SAE	0.5833	0.5781	0.5747	0.5646	0.5828	0.5789	<b>0.5872</b>
	EL-SDAE	0.5672	0.5609	0.5417	0.5359	<b>0.5710</b>	0.5570	0.5042
Average valence accuracy	LSSVM	0.5578*	0.5596*	0.5461*	0.5547*	0.5581*	0.5427*	<b>0.6510</b>
	NB	0.5125*	0.5148*	0.5081*	0.5156*	0.5125*	0.5070*	<b>0.6732</b>
	LR	0.5693*	0.5646*	0.5609*	0.5797*	0.5799*	0.5424*	<b>0.6487</b>
	KNN	0.5466*	0.5555*	0.5427*	0.5437*	0.5516*	0.5599*	<b>0.6513</b>
	ELM	0.5625*	0.5716*	0.5604*	0.5656*	0.5604*	0.5352*	<b>0.6378</b>
	ANN	0.5674*	0.5651*	0.5698*	0.5682*	0.5523*	0.5464*	<b>0.6560</b>
	SAE	0.5635	0.5482	0.5497	0.5568	0.5484	0.5182*	<b>0.5750</b>
F1-score	EL-SDAE	0.5404	0.5534	0.5451	0.5375	<b>0.5609</b>	0.5388	0.5234
	LSSVM	0.5724*	0.5729*	0.5568*	0.5844*	0.5758*	0.5437*	<b>0.6797</b>
	NB	0.5281*	0.5130*	0.5228*	0.5254*	0.5218*	0.5245*	<b>0.6188</b>
	LR	0.5229*	0.5326*	0.4896*	0.5287*	0.5085*	0.4714*	<b>0.5869</b>
	KNN	0.5165	0.5047	0.5050	<b>0.5202</b>	0.4978	0.5043	0.5154
	ELM	0.4939*	0.5237*	0.5126*	0.5131*	0.5138*	0.4808*	<b>0.6080</b>
	ANN	0.5187*	0.4924*	0.5140*	0.5180*	0.4897*	0.4937*	<b>0.5909</b>
Average valence F1-score	SAE	0.5025	0.4723	0.4252	<b>0.5034</b>	0.4078	0.4672	0.4228
	EL-SDAE	0.5030	0.5173	0.5167	<b>0.5222</b>	0.5277	0.5181	0.5120
	LSSVM	0.5310*	0.5146*	0.4891*	0.5117*	0.5110*	0.4672*	<b>0.6299</b>
	NB	0.5491*	0.5529*	0.5431*	0.5502*	0.5457*	0.5206*	<b>0.6689</b>
	LR	0.5597*	0.5469*	0.5478*	0.5682*	0.5710*	0.5259*	<b>0.6381</b>
	KNN	0.5119*	0.5181*	0.5142*	0.5137*	0.5079*	0.5281*	<b>0.6388</b>
	ELM	0.5488*	0.5572*	0.5443*	0.5510*	0.5450*	0.5154*	<b>0.6308</b>
F1-score	ANN	0.5512*	0.5405*	0.5473*	0.5433*	0.5282*	0.5122*	<b>0.6337</b>
	SAE	0.5357	0.5280	0.5232	<b>0.5402</b>	0.5296	0.4959	0.4582
	EL-SDAE	0.5017	0.5193	0.5232	0.5224	0.5405	0.5163	0.4864
	LSSVM	0.5622*	0.5631*	0.5440*	0.5723*	0.5654*	0.5265*	<b>0.6689</b>

Note: The highest performance metric in each row is in boldface. The star marker denotes the LRFS combined with the current classifier significantly outperforms the marked feature selection method based on a two-tailed, paired t-test.

For DEAP's results shown in [Table 3](#), it is observed that the LRFS achieves the highest average performance when combined with the NB, LR, ELM, ANN, and LSSVM methods. The EL-SDAE is comparable with SDAE and achieves the highest accuracy when combining with the LLC approach. The combination of the LRFS and EL-SDAE does not significantly improve the generalization capability. Similar observation is found in the results of the MAHNOB-HCI database summarized in [Table 4](#). The highest F1-score is derived from the hybrid classifier of the LRFS and EL-SDAE. Under the same condition, the EL-SDAE is also superior to the classical SAE. However, all the improvements are insignificant according to the *t*-test. In particular, the combination between the LRFS and LSSVM achieves the highest performance except for the case that the NB and KNN are used as base learners to recognize the valence levels. A two-tailed, paired *t*-test is carried out to compare the LRFS with the other feature selection methods. The significance level is generated according to Bonferroni correction as  $0.05/6 = 0.0083$ . A significant improvement of the LRFS with the baseline condition, CC, DGU-FS, INF-FS, LLC, and RFE is observed for the NB, LR, and LSSVM. For ELM and ANN, an insignificance of the LRFS is found when recognizing MAHNOB-HCI's arousal levels. The LRFS-based KNN fails to recognize arousal levels. When implementing SAE as base classifiers, LRFS is comparable with other feature selection methods.

#### 4.4. Performance of the LRFS with different hyper-parameters

Two critical hyper-parameters, i.e., number of the neighbor subjects  $P$  and width of the Gaussian function of feature ranking ensemble  $\sigma$ , affect LRFS's performance.

By setting the experimental condition and the other parameters as same as those in [Section 4.2](#), we investigate the accuracy and F1 score of the hybrid emotion classifier, the LSSVM combined with the LRFS, with different choices of  $\sigma$  in [Fig. 11](#). A small value of  $\sigma$  implies that only a few base learners of high ranks are utilized to make decisions of emotions. On the other hand, a large value indicates that all base learners share similar output weights. For the DEAP (or MAHNOB-HCI) database, a large (or moderate)  $\sigma$  within a range of [8, 12] (or [4, 6]) is preferable. Note that a small value of  $\sigma$  undermines LRFS's performance for both databases. These observations also suggest the accuracy on the MAHNOB-HCI is more sensitive to variation of  $\sigma$  than that of DEAP.

In [Fig. 12](#), the number of the neighbor subjects was examined within a set {1, 2, 3, 4, 5} under the same condition. For both datasets and emotional dimensions, two neighbor subjects could help the LRFS achieve the highest performance. It is observed that setting too few or too many neighbor subjects impairs LRFS's performance.

#### 4.5. Computational cost analysis

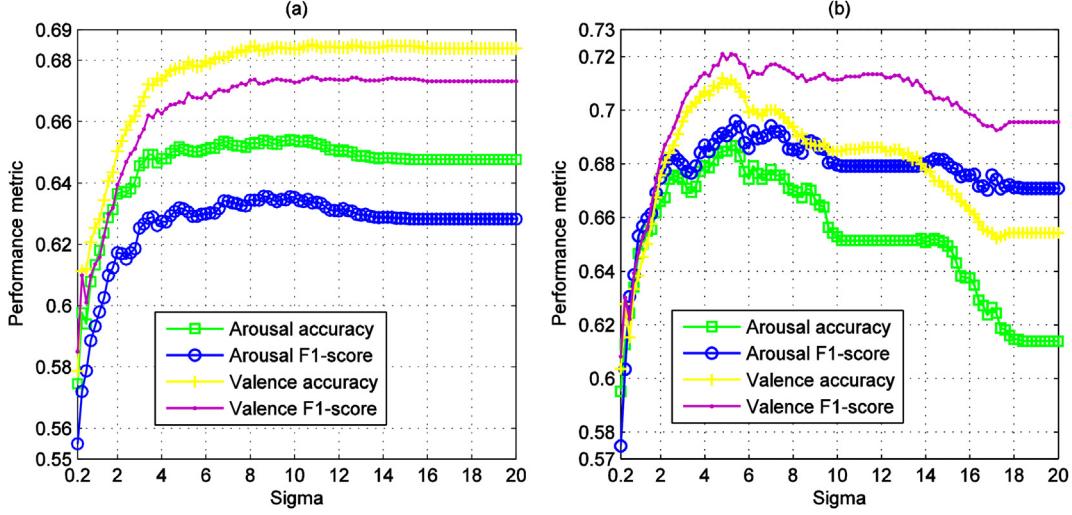
Computational costs of all feature selection approaches evaluated by CPU time for generating feature rankings on two databases are shown in [Table 5](#). We tested all algorithms via ten same trials and compute mean and standard deviation of the time costs. The algorithms were implemented based on Matlab® 2011b software and a laptop computer with Intel® i7 CPU 2.5GHZ and 8GRAM configurations. From the table, we observe that the LRFS outperforms the DGU-FS, LLC, and RFE on the DEAP database and is slightly better than RFE but worse than the remaining approaches on

**Table 4**

Classification performance comparison across different combinations of feature selection methods and machine learning classifiers on the MAHNOB-HCI database.

Performance	Classifier	Feature selection method					
		Baseline	CC	DGU-FS	INF-FS	LLC	RFE
Average arousal accuracy	NB	0.5215*	0.5222*	0.5174*	0.5250*	0.5375*	0.5340*
	LR	0.5347	0.5299	0.5326	0.5410	0.5313	0.5514
	KNN	0.5243*	0.5306*	0.5167*	0.5188*	0.5215*	0.5208*
	ELM	0.5514	0.5417	0.5368	0.5201*	0.5049*	0.5472
	ANN	0.5431	0.5563	0.5326	0.5563	0.5417	0.5625
	SAE	0.5597	0.5403	0.5299	0.5375	0.5312	0.5153
	EL-SDAE	<b>0.5611</b>	0.5486	0.5389	0.5451	0.5069	0.5451
Average valence accuracy	LSSVM	0.5396*	0.5326*	0.5271*	0.5208*	0.5319*	0.5583*
	NB	0.5215*	0.5285*	0.5181*	0.5264*	0.5472*	0.5403*
	LR	0.5104*	0.5285*	0.5444*	0.5243*	0.4979*	0.5090*
	KNN	0.5806*	0.5590*	0.5681*	0.5743*	0.5576*	0.5569*
	ELM	0.5542*	0.4979*	0.5507*	0.5542*	0.5333*	0.5278*
	ANN	0.5576*	0.5521*	0.5285*	0.5222*	0.5437*	0.5229*
	SAE	0.5479	0.5222	0.5549	0.5299	0.5382	0.5160
F1-score	EL-SDAE	0.5576	0.5563	0.5451	0.5375	0.5514	0.5229
	LSSVM	0.5264*	0.5278*	0.5424*	0.5250*	0.5021*	0.5118*
	NB	0.5418*	0.5453*	0.5392*	0.5460*	0.5611*	0.5526*
	LR	0.5422*	0.5387*	0.5347*	0.5438	0.5434	0.5623
	KNN	0.5351*	0.5398*	0.5275*	0.5286*	0.5353*	0.5307*
	ELM	0.5593*	0.5437*	0.5429*	0.5290*	0.5160*	0.5538*
	ANN	0.5515*	0.5713	0.5510*	0.5735	0.5556*	0.5760
Average valence F1-score	SAE	<b>0.5596</b>	0.5581	0.5459	0.5514	0.5252	0.4978
	EL-SDAE	0.5504	0.5691	0.5553	0.5235	0.5170	0.5530
	LSSVM	0.5461*	0.5430*	0.5349*	0.5282*	0.5433*	0.5704*
	NB	0.5221*	0.5293*	0.5163*	0.5355*	0.5490*	0.5248*
	LR	0.5053*	0.5265*	0.5438*	0.5197*	0.4939*	0.5037*
	KNN	0.5768*	0.5504*	0.5619*	0.5691*	0.5477*	0.5532*
	ELM	0.5532*	0.4953*	0.5479*	0.5542*	0.5326*	0.5210*
F1-score	ANN	0.5571*	0.5423*	0.5161*	0.4535*	0.5389*	0.5114*
	SAE	0.5472	0.5194	<b>0.5561</b>	0.5302	0.5278	0.5097
	EL-SDAE	0.5610	0.5526	0.5392	0.5331	0.5465	0.5185
	LSSVM	0.5217*	0.5256*	0.5398*	0.5214*	0.4989*	0.5062*
	NB	0.5221*	0.5293*	0.5163*	0.5355*	0.5490*	0.5248*
	LR	0.5053*	0.5265*	0.5438*	0.5197*	0.4939*	0.5037*
	KNN	0.5768*	0.5504*	0.5619*	0.5691*	0.5477*	0.5532*
Average arousal F1-score	ELM	0.5532*	0.4953*	0.5479*	0.5542*	0.5326*	0.5210*
	ANN	0.5571*	0.5423*	0.5161*	0.4535*	0.5389*	0.5114*
	SAE	0.5472	0.5194	<b>0.5561</b>	0.5302	0.5278	0.5097
	EL-SDAE	0.5610	0.5526	0.5392	0.5331	0.5465	0.5185
	LSSVM	0.5217*	0.5256*	0.5398*	0.5214*	0.4989*	0.5062*

Note: All notations are consistent with Table 3.



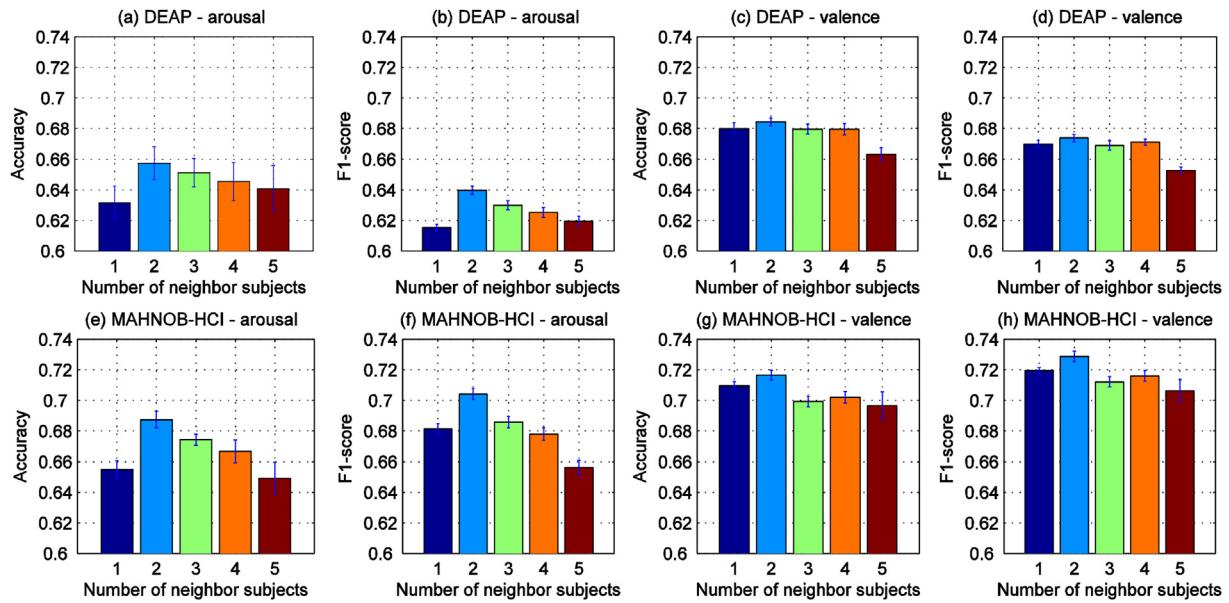
**Fig. 11.** Leave-one-subject-out performance of the LSSVM combined with the LRFS at different width of the Gaussian function (denoted as sigma) of classifier ensemble. Results of the (a) DEAP and (b) MAHNOB-HCI databases are shown.

MAHNOB-HCI. Considering a higher accuracy of the LRFS shown in the previous sections, the proposed method is competitive on the issue of individual-independent classification of emotions.

## 5. Discussions

In general, the proposed LRFS has capability to improve the accuracy of individual-independent emotion recognition on both

physiological databases. The existing feature selection approaches attempt to find generic EEG features. However, such task could be difficult since individual difference exists in specific feature PDFs as shown in Figs. 8 and 9. The essential of the LRFS is to determine the robust EEG features within a subset of the subject pool. Each subset consists of several individuals with consistent feature properties. In particular, the optimum of indicator  $\rho_1$  (see Fig. 10) implies that the inter-individual consistency of EEG features make



**Fig. 12.** Leave-one-subject-out performance of LSSVM combined with the LRFS under different values of neighbor subjects. The error bar represents variance of the classification metrics of all subjects in a database.

**Table 5**

Average CPU time (in sec) for eliciting EEG feature rankings of each database. Mean and standard deviation (s.d.) are computed across 10 repeated trials.

Feature selection methods	DEAP		MAHNOB-HCI	
	Average	s.d.	Average	s.d.
CC	0.06	0.02	0.03	0.01
DGU-FS	389.63	5.48	29.15	0.29
INF-FS	0.12	0.02	0.07	0.02
LLC	1233.65	21.85	100.99	0.85
RFE	1509.92	41.45	125.83	0.53
<b>LRFS</b>	<b>216.63</b>	<b>5.68</b>	<b>117.56</b>	<b>0.72</b>

Note: The results of the proposed method are in boldface.

more contribution than the inter-class discriminating capability for individual-independent classification.

Table 6 lists performances of existing emotion recognition systems with the same databases used. For the DEAP database, the proposed method outperforms (Koelstra et al., 2012) and (Wang & Shang, 2013) and is comparable with (Chen et al., 2015) on binary emotion classification. Since the individual-specific paradigm adopted much less training data than the individual-independent case, high dimensional EEG features may lead to overfitting. Higher accuracies were achieved by (Atkinson & Daniel, 2016) and (Yin, Liu, & et al., 2017). For the latter case, the D-RFE method has to employ the baseline EEG data from the same testing individual to calibrate the training set. This mechanism cannot facilitate an individual-independent paradigm. For the MAHNOB-HCI database, the proposed method outperforms (Zhu, Wang, & Ji, 2014) and the arousal recognition results of (Pereira, Gomes, Veloso, & Mota, 2018). Considering that lower accuracies with three emotional classes were achieved by (Soleymani, Lichtenauer, Pun, & Pantic, 2012) and (Huang et al., 2016), the LRFS based classifier is comparable with these methods. Promising performance was obtained by (Nakisa et al., 2018). However, only 15 subjects were randomly selected for classification in their work. Although it is impractical to examine all accuracies quantitatively because of different experimental environments, the LRFS feature selection method show specific competitiveness under the individual-independent paradigm. The elicited results are also comparable with a part of the existing works under individual-specific and generic conditions.

Another novelty of the LRFS is an application of classifier ensemble for feature ranking fusion. Under the leave-one-subject-out paradigm, a subject was tested based on the ranking sequence generated from his/her neighbor subjects. To exploit multiple ranking sequences, emotions were separately predicted by different selected features and these decisions were combined based on the weights coded by the average inter-individual consistency of EEG features. As shown in Table 5, this mechanism reduces the computational cost of the LRFS. The reason is that the main computational complexity of RFE and LLC is  $O(n^2 \cdot D^2)$  (with  $n$  EEG instances of  $D$  features) since kernel and similarity matrices are computed across each two data points. On the other hand, the main computational complexity of the LRFS is  $O(n^2 \cdot D^2 / K^2) \cdot K^2$  (with  $K$  subjects). It implies that the LRFS cost less time than LLC and RFE with an increased value of  $n$  and a fixed value of  $K$ .

In a recent study reported by (Wang, Chen, & Cao, 2020), convolutional neural networks (CNNs) and recurrent neural networks were combined to process facial expression and speech signals to recognize six emotions. The highest classification accuracies on two modalities are 88.96% and 85.46%, respectively. A visual-audio emotion recognition system was proposed based on ensemble learning framework (Hao et al., 2020), where the multi-task CNN has been applied. The individual-independent and individual-dependent accuracies on the eENTERFACCE database were achieved at 81.36% and 78.42%, respectively. In (Rao, Rao, &

**Table 6**

Accuracy comparison of the proposed method with some of EEG-based emotion recognition systems validated by using the DEAP and MAHNOB databases.

Reference index	Features	Feature selection or fusion	Number of classes & Classifier	Database	Training and testing scheme	Accuracy
Koelstra et al. (2012)	EEG power	Correlation coefficient	2-class, NB	DEAP	Individual specific	Arousal: 0.62 Valence: 0.58
Liu and Sourina (2012)	EEG fractal dimension	–	4-class, SVM	DEAP	Individual specific	Arousal: 0.76 Valence: 0.51
Yoon and Chung (2013)	EEG power.	Correlation coefficient	2/3-class, Modified NB with weighted-log-posterior function	DEAP	Individual specific	Arousal: 0.71 Valence: 0.70
Wang and Shang (2013)	Raw EEG	–	2-class, DBN	DEAP	Individual generic	Arousal: 0.51 Valence: 0.61
Chen et al. (2015)	EEG entropy, temporal statistics.	Gain ratio	2-class, C4.5	DEAP	Individual specific	Arousal: 0.69 Valence: 0.68
Atkinson & Daniel (2016)	EEG power, fractal dimension	mRMR	2-class, SVM	DEAP	Individual specific	Arousal: 0.73 Valence: 0.73
Yin, Liu, and et al. (2017)	EEG power, temporal statistics	D-RFE	2-class, LSSVM	DEAP	Individual generic	Arousal: 0.79 Valence: 0.71
Soleymani, Lichtenauer, Pun, and Pantic (2012)	EEG power	ANOVA	3-class, SVM	MAHNOB-HCI	Individual independent	Arousal: 0.52 Valence: 0.57
Zhu et al. (2014)	EEG power	CCA	2-class, SVM	MAHNOB-HCI	Individual generic	Arousal: 0.60 Valence: 0.56
Huang et al. (2016)	EEG power	KLPP	3-class, SVM	MAHNOB-HCI	Individual independent	Arousal: 0.62 Valence: 0.61
Pereira, Gomes, Veloso, & Mota (2018)	High order crossings	–	2-class, SVM	MAHNOB-HCI	Individual specific	Arousal: 0.59 Valence: 0.76
Nakisa et al. (2018)	Temporal/Frequency statistics	Evolutionary computation	4-class, PNN	DEAP/ MAHNOB-HCI	Individual generic	0.67/0.97
Proposed method	Temporal/Frequency statistics	LRFS	2-class, LSSVM, NB	DEAP/ MAHNOB-HCI	Individual independent	Arousal: 0.65/0.67 Valence: 0.68/0.70

(Chowdary, 2019), a framework using speech (Mel frequency cepstral coefficients, MFCS) and facial features (maximally stable extremal regions, MSER) was proposed. Accuracies computed via the MFCS and MSER were 73.01% and 77%, respectively. In Hung, Lin, and Lai (2019), the CNN model was combined with transfer learning principle to evaluate facial emotion variations. The average accuracy of 91.93% was achieved.

It is worth noted that both speech and facial expressions have a capability to reach higher performance on multiclass emotion recognition (from 73% to 91%). In comparison, the proposed BCI approach achieves relatively lower performance because of high individual difference of EEG features and its sensitivity to movement noise. However, the latter is able to measure emotional clues from human with diseases of facial paralysis or muscular disorders. It could be also continuously recorded in a silent task environment of human machine collaboration. Specifically, cortical responses of emotions can be detected by the BCI before triggering activations of muscular. Therefore, BCI approaches for emotion recognition have advantages in applications with the medical care system, driving system, and safety-critical tasks.

## 6. Conclusion

In this study, the proposed LRFS framework has made a contribution to achieve proper feature selection for emotion recognition under individual-independent paradigm. It is practical since the individual-specific and individual-generic classification frameworks have disadvantage in online applications. The essential of the LRFS is to find the robust EEG features within a subset of all individuals. The DEAP and MAHNOB-HCI databases are used to validate the classification performance on 32 and 24 subjects, respectively. Based on 364 EEG features, the LRFS achieves competitive classification performance combined with multiple machine learning classifiers. It should be noted that the main limitation of the

LRFS is an employment of two critical hyper-parameters, i.e., the number of the neighbor-subjects and width of the Gaussian function of the classifier ensemble. They significantly affect the accuracy and are required to be carefully predetermined before implementing the algorithms. In future works, we believe that better analysis of inter-individual similarity of EEG features could further improve generalization capacity for the proposed method since the current accuracy is far from the perfect.

## CRediT authorship contribution statement

**Zhong Yin:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **Lei Liu:** Software, Data curation, Investigation. **Jianing Chen:** Resources, Visualization. **Boxi Zhao:** Resources, Visualization. **Yongxiong Wang:** Software, Writing - review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work is sponsored by the National Natural Science Foundation of China under Grant No. 61703277, Shanghai Sailing Program under Grant No. 17YF1427000, National Natural Science Foundation of China under Grant No. 61673276 and Natural Science Foundation of Shanghai under Grant No. 17ZR1419000. The research in this paper uses the MAHNOB Database collected by Professor Pantic and the iBUG group at Imperial College London, and in part collected in collaboration with Prof. Pun and his team of University of Geneva, in the scope of MAHNOB project financially supported by

the European Research Council under the European Community's 7th Framework Programme (FP7/20072013)/ERC Starting Grant agreement No. 203143.

## References

- Alazraia, R., Alwanni, H., & Daoud, M. I. (2019). EEG-based BCI system for decoding finger movements within the same hand. *Neuroscience Letters*, 698, 113–120.
- Atkinson, J., & Daniel, C. (2016). Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Systems with Applications*, 47, 35–41.
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience and Biobehavioral Reviews*, 44, 58–75.
- Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1, 18–37.
- Chen, J., Hua, B., Moore, P., Zhang, X., & Ma, X. (2015). Electroencephalogram-based emotion assessment system using ontology and data mining techniques. *Applied Soft Computing*, 30, 663–674.
- Christensen, J., Estepp, J., Wilson, G., & Russell, C. (2012). The effects of day-to-day variability of physiological data on operator functional state classification. *NeuroImage*, 59, 57–63.
- Critchley, H., Mathias, C., Josephs, O., O'Doherty, J., Zanini, S., Dewar, B.-K., ... Dolan, R. (2003). Human cingulate cortex and autonomic control: Converging neuroimaging and clinical evidence. *Brain*, 126, 2139–2152.
- Gundel, A., & Wilson, G. (1992). Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain Topography*, 5, 17–25.
- Guo, J., & Zhu, W. (2018). Dependence guided unsupervised feature selection. In Conference on artificial intelligence (pp. 2232–2239). AAAI.
- Gurker, J. W., & Isaacowitz, D. M. (2019). Emotion regulation and emotion perception in aging: A perspective on age-related differences and similarities. *Progress in Brain Research*. <https://doi.org/10.1016/bs.pbr.2019.02.007> (in press).
- Gramacki, A. (2019). *Nonparametric kernel density estimation and its computational aspects*. Springer.
- Hao, M., Cao, W.-H., Liu, Z.-T., Wu, M., & Xiao, P. (2020). Visual-audio emotion recognition based on multi-task and ensemble learning with multiple features. *Neurocomputing*, 391, 42–51.
- Hassan, M. M., Alam, G. R., Uddin, Z., Huda, S., Almogren, A., & Fortino, G. (2019). Human emotion recognition using deep belief network architecture. *Information Fusion*, 51, 10–18.
- Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70, 489–501.
- Huang, X., Kortelainen, J., Zhao, G., Li, X., Moilanen, A., Seppänen, T., & Pietikäinen, M. (2016). Multi-modal emotion analysis from facial expressions and electroencephalogram. *Computer Vision and Image Understanding*, 147, 114–124.
- Hung, J. C., Lin, K.-C., & Lai, N.-X. (2019). Recognizing learning emotion based on convolutional neural networks and transfer learning. *Applied Soft Computing*, 84, 105724.
- Iacoviello, D., Petracca, A., Spezialetti, M., & Placidi, G. (2015). A real-time classification algorithm for EEG-based BCI driven by self-induced emotions. *Computer Methods and Programs in Biomedicine*, 122, 293–303.
- Jain, D. K., Shamsolmoali, P., & Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. *Pattern Recognition Letters*, 120, 69–74.
- King, R. B., & dela Rosa, E. D. (2019). Are your emotions under your control or not? Implicit theories of emotion predict well-being via cognitive reappraisal. *Personality and Individual Differences*, 138, 177–182.
- Kober, H., Barrett, L., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical-subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage*, 42, 998–1031.
- Koelstra, S., Muehl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T. E., ... Patras, I. Y. (2012). DEAP: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing*, 3, 18–31.
- LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, 73, 653–676.
- Liang, Z., Shigeyuki, O., & Shin, I. (2019). An unsupervised EEG decoding system for human emotion recognition. *Neural Networks*, 116, 257–268.
- Liu, Y., & Sourina, O. (2012). In *EEG-based valence level recognition for realtime applications* (pp. 53–60). IEEE.
- Lopez-de-Ipiña, K., Alonso, J. B., Solé-Casals, J., Barroso, N., Henriquez, P., Faun-de-Zanuy, M., ... Eguiraun, H. (2013). On automatic diagnosis of Alzheimer's disease based on spontaneous speech analysis and emotional temperature. *Cognitive Computation*, 7, 44–55.
- Mano, L. Y., Mazzo, A., Neto, J. R. T., Meska, M. H. G., Giancristofaro, G. T., Ueyama, J., & Júnior, G. A. P. (2019). Using emotion recognition to assess simulation-based learning. *Nurse Education in Practice*, 36, 13–19.
- Menon, V., Rivera, S. M., White, C. D., Glover, G. H., & Reiss, A. L. (2000). Dissociating pre-frontal and parietal cortex activation during arithmetic processing. *NeuroImage*, 12, 357–365.
- Nakisa, B., Rastgoor, M. N., Tjondronegoro, D., & Chandran, V. (2018). Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors. *Expert Systems with Applications*, 93, 143–155.
- Panksepp, J., Lane, R. D., Solms, M., & Smith, R. (2017). Reconciling cognitive and affective neuroscience perspectives on the brain basis of emotional experience. *Neuroscience & Biobehavioral Reviews*, 76, 187–215.
- Pereira, E. T., Gomes, H. M., Veloso, L. R., & Mota, M. A. (2018). Empirical evidence relating EEG signal duration to emotion classification performance. *IEEE Transactions on Affective Computing*, in press. <https://doi.org/10.1109/TACFFC.2018.2854168>.
- Quan, C., & Ren, F. (2016). Weighted high-order hidden Markov models for compound emotions recognition in text. *Information Sciences*, 329, 581–596.
- Rao, K. P., Rao, M. V. P. C. S., & Chowdary, N. H. (2019). An integrated approach to emotion recognition and gender classification. *Journal of Visual Communication and Image Representation*, 60, 339–345.
- Raza, H., Rathee, D., Zhou, S.-M., Cecotti, H., & Prasad, G. (2019). Covariate shift estimation based adaptive ensemble learning for handling non-stationarity in motor imagery related EEG-based brain-computer interface. *Neurocomputing*, 343, 154–166.
- Richhariya, B., Tanveer, M., & Rashid, A. H. (2020). Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE). *Biomedical Signal Processing and Control*, 59, 101903.
- Roffo, G., Melzi, S., & Cristani, M. (2015). In *Infinite feature selection* (pp. 4202–4210). IEEE.
- Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3, 42–55.
- Suykens, J. A. K., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural Processing Letters*, 9, 293–300.
- Taran, S., & Varun, B. (2019). Emotion recognition from single-channel EEG signals using a two-stage correlation and instantaneous frequency-based filtering method. *Computer Methods and Programs in Biomedicine*, 173, 157–165.
- Torres, E. N., Wei, W., Hua, N., & Chen, P. J. (2019). Customer emotions minute by minute: How guests experience different emotions within the same service environment. *International Journal of Hospitality Management*, 77, 128–138.
- Wang, X., Chen, X., & Cao, C. (2020). Human emotion recognition by optimally fusing facial expression and speech feature. *Signal Processing: Image Communication*, 84, 115831.
- Wang, D., & Shang, Y. (2013). Modeling physiological data with deep belief networks. *International Journal of Information & Education Technology*, 3, 505–511.
- Wu, Y., Wang, C., Bu, J., & Chen, C. (2016). Group sparse feature selection on local learning based clustering. *Neurocomputing*, 171, 1118–1130.
- Yang, S., Yin, Z., Wang, Y., Zhang, W., Wang, Y., & Zhang, J. (2019). Assessing cognitive mental workload via EEG signals and an ensemble deep learning classifier based on denoising autoencoders. *Computers in Biology and Medicine*, 109, 159–170.
- Yin, Z., & Zhang, J. (2014). Operator functional state classification using least-square support vector machine based recursive feature elimination technique. *Computer Methods and Programs in Biomedicine*, 113, 101–115.
- Yin, Z., Zhao, M., Wang, Y., Yang, J., & Zhang, J. (2017). Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer Methods and Programs in Biomedicine*, 140, 93–110.
- Yin, Z., Liu, L., Liu, L., Zhang, J., & Wang, Y. (2017). Dynamical recursive feature elimination technique for neurophysiological signal-based emotion recognition. *Cognition, Technology & Work*, 19, 667–685.
- Yin, Z., Zhao, M., Zhang, W., Wang, Y., Yang, J., & Zhang, J. (2019). Physiological-signal-based mental workload estimation via transfer dynamical autoencoders in a deep learning framework. *Neurocomputing*, 347, 212–229.
- Yogesh, C. K., Hariharan, M., Ngadiran, R., Adom, A. H., Yaacob, S., Berkai, C., & Polat, K. (2017). A new hybrid PSO assisted biogeography-based optimization for emotion and stress recognition from speech signal. *Expert Systems with Applications*, 69, 149–158.
- Yoon, H. J., & Chung, S. Y. (2013). EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm. *Computers Biology and Medicine*, 43, 2230–2237.
- Zhu, Y., Wang, S., & Ji, Q. (2014). Emotion recognition from users' EEG signals with the help of stimulus VIDEOS. In 2014 IEEE international conference on multimedia and expo (ICME) (pp. 1–6).