



# Facial expression distribution prediction based on surface electromyography

Xugang Xi<sup>a,\*</sup>, Yan Zhang<sup>a</sup>, Xian Hua<sup>b</sup>, Seyed M. Miran<sup>c</sup>, Yun-Bo Zhao<sup>d</sup>, Zhizeng Luo<sup>a</sup>

<sup>a</sup> School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China

<sup>b</sup> Jinhua People's Hospital, Jinhua 321000, China

<sup>c</sup> Biomedical Informatics Center, George Washington University, Washington, DC 20052, USA

<sup>d</sup> Department of Automation, Zhejiang University of Technology, Hangzhou 310023, China

## ARTICLE INFO

### Article history:

Received 7 July 2019

Revised 31 May 2020

Accepted 19 June 2020

Available online 29 June 2020

### Keywords:

Facial expression recognition

Emotion distribution learning

Surface electromyography

Principal component analysis

## ABSTRACT

Facial expression recognition plays an important role in research on human–computer interaction. The common facial expressions are mixtures of six basic emotions: anger, disgust, fear, happiness, sadness, and surprise. The current study, however, focused on a single basic emotion on the basis of physiological signals. We proposed emotion distribution learning (EDL) based on surface electromyography (sEMG) for predicting the intensities of basic emotions. We recorded the sEMG signals from the depressor supercillii, zygomaticus major, frontalis medial, and depressor anguli oris muscles. Six features were extracted in the frequency, time, time–frequency, and entropy domains. Principal component analysis (PCA) was used to select the most representative features for prediction. The key idea of EDL is to learn a function that maps the PCA-selected features to the facial expression distributions such that the special description degrees of all basic emotions for an emotion can be learned by EDL. Simultaneously, Jeffrey's divergence considered the relationship between different basic emotions. The performance of EDL was compared with that of multilabel learning based on PCA-selected features. Predicted results were measured by six indices, which could reflect the distance or similarity degree between distributions. We conducted an experiment on six different emotion distributions. Experimental results show that the EDL can predict the facial expression distribution more accurately than the other methods.

© 2020 Elsevier Ltd. All rights reserved.

## 1. Introduction

In recent years, facial expression recognition has become a popular research topic in human–computer interaction (HCI). Facial expression plays an important role in affective computing. Emotion can show the cognitive activity and psychopathology of a person. Surface electromyography (sEMG) can be recorded from motions of facial muscles, which can indicate the change in emotional state under internal and external stimuli (Jiang, Rahmani, Westerlund, Liljeberg, & Tenhunen, 2015). Except for sEMG signals, other approaches are mostly based on computer vision (Chen, Yang, & Wang, 2015). Facial sEMG-based methods outperform methods that are based on computer vision in many aspects, as discussed below.

Ekman, Friesen, and Ellsworth (1972) proposed six basic emotions: anger, disgust, fear, happiness, sadness, and surprise. In a

previous study, the accuracy of facial expression recognition was improved significantly in a single-emotion problem. However, emotion theory indicates that common facial expressions are mixtures of basic emotions (Plutchik, 1980). Hence, the degree of all basic emotions should be calculated. Label distribution learning (LDL) is suitable for facial expression distribution to match the fact that one facial expression is a mixture of different intensities of basic emotions (Geng, Yin, & Zhou, 2013). LDL, first proposed by Geng, Yin, and Zhou (2013), can address the importance of multiple labels. LDL has been applied to age estimation, head pose estimation, and other fields using computer vision (Geng, 2016). Emotion distribution learning (EDL), which is an LDL algorithm, performs more efficiently with the idea of the limited-memory quasi-Newton method L-BFGS than the standard LDL (Geng, Yin, & Zhou, 2013). A local low-rank structure is proposed to capture correlations of the local label on facial emotion distribution learning, and experiments demonstrate that the method can better deal with distribution recognition (Jia, Zheng, Li, Zhang, & Li, 2019).

Many single-label learning methods have been proposed and applied to facial expression recognition. Most methods for

\* Corresponding author.

E-mail addresses: [xixi@hdu.edu.cn](mailto:xixi@hdu.edu.cn) (X. Xi), [miran@gwu.edu](mailto:miran@gwu.edu) (S.M. Miran), [ybzhao@ieee.org](mailto:ybzhao@ieee.org) (Y.-B. Zhao), [luo@hdu.edu.cn](mailto:luo@hdu.edu.cn) (Z. Luo).

expression recognition are based on computer vision, such as images or videos. The support vector machine (SVM) (Song, Liu, & Wang, 2013) and hidden Markov model (Wang & Lien, 2009) were adopted to recognize a single basic emotion, and a high recognition rate was obtained. Computer vision-based methods are easy and inexpensive and can achieve acceptable results. However, the results of the recognition rate depend on the light level, camera resolution, and other external factors. Moreover, real facial expressions can be hidden or masked under certain circumstances, and the recognition result may be unaccompanied by the real result.

In HCI, physiological signals play an important role as sources of information in the reaction of human beings (Santamaria-Granados, Munoz-Organero, Ramirez-Gonzalez, Abdulhay, & Arunkumar, 2018). Physiological signals include electromyography, electroencephalogram, and galvanic skin response (Dey, Ashour, Mohamed, & Nguyen, 2019). sEMG has been widely applied in gesture recognition, sign language recognition, and motion recognition of upper limbs (Du, Lin, Shyu, & Chen, 2010; Pancholi & Joshi, 2019). Moreover, sEMG has been found to be useful for facial expression recognition (Kehri, Ingle, Patil, & Awale, 2019). sEMG can truly reflect neuromuscular activities, which are controlled by the autonomic nervous system (Becker, von Werder, Lassek, & Disselhorst-Klug, 2019). Schmidt and Cohn (2001) found that a smile is related to the activation of the zygomaticus major muscle in 95 individuals. Hamed, Salleh, and Swee (2011) applied the sEMG of frontalis and right and left temporalis to recognize a smile, smile with right/left side, and anger, and they obtained a recognition rate of 80.4% by SVM. Chen, Yang, and Wang (2015) achieved a recognition rate of up to 95.56% on the basis of the sEMG of eyebrow expression. sEMG-based methods have advantages compared with computer vision methods. Without imposing limits on the external environment, sEMG-based methods are not only unaffected by head movements but also provide nonvisual and unbiased information (Chowdhury, Reaz, Ali, Bakar, Chellappan, & Chang, 2013; Mesa, Rubio, Tubia, De No, & Diaz, 2014). Principal component analysis (PCA) with Karhunen-Loeve transform has been used in signal and image processing and applied to recognize upper limb motions on the basis of sEMG (Veer & Vig, 2018). PCA can not only overcome the limitation of dimensionality (Loconsole, Cascarano, Brunetti, Trotta, Losavio, & Bevilacqua, 2019) but can also identify the most representative features for classification.

Facial expressions, which are mixtures of basic emotions, have not yet been explored on the basis of physiological signals in previous studies. Therefore, our study aims to obtain the most representative features and the best distribution prediction algorithm for an emotion simultaneously on the basis of sEMG. To that end, four electrodes were placed on the depressor supercillii, zygomaticus major, frontalis medial, and depressor anguli oris muscles of the subjects. For each channel, six features were extracted in the frequency, time, time–frequency, and entropy domains from each sEMG signal, and PCA was applied to select the most representative features for the prediction. Then, EDL was applied to the predicted distribution with the intensities of basic emotions for an expression. Our experiment with 12 subjects showed that EDL with PCA-selected features has a more accurate distribution prediction than other multilabel learning (MLL) algorithms.

## 2. Materials

### 2.1. Participants

Twelve subjects (eight males and four females: age, 24–26 years; height, 155–180 cm; weight 40–75 kg) participated in the

experiment. All subjects were healthy and had no family history of facial neuromuscular and nervous disorders. They provided written informed consent approved by an institutional review board.

### 2.2. Expressions and the muscles

Each subject wore sensors on four facial muscles (depressor supercillii, zygomaticus major, frontalis medial, and depressor anguli oris). The muscles were tightly related to six basic emotions (Edmonds, Couture, Paloheimo, & Rigor, 1988). We recorded the sEMG signals by Trigno™ Wireless EMG (Delsys Inc, Natick, MA, USA) with a sampling rate of 1000 Hz. The left and right faces were assumed to be symmetrical. Thus, one side of the face was selected as the recorded point. The locations of the sensors are shown in Fig. 1.

### 2.3. Experimental procedure

The experiment was divided into three steps: preparation, task, and statistics. In the preparation step, all subjects were required to keep their faces clean and smooth to reduce interference from other noises. We introduced the procedures and the equipment to the subjects. We conducted a simple tutorial for the subjects. When we showed pictures that can trigger mood change, the subjects showed facial expressions.

In the task step, 10 different pictures that trigger different facial expressions from the Radboud Faces Database were prepared as a set (Langner, Dotsch, Bijlstra, Wigboldus, Hawk, & Van Knippenberg, 2010). As previously stated, the common facial expressions are mixtures of six basic emotions: anger, disgust, fear, happiness, sadness, and surprise. At the end of each expression, the subjects were required to make a real-time self-assessment, which involved scoring each basic emotion. The sequence of photos was random, and each photo appeared 10 times. Each picture was shown for 4 s, each emotion lasted for 2 s, and the sampling frequency of the equipment was set as 1000 Hz. Subjects were asked to rest for 10 min after finishing five emotions to avoid muscle fatigue. This experiment was repeated the next day. Fig. 2 shows the emotion and emotion distribution.

In the last step, statistics from the subjects were recorded. The records of expressions from each subject reacting to each picture were normalized.

### 2.4. Data normalization

The dataset was randomly divided into two subsets: a training set, which contained 70% of the data, and a test set, which contained the remaining data. We took 10-fold cross validation for the training set. As the sEMG signals were recorded from four

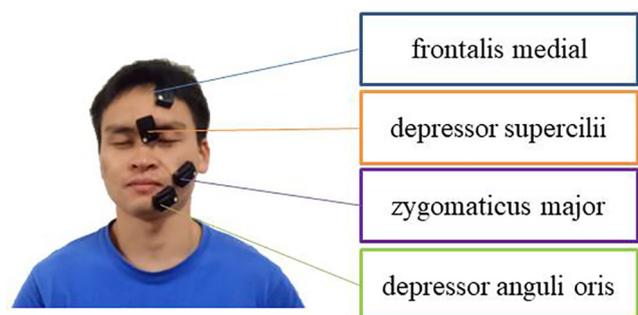


Fig. 1. Locations of sEMG sensors.

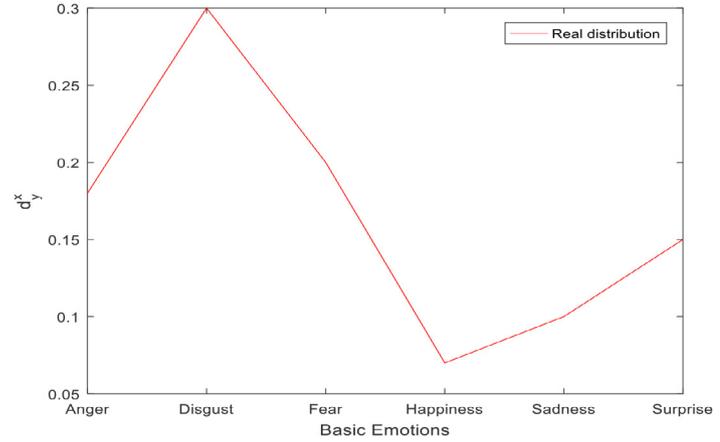


Fig. 2. Emotion and emotion distribution.

channels, the data were normalized within each channel using the following formula:

$$ND(i) = \text{Amp}(i)D(i) + \text{Off}(i), \tag{1}$$

where  $ND(i)$  denotes the  $i$ th normalized sEMG and  $D(i)$  represents the  $i$ th raw sEMG.

$\text{Amp}(i)$  and  $\text{Off}(i)$  are the amplitude and offset of the  $i$ th channel, respectively, as follows:

$$\text{Amp}(i) = \frac{x_k - x_l}{\text{Max}(i) - \text{Min}(i)}, \tag{2}$$

$$\text{Off}(i) = x_k - \text{Amp}(i)\text{Max}(i), \tag{3}$$

where  $\text{Max}(i)$  and  $\text{Min}(i)$  are the maximum and minimum values of the  $i$ th channel, respectively.  $x_k$  and  $x_l$  denote the upper and the lower limits for values of the  $i$ th channel, respectively.

### 3. Method

#### 3.1. Feature extraction

sEMG features were computed with 2 s epochs. Six features were extracted in the time, frequency, time–frequency, and entropy domains, which present the characterization of the signal in different aspects (Sun, Ovsjanikov, & Guibas, 2009).

Six different features were selected for comparison: Wilson amplitude (WAMP), fuzzy entropy (FE), energy of wavelet packet coefficient (EWP), autoregressive coefficient (AR), mean power frequency (MPF), and mean frequency (MF), as shown in Table 1. These features were proven to be effective (Xi, Tang, & Luo, 2018).

The vector space of all the considered features has a total of 13 dimensions. The PCA approach was applied to select the most representative of the principal components. It was decided to consider the most significant components that account for at least 90% of the variance in data, and thus the first 7 components out of 13 have been included in Fig. 3.

Table 1  
Feature extractions.

ID	Feature	Dimension	Domain
1	WAMP	1	Time domain
2	AR	3	Time domain
3	MF	1	Frequency domain
4	MPF	1	Frequency domain
5	EWP	6	Time-frequency domain
6	FE	1	Entropy domain

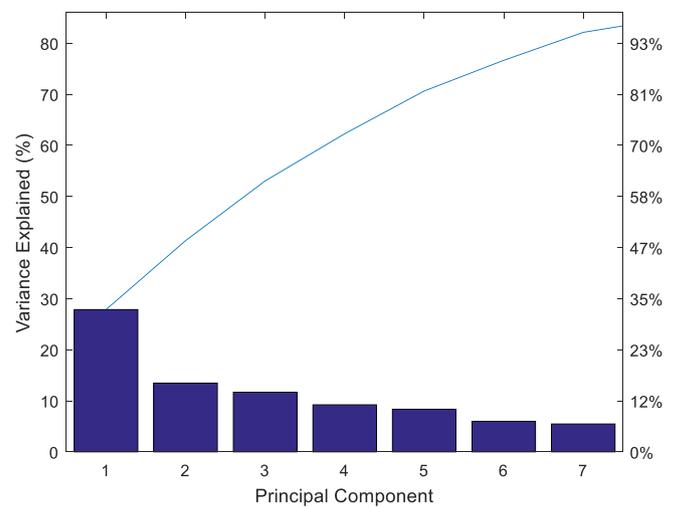


Fig. 3. The contribution rate of the first seven principal components.

#### 3.2. EDL

EDL is a new machine learning paradigm. Traditional single-label learning and multilabel learning can be regarded as special cases of the paradigm (Geng, Yin, & Zhou, 2013). EDL uses description degree  $d_x^y \in [0, 1]$  to present the relation between basic emotion  $y$  and expression  $x$ . The higher the number is, the higher the intensity of basic emotion  $y$ . The description degrees of all basic emotions add up to 1. The description degree is similar to the probability distribution. Therefore, we call it emotional distribution.

For a training set  $T = \{(\mathbf{x}_1, M_1), (\mathbf{x}_2, M_2), \dots, (\mathbf{x}_n, M_n)\}$ ,  $\mathbf{x}_i \in \mathbf{X}$  presents an emotion instance, and  $M_i = \{d_{x_i}^{y_1}, d_{x_i}^{y_2}, \dots, d_{x_i}^{y_c}\}$  is the distribution of the sample emotion  $\mathbf{x}_i$  and label value  $y$ .  $d_{x_i}^y$  is represented by the form of conditional probability. The goal of EDL is to learn functions  $d_{x_i}^y = p(y|x_i)$  for a training set  $T$ , where  $x_i \in X, y \in Y$ .  $p(y|x_i)$  is assumed to be a parametric model  $p(y|x_i; \theta)$ , where  $\theta$  is a parameter vector.

Kullback-Leibler divergence was used to measure the distance between distributions  $M_a^j$  and  $M_b^j$ , defined as follows:

$$D_J(M_a || M_b) = \sum_j (M_a^j - M_b^j) \log \frac{M_a^j}{M_b^j}, \tag{4}$$

where  $M_a^j$  and  $M_b^j$  are the  $j$ th elements of distributions  $M_a$  and  $M_b$ , respectively.  $\parallel$  indicates that the relationship of the two distributions is balanced. It also means  $D_J(M_a||M_b) = D_J(M_b||M_a)$ .

Eq. (4) only considers the same basic emotion of two distributions; thus, the sum of distances between the same superscripts of  $M_a$  and  $M_b$  are calculated. The relationship between different basic emotions should be considered. Different basic emotions are closely related to muscular activity. Some basic emotions, such as happiness and disgust, cannot appear simultaneously. Thus, weighted Jeffrey's divergence is proposed using the relationship between basic emotions.

$$D_{wJ}(M_a||M_b) = \sum_{j,k} w_{jk} (M_a^j - M_b^j) \log \frac{M_a^j}{M_b^j}, \tag{5}$$

where  $w_{jk}$  is the weight between the  $j$ th basic emotion and the  $k$ th basic emotion of the distribution, which can be calculated as follows:

$$w_{jk} = \begin{cases} \frac{1}{\Lambda_j} (\rho_{jk})^\eta & |\rho_{jk}| \geq \varepsilon \\ 0 & \text{other} \end{cases}, \tag{6}$$

where  $\rho_{jk} = \frac{\sum_i (d_{x_i}^{y_j} - \bar{d}_{x_i}^{y_j})(d_{x_i}^{y_k} - \bar{d}_{x_i}^{y_k})}{\sqrt{\sum_i (d_{x_i}^{y_j} - \bar{d}_{x_i}^{y_j})^2} \sqrt{\sum_i (d_{x_i}^{y_k} - \bar{d}_{x_i}^{y_k})^2}}$  is the correlation coefficient between the  $j$ th basic emotion and the  $k$ th basic emotion

and  $\Lambda_j = \sum_k (\rho_{jk})^\eta$  presents the normalization factor, which can ensure  $\sum_k \omega_{jk} = 1$ .  $\eta$  can control the correlation coefficient, which is a positive odd number.  $\varepsilon$  is a threshold. If the value of the correlation coefficient is smaller than the value of  $\varepsilon$ , the two basic emotions have no relationship. The best values of  $\eta$  and  $\varepsilon$  are set as 5 and 0.25, respectively (Geng, Yin, & Zhou, 2013). This was also confirmed by our experiment.

Finally, the target function  $\theta^*$  is defined as follows:

$$\begin{aligned} \theta^* = \operatorname{argmin}_\theta \sum_i D_{\omega J}(M_i||\hat{M}_i) - \xi_1 \frac{1}{n} \sum_k \|\theta_k - \bar{\theta}\|_2^2 \\ + \frac{1}{2} \xi_2 \sum_{k,r} \theta_{kr}^2 = \operatorname{argmin}_\theta \sum_{i,j,k} \omega_{ij} (d_{x_i}^{y_j} - p(y_k|x_i; \theta)) \\ \times (\ln d_{x_i}^{y_j} - \ln p(y_k|x_i; \theta)) - \xi_1 \frac{1}{n} \sum_k \|\theta_k - \bar{\theta}\|_2^2 + \frac{1}{2} \xi_2 \sum_{k,r} \theta_{kr}^2, \end{aligned} \tag{7}$$

where  $M_i$  is the distribution of the  $i$ th sample and  $\hat{M}_i$  is the predicted distribution of the  $i$ th sample by using  $p(y|x_i; \theta)$ . The second item is a regularized item used to emphasize the important emotions. The third term is another regularized item for preventing unstable output.  $\xi_1$  and  $\xi_2$  are balance factors.  $\xi_1$  and  $\xi_2$  for EDL are set as 0.0002 and 0.002 (Geng, Yin, & Zhou, 2013).

Assuming that  $p(y|x_i; \theta)$  is a maximum entropy model,

$$p(y_k|x_i; \theta) = \frac{1}{Z_i} \exp\left(\sum_r \theta_{kr} x_i^r\right), \tag{8}$$

where  $Z_i = \sum_k \exp(\sum_r \theta_{kr} x_i^r)$  is the normalizing factor,  $x_i^r$  is the  $r$ th feature of  $x_i$ , and  $\theta_{kr}$  is the element of the row  $k$  column  $r$  in parameter  $\theta$ . The target function of  $\theta$  is  $T(\theta)$ .

$$\begin{aligned} T(\theta) = \sum_i Z_i + \sum_{i,j,k} \omega_{jk} \left[ \frac{1}{Z_i} \exp\left(\sum_r \theta_{kr} x_i^r\right) \right. \\ \left. \left( \sum_r \theta_{kr} x_i^r - \ln Z_i - \ln d_{x_i}^{y_j} \right) - d_{x_i}^{y_j} \sum_r \theta_{kr} x_i^r \right] \\ - \xi_1 \frac{1}{n} \sum_k \|\theta_k - \bar{\theta}\|_2^2 + \frac{1}{2} \xi_2 \sum_{k,r} \theta_{kr}^2, \end{aligned} \tag{9}$$

The limited-memory quasi-Newton method L-BFGS is used in the minimization of function  $T(\theta)$  to increase computing efficiency. The computation of L-BFGS is related to the first-order gradient of  $T(\theta)$ , which can be achieved by

$$\begin{aligned} \frac{\partial T(\theta)}{\partial \theta_{kr}} = \sum_{i,j,k} \omega_{jk} \left[ p_{ik} x_i^k (1 - p_{ik}) \left( \sum_r \theta_{kr} x_i^r - \ln Z_i - \ln d_{x_i}^{y_j} + 1 \right) \right. \\ \left. - \sum_i x_i^r (1 - p_{ik}) - \xi_1 \frac{1}{n} \left[ (\theta_{kr} - \bar{\theta}_r) - \frac{1}{c} \sum_k (\theta_{kr} - \bar{\theta}_r) \right] \right. \\ \left. + \xi_2 \sum_{k,r} \theta_{kr}, \right] \end{aligned} \tag{10}$$

where  $p_{ik} = \frac{1}{Z_i} \exp\left(\sum_r \theta_{kr} x_i^r\right)$ .

### 4. Experiment

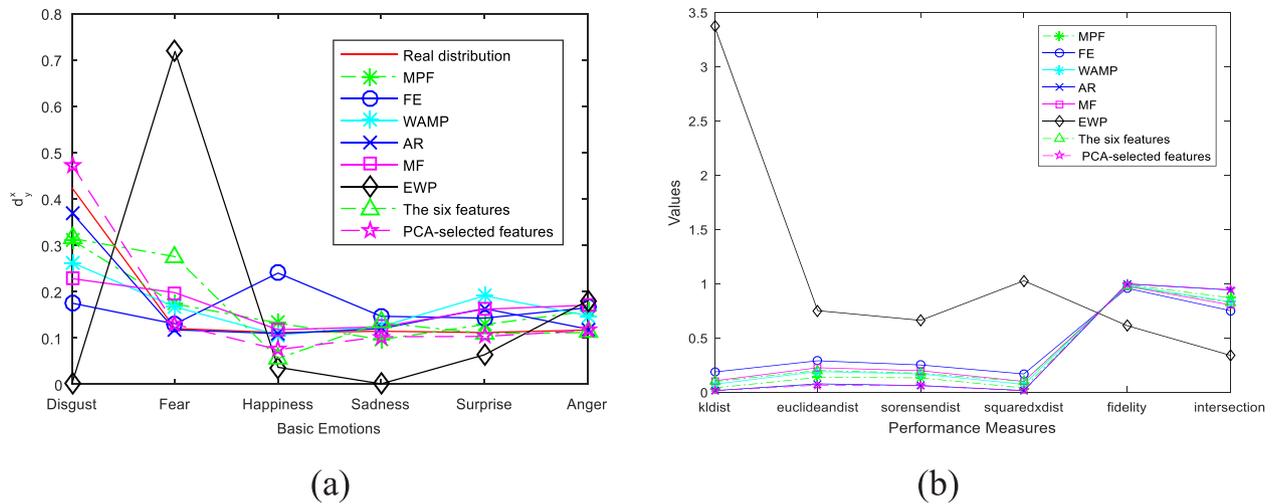
PCA was adopted to select the most representative features in this study. The different features were applied to five different prediction algorithms to demonstrate the effectiveness of the PCA-selected features. EDL is a prediction algorithm. The other distribution prediction algorithms are algorithm adaptation k-nearest neighbor (AA-KNN), problem transformation Bayes (PT-Bayes), problem transformation support vector machine (PT-SVM), and algorithm adaptation backpropagation algorithm (AABP). The basic idea of PT-Bayes and PT-SVM algorithms is to transform the label distribution into an MLL. Moreover, AA-KNN and AABP algorithms, which are natural extensions, are adopted to address label distribution (Geng, 2016). In this manner, we can find the best distribution prediction algorithm. In EDL, parameters  $\eta$ ,  $\varepsilon$ ,  $\xi_1$ , and  $\xi_2$  are set as 5, 0.25, 0.0002, and 0.002, respectively. For each compared algorithm, several parameters must be set to obtain the best performance. In AA-KNN,  $k$  is set to 6. In PT-SVM, a linear kernel is used. The number of hidden-layer neurons is 50 in the AABP. For all datasets, tenfold cross validation is performed in each algorithm.

#### 4.1. Measurement

The evaluation indices are divided into distance and similarity measures, which are used to evaluate the distance or similarity degree between distributions. The first four measures are distance measures, and the last two measures are similarity measures Table 2. In distance measures, the values of indices are lowered with the enhanced effect of the algorithm. By contrast, as the values of the indices of similarity measures increase, the prediction of the algorithm improves.

**Table 2**  
Measures.

	Measures	Equation
Distance measures	KL-div↓	$dist_1(P, Q) = \sum_{j=1}^c P_j \ln \frac{P_j}{Q_j}$
	Euclidean↓	$dist_2(P, Q) = \sqrt{\sum_{j=1}^c (P_j - Q_j)^2}$
	Sorensen↓	$dist_3(P, Q) = \frac{\sum_{j=1}^c  P_j - Q_j }{\sum_{j=1}^c (P_j + Q_j)}$
	Squared $\chi^2$ ↓	$dist_4(P, Q) = \sum_{j=1}^c \frac{(P_j - Q_j)^2}{P_j + Q_j}$
Similarity measures	Fidelity↑	$sim_1(P, Q) = \sum_{j=1}^c \sqrt{P_j Q_j}$
	Intersection↑	$sim_2(P, Q) = \sum_{j=1}^c \min(P_j, Q_j)$



**Fig. 4.** Results of distribution prediction for emotion in EDL. (a) Comparison of real and predicted distributions by using different features in EDL; (b) values of different measurements using different features in EDL.

4.2. Components selection

PCA was used to select the most representative components as inputs to the distribution prediction algorithms. All results were measured and compared in the experiment. Fig. 4 shows the predicted distributions by different features in EDL and the results of the evaluation index. The indices of distance measures are low, and the indices of similarity measures are high for the PCA-selected features under the same indicators. In Figs. 5–8, the experimental results show that the PCA-selected components obtain well-predicted distributions in the MLL prediction algorithms. In the AABP, the algorithm fails, but the PCA-selected features obtain the best results in Fig. 8. In general, the PCA-selected components exhibit the best performance on five different prediction algorithms.

4.3. Results of predicted distribution

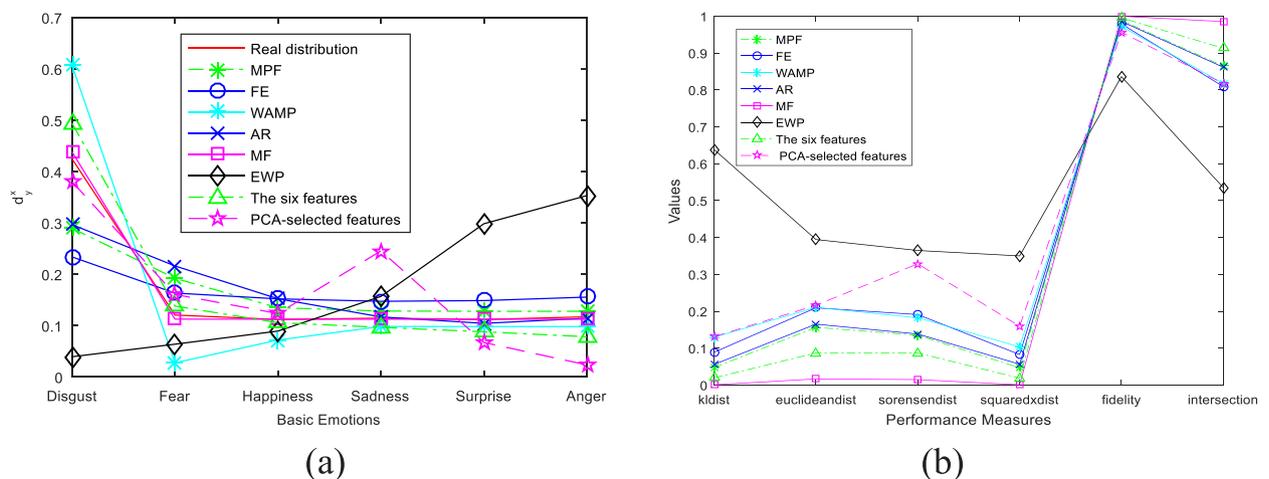
To verify the effectiveness of the EDL algorithm, the four other MLL algorithms were used for the comparison.

Figs. 9–14 show the comparison results of EDL with four MLL algorithms. EDL performs best on all measures. Figs. 9–14 present six different representative emotions (a, the comparison of real and predicted distributions by using different algorithms; b, the values

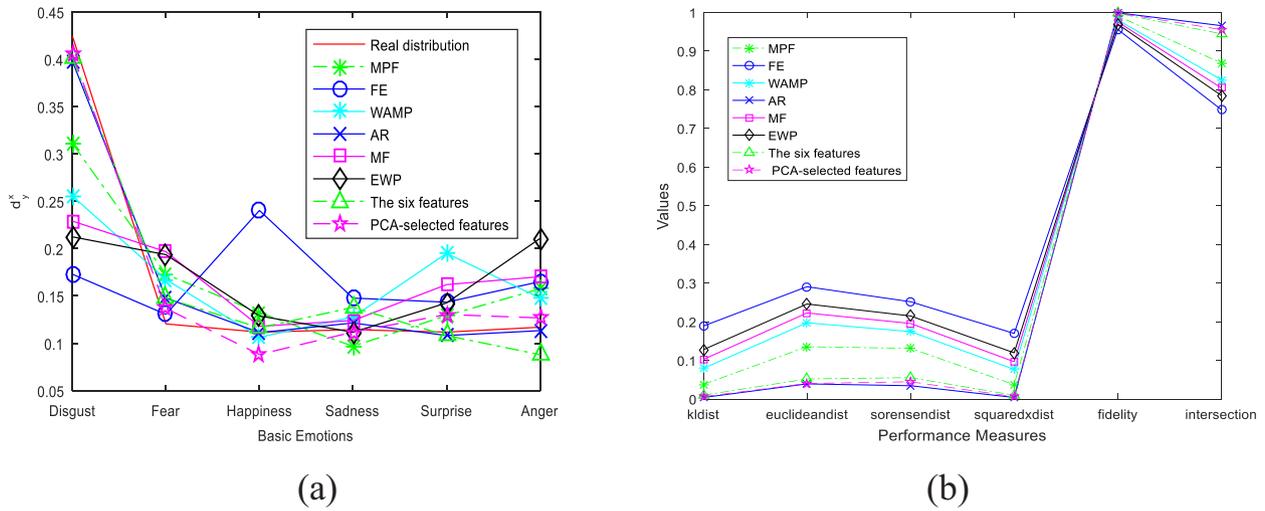
of different measurements for different features by using various algorithms). Each of the six typical emotions represents a different meaning with different intensities of basic emotions. EDL can not only match further complex cases, but can also perform better than the other four MLL algorithms due to the computation of the intensities of the basic emotions. Happiness and disgust cannot appear with the same intensities in real distributions. EDL accurately predicts the distribution and shows the relationship between different basic emotions. The results show that AABP performs worst in all tests. Table 3 also indicates that the rankings of the five algorithms on six measures are almost consistent. Based on six different distributions, the five algorithms can be ranked as EDL ≥ PT-SVM ≈ PT-Bayes > AA-kNN > AABP.

5. Discussion and conclusion

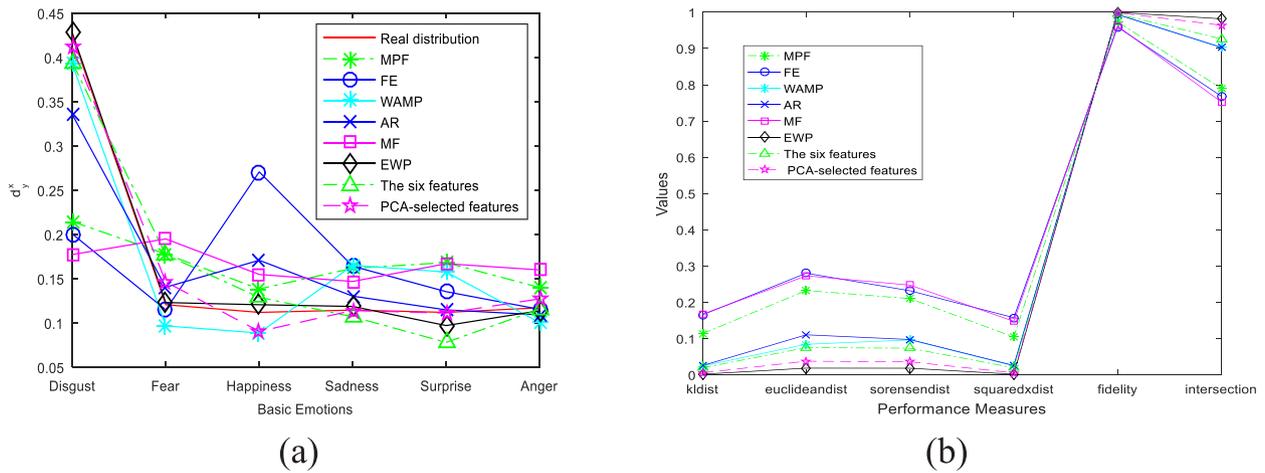
Results. Figs. 4–14, show some typical examples that demonstrate the superiority of EDL. In Fig. 4, the value of the description degree of disgust is the highest, and the other description degrees are low. However, in single-emotion learning and MLL, disgust may be the only effective expression with an information loss of all five other emotions. In Fig. 14, the value of happiness is consistent with surprise but significantly higher than that of the others. MLL may lose useful information about the intensities of the relevant



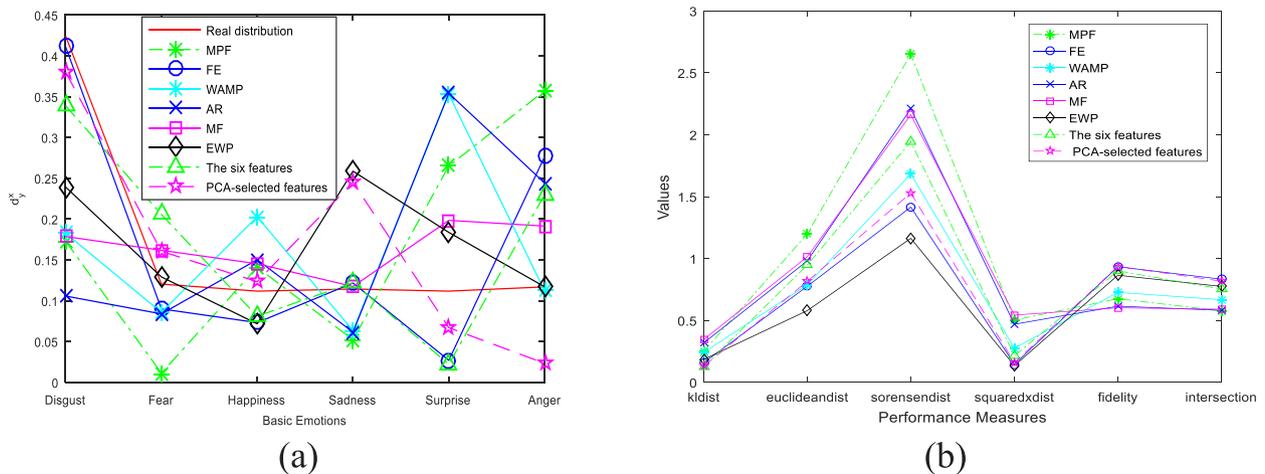
**Fig. 5.** Results of distribution prediction for emotion in AA-KNN. (a) Comparison of real and predicted distributions by using different features in AA-KNN; (b) values of different measurements using different features in AA-KNN.



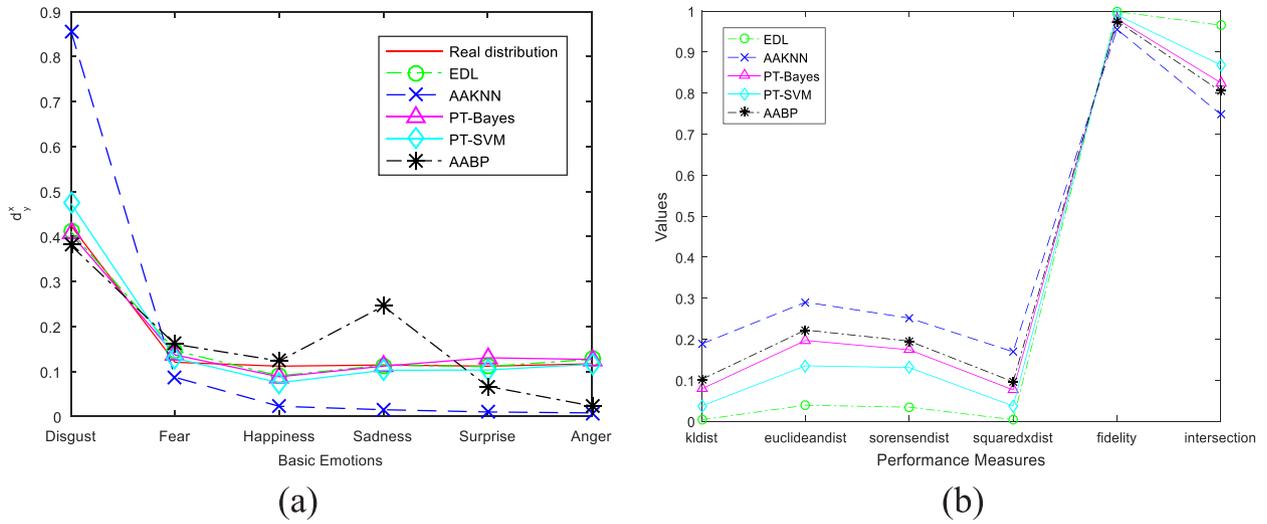
**Fig. 6.** Results of distribution prediction for the emotion in PT-Bayes. (a) Comparison of real and predicted distributions by using different features in PT-Bayes; (b) values of different measurements using different features in PT-Bayes.



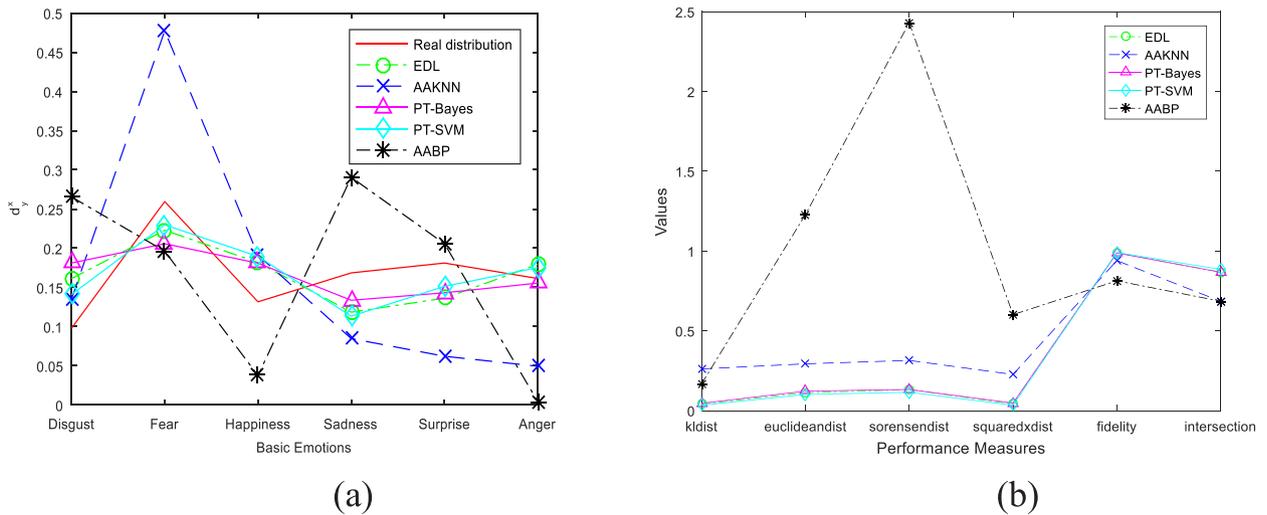
**Fig. 7.** Results of distribution prediction for the emotion in PT-SVM. (a) Comparison of real and predicted distributions by using different features in PT-SVM; (b) values of different measurements using different features in PT-SVM.



**Fig. 8.** Results of distribution prediction for emotion in AABP. (a) Comparison of real and predicted distributions using different features in AABP; (b) values of different measurements using different features in AABP.



**Fig. 9.** Results of distribution prediction for test 1 (the intensity of disgust is the highest). (a) Comparison of real and predicted distributions by using different algorithms; (b) values of different measurements using different algorithms.



**Fig. 10.** Results of distribution prediction for test 2 (the intensity of fear is the highest). (a) Comparison of real and predicted distributions by using different algorithms; (b) values of different measurements using different algorithms.

emotions, and single-emotion learning cannot deal with the problem. Therefore, EDL is an appropriate method for predicting the distribution.

In Figs. 4–8, PCA is an efficient method for obtaining a transformation matrix to select improved features. The data decrease from high-dimensional to low-dimensional space, and similar features are merged for the variance. With reduced data, the number of features decreases, which is beneficial in preventing the occurrence of overfitting. In Figs. 9–14, EDL achieves the best performance on the basis of the six different measures compared with MLL. The reason is that EDL directly minimizes the distance and enlarges the similarity between the real and predicted distributions. PT-Bays and PT-SVM obtain relatively good performance. The ideas of PT-Bays and PT-SVM are appropriate for the datasets, which are the Gaussian assumption for PT-Bays and decomposed distribution by weighted resampling for PT-SVM. Additionally, keeping the labeling structure of each example of AA-KNN is inappropriate for the datasets. AABP performs worst due to overfitting because it needs additional examples with several parameters to learn.

Compared to the related work listed in Table 4, our proposed framework revealed at least two comparative merits.

- 1) Compared with images (Zhou, Xue, & Geng, 2015), we find that the predicted distributions are more similar to real distributions. Simultaneously, standardized data databases are more persuasive for showing the performance of recognition methods (Jia, Zheng, Li, Zhang, & Li, 2019). The performance of the proposed method using sEMG is not worse than that of the methods using images. However, someone tries to hide real emotion, which leads to misleading results of mental state assessment with exceptional circumstances using images.
- 2) Compared with existing related work using biological electrical signals, the proposed method can address the importance of multiple labels. Current single-label methods have achieved good results. For example, Mithbavkar and Shah (2019) achieved good accuracy of up to 97% using the EMG from AUBT, which contains a single person in four diverse

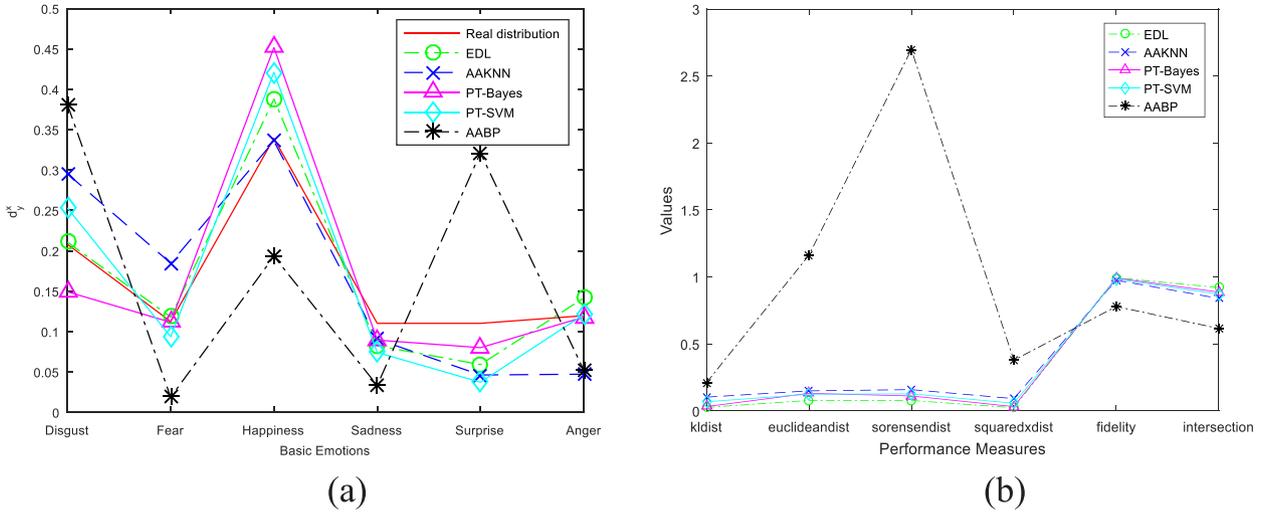


Fig. 11. Results of distribution prediction for test 3 (the intensity of happiness is the highest). (a) Comparison of real and predicted distributions by using different algorithms; (b) values of different measurements using different algorithms.

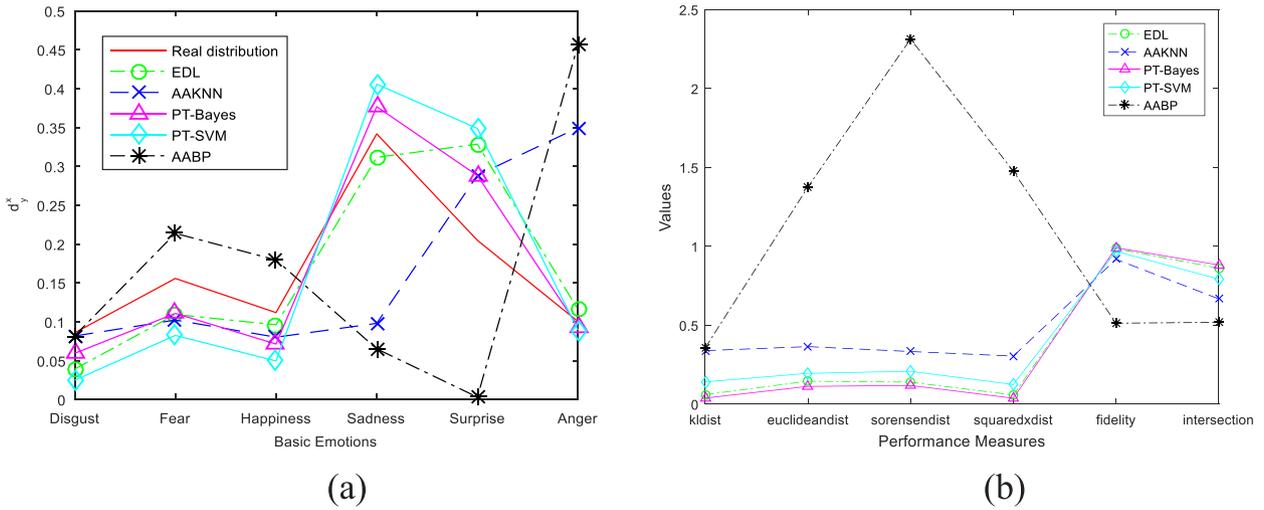


Fig. 12. Results of distribution prediction for test 4 (the intensity of sadness is the highest). (a) Comparison of real and predicted distributions by using different algorithms; (b) values of different measurements using different algorithms.

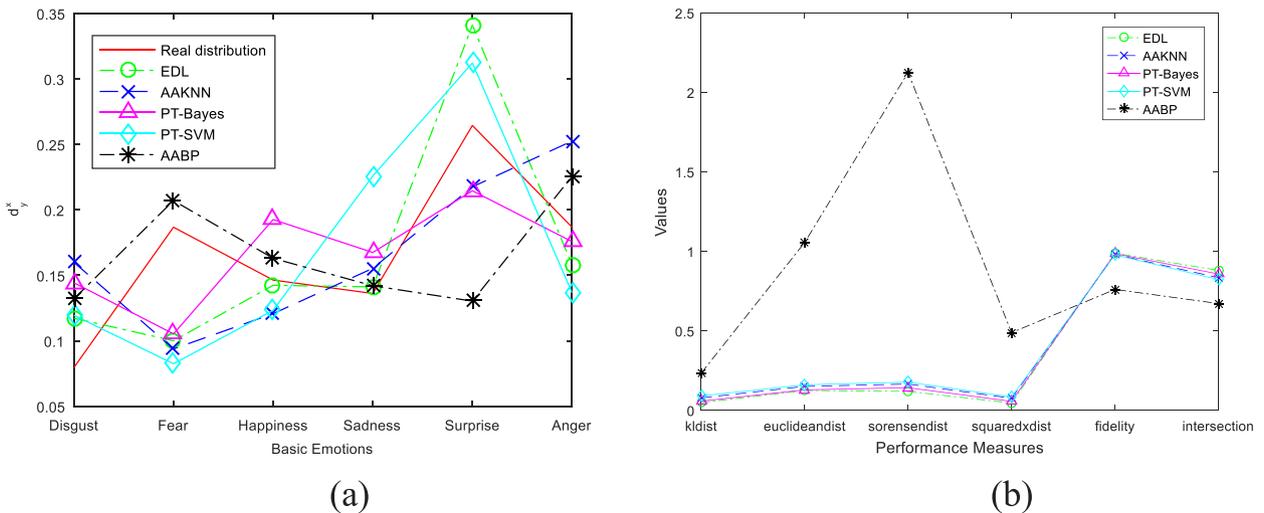
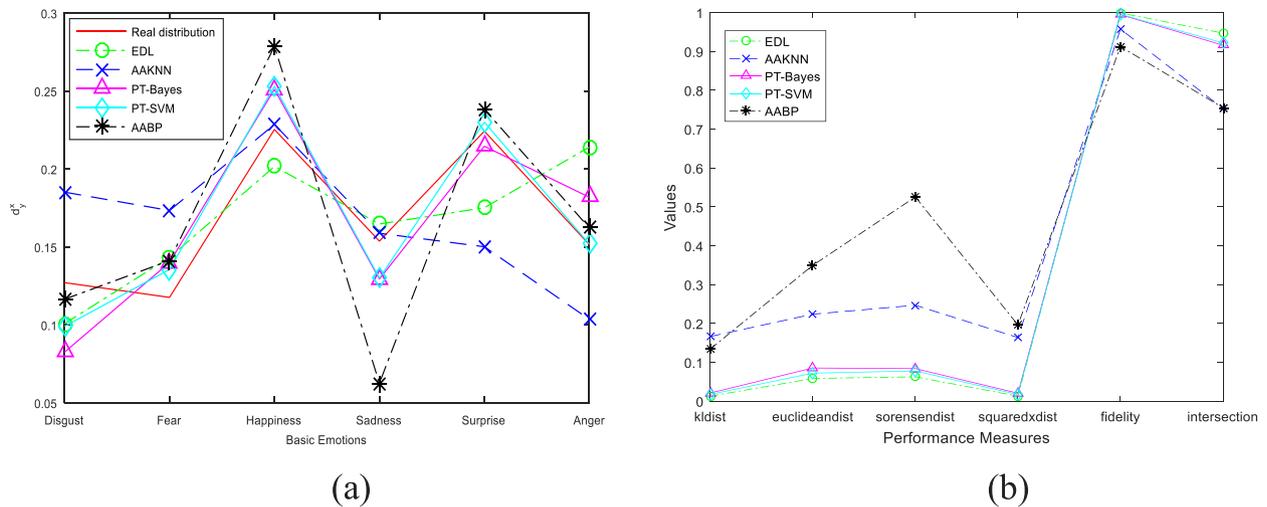


Fig. 13. Results of distribution prediction for test 5 (the intensity of surprise is the highest). (a) Comparison of real and predicted distributions by using different algorithms; (b) values of different measurements using different algorithms.



**Fig. 14.** Results of distribution prediction for test 6 (the intensities of disgust and surprise are close and high). (a) Comparison of real and predicted distributions by using different algorithms; (b) values of different measurements using different algorithms.

**Table 3**  
The result of predicted distributions using different algorithms.

Emotions	Prediction Algorithms	Measures					
		kldist	euclideanist	sorensendist	squaredxdist	fidelity	intersection
test 1	EDL	<b>0.003</b>	<b>0.039</b>	<b>0.034</b>	<b>0.003</b>	<b>0.999</b>	<b>0.965</b>
	AA-KNN	0.188	0.290	0.251	0.169	0.955	0.748
	PT-Bayes	0.079	0.197	0.174	0.076	0.999	0.837
	PT-SVM	0.037	0.135	0.131	0.037	0.990	0.868
	AABP	0.102	0.222	0.525	0.096	0.975	0.804
test 2	EDL	0.042	0.112	0.131	0.042	0.989	0.868
	AA-KNN	0.261	0.293	0.314	0.226	0.940	0.685
	PT-Bayes	0.046	0.122	0.133	0.047	0.988	0.867
	PT-SVM	<b>0.034</b>	<b>0.101</b>	<b>0.115</b>	<b>0.034</b>	<b>0.991</b>	<b>0.885</b>
	AABP	0.167	1.228	2.430	0.601	0.813	0.685
test 3	EDL	<b>0.027</b>	<b>0.079</b>	<b>0.079</b>	<b>0.025</b>	<b>0.994</b>	<b>0.921</b>
	AA-KNN	0.103	0.150	0.160	0.092	0.976	0.840
	PT-Bayes	0.033	0.132	0.112	0.033	0.992	0.888
	PT-SVM	0.069	0.125	0.128	0.058	0.985	0.872
	AABP	0.211	1.162	2.698	0.379	0.780	0.616
test 4	EDL	0.060	0.145	0.139	0.058	0.985	0.861
	AA-KNN	0.339	0.364	0.333	0.304	0.918	0.667
	PT-Bayes	<b>0.038</b>	<b>0.112</b>	<b>0.118</b>	<b>0.037</b>	<b>0.991</b>	<b>0.882</b>
	PT-SVM	0.141	0.195	0.208	0.124	0.968	0.792
	AABP	0.356	1.376	3.313	1.475	0.513	0.518
test 5	EDL	<b>0.049</b>	<b>0.125</b>	<b>0.119</b>	<b>0.046</b>	<b>0.988</b>	<b>0.881</b>
	AA-KNN	0.078	0.151	0.165	0.076	0.981	0.835
	PT-Bayes	0.058	0.129	0.142	0.057	0.986	0.858
	PT-SVM	0.091	0.161	0.177	0.084	0.978	0.823
	AABP	0.234	1.056	2.125	0.489	0.761	0.672
test 6	LDL	<b>0.013</b>	<b>0.059</b>	<b>0.063</b>	<b>0.013</b>	<b>0.999</b>	<b>0.947</b>
	AA-KNN	0.167	0.224	0.247	0.164	0.958	0.753
	PT-Bayes	0.021	0.086	0.084	0.021	0.995	0.916
	PT-SVM	0.018	0.072	0.078	0.017	0.996	0.922
	AABP	0.134	0.350	0.526	0.196	0.913	0.755

The bold values indicate the best performance.

expressive states, namely, angry, joy, pleasure, and sad, under musical circumstances. EEG is also applied to emotion recognition (Chen, Zhang, Mao, Huang, Jiang, & Zhang, 2019), and the accuracy is up to 85.57%. However, bioelectricity databases are not suitable for distribution studies.

In this study, the facial sEMG-based method was proposed to predict the intensities of all basic emotions in an expression. The six features were extracted in different domains (frequency, time,

time–frequency, and entropy domains), and PCA was applied to select the most representative features. The objective of this study was to predict the distribution with the intensities of the basic emotions. Simultaneously, Jeffrey’s divergence considered the relationship between different basic emotions. The effectiveness of EDL with the PCA-selected features was verified experimentally. In addition, the performances of EDL and other MLL models were compared on the basis of the PCA-selected features. The results show that the predicted results of EDL are better than those of

**Table 4**  
List of some typical related work.

Article	Signal	Label	Method	Databases	Recognition
Zhou, Xue, and Geng (2015)	Image	Distribution	EDL	s-JAFFE	Kldist 0.0957; Euclideanist 0.1002; Sorensendist0.0339; Squaredxdist0.034; Fidelity 0.8998; Intersection 0.9914
				BU 3DFE	Kldist 0.1055; Euclideanist 0.1061; Sorensendist0.0402; Squaredxdist0.04; Fidelity 0.8939; Intersection 0.9898
Jia, Zheng, Li, Zhang, and Li (2019)	Image	Distribution	EDL-LRL	s-JAFFE	Kldist 0.0806; Euclideanist 0.3008; Sorensendist0.6134; Squaredxdist0.0361; Fidelity 0.966; Intersection 0.897
				BU 3DFE	Kldist 0.0951; Euclideanist 0.3556; Sorensendist0.7463; Squaredxdist0.0694; Fidelity 0.9626; Intersection 0.8686
Jain, Shamsolmoali, and Sehdev (2019)	Image	Single-label	Extended DNN	JAFFE	Accuracy 95.23%
Renda, Barsacchi, and Bechini (2019)	Image	Single-label	Emsemble CNNs	FER2013	The best strategies the Preprocessing Strategy and Pretraining Strategy, accuracy above 72%
Zangeneh, Rahmati, and Mohsenzadeh (2020)	Image	Single-label	DCNNs	Training dataset:FERET Evaluation datasets:LFW, MBGC,and FERET	Accuracy: FERET(6 × 6) 81.4%; FERET(12 × 12) 92.1%; FERET(24 × 24) 96.7%. LFW(8 × 8):76.3% MBGC(12 × 12) 68.64%
Mithbavkar and Shah (2019)	EMG	Single-label	NARX	AUBT	Accuracy 91% to 97%
Chen, Zhang, Mao, Huang, Jiang, and Zhang (2019)	EEG	Single-label	deep CNN	DEAP	Accuracy up to 85.57%
Our Work	sEMG	Distribution	EDL	Obtained by the experiment	Kldist 0.0323; Euclideanist 0.0932; Sorensendist0.0942; Squaredxdist0.0312; Fidelity 0.9923; Intersection 0.9072

other methods. HCI has good prospects in the future. Thus, our future work will be devoted to developing a wearable device that can understand the mentality of people.

#### CRedit authorship contribution statement

**Xugang Xi:** Conceptualization, Funding acquisition, Writing - review & editing. **Yan Zhang:** Methodology, Data curation, Writing - original draft. **Xian Hua:** Resources, Investigation. **Seyed M. Miran:** Writing - review & editing. **Yun-Bo Zhao:** Software, Validation. **Zhizeng Luo:** Supervision.

#### 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 was supported by the National Natural Science Foundation of China (Nos. 61971169, 61671197, 61673350 and 60903084), Zhejiang Public Welfare Technology Research (No. LGF18F010006) and Jinhua Science and Technology Research Project (No. 2019-3-020).

#### References

- Becker, S., von Werder, S. C., Lassek, A. K., & Disselhorst-Klug, C. (2019). Time-frequency coherence of categorized sEMG data during dynamic contractions of biceps, triceps, and brachioradialis as an approach for spasticity detection. *Medical & Biological Engineering & Computing*, 57(3), 703–713.
- Chen, Y., Yang, Z., & Wang, J. (2015). Eyebrow emotional expression recognition using surface EMG signals. *Neurocomputing*, 168, 871–879.
- Chen, J. X., Zhang, P. W., Mao, Z. J., Huang, Y. F., Jiang, D. M., & Zhang, Y. N. (2019). Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks. *IEEE Access*, 7, 44317–44328.
- Chowdhury, R. H., Reaz, M. B., Ali, M. A. B. M., Bakar, A. A., Chellappan, K., & Chang, T. G. (2013). Surface electromyography signal processing and classification techniques. *Sensors*, 13(9), 12431–12466.
- Dey, N., Ashour, A. S., Mohamed, W. S., & Nguyen, N. G. (2019). Acoustic sensors in biomedical applications. In *Acoustic sensors for biomedical applications* (pp. 43–47). Cham: Springer.
- Du, Y. C., Lin, C. H., Shyu, L. Y., & Chen, T. (2010). Portable hand motion classifier for multi-channel surface electromyography recognition using grey relational analysis. *Expert Systems with Applications*, 37(6), 4283–4291.
- Edmonds, J. H., Couture, L. J., Paloheimo, M. P., & Rigor, S. B. (1988). Objective assessment of opioid action by facial muscle surface electromyography (sEMG). *Progress in Neuro-psychopharmacology & Biological Psychiatry*, 12(5), 727–738.
- Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). Emotion in the human face: guidelines for research and an integration of findings: guidelines for research and an integration of findings. *Pergamon*, 181–187.
- Geng, X. (2016). Label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 28(7), 1734–1748.
- Geng, X., Yin, C., & Zhou, Z. H. (2013). Facial age estimation by learning from label distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(10), 2401–2412.
- Hamedi, M., Salleh, S. H., & Swee, T. T. (2011). Surface electromyography-based facial expression recognition in bi-polar configuration. *Journal of Computer Science*, 7(9), 1407–2141.
- Jain, D. K., Shamsolmoali, P., & Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. *Pattern Recognition Letters*, 120, 69–74.
- Jia, X., Zheng, X., Li, W., Zhang, C., & Li, Z. (2019). Facial emotion distribution learning by exploiting low-rank label correlations locally. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 9841–9850).
- Jiang, M., Rahmani, A. M., Westerlund, T., Liljeborg, P., & Tenhunen, H. (2015). Facial expression recognition with sEMG method. In *2015 IEEE international conference on computer and information technology: Ubiquitous computing and communications; Dependable, autonomic and secure computing; Pervasive intelligence and computing* (pp. 981–988). IEEE.
- Kehri, V., Ingle, R., Patil, S., & Awale, R. N. (2019). Analysis of facial emg signal for emotion recognition using wavelet packet transform and svm. In *Machine intelligence and signal analysis* (pp. 247–257). Singapore: Springer.
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & Van Knippenberg, A. D. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, 24(8), 1377–1388.

- Loconsole, C., Cascarano, G. D., Brunetti, A., Trotta, G. F., Losavio, G., Bevilacqua, V., et al. (2019). A model-free technique based on computer vision and sEMG for classification in Parkinson's disease by using computer-assisted handwriting analysis. *Pattern Recognition Letters*, 121, 28–36.
- Mesa, I., Rubio, A., Tubia, I., De No, J., & Diaz, J. (2014). Channel and feature selection for a surface electromyographic pattern recognition task. *Expert Systems with Applications*, 41(11), 5190–5200.
- Mithbavkar, S. A., & Shah, M. S. (2019). Recognition of emotion through facial expressions using EMG signal. In *2019 International Conference on Nascent Technologies in Engineering (ICNTE)* (pp. 1–6). IEEE.
- Pancholi, S., & Joshi, A. M. (2019). Electromyography-based hand gesture recognition system for upper limb amputees. *IEEE Sensors Letters*, 3(3), 1–4.
- Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In *Theories of emotion* (pp. 3–33). Academic press.
- Renda, A., Barsacchi, M., Bechini, A., et al. (2019). Comparing ensemble strategies for deep learning: An application to facial expression recognition. *Expert Systems with Applications*, 136, 1–11.
- Santamaria-Granados, L., Munoz-Organero, M., Ramirez-Gonzalez, G., Abdulhay, E., & Arunkumar, N. J. I. A. (2018). Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS). *IEEE Access*, 7, 57–67.
- Schmidt, K. L., & Cohn, J. F. (2001). Dynamics of facial expression: Normative characteristics and individual differences. *IEEE international conference on multimedia and Expo, 2001. ICME 2001*. 140–140.
- Song, C., Liu, W., & Wang, Y. (2013). Facial expression recognition based on hessian regularized support vector machine. In *Proceedings of the fifth international conference on internet multimedia computing and service* (pp. 264–267).
- Sun, J., Ovsjanikov, M., & Guibas, L. (2009). A concise and provably informative multi-scale signature based on heat diffusion No. 5. In *Computer graphics forum* (pp. 1383–1392). Oxford, UK: Blackwell Publishing Ltd.
- Veer, K., & Vig, R. (2018). Identification and classification of upper limb motions using PCA. *Biomedical Engineering/Biomedizinische Technik*, 63(2), 191–196.
- Wang, T. H., & Lien, J. J. J. (2009). Facial expression recognition system based on rigid and non-rigid motion separation and 3D pose estimation. *Pattern Recognition*, 42(5), 962–977.
- Xi, X., Tang, M., & Luo, Z. (2018). Feature-level fusion of surface electromyography for activity monitoring. *Sensors*, 18(2), 614.
- Zangeneh, E., Rahmati, M., & Mohsenzadeh, Y. (2020). Low resolution face recognition using a two-branch deep convolutional neural network architecture. *Expert Systems with Applications*, 139 112854.
- Zhou, Y., Xue, H., & Geng, X. (2015, October). Emotion distribution recognition from facial expressions. In *Proceedings of the 23rd ACM international conference on multimedia* (pp. 1247–1250).