

An Emotion Recognition System Based on Physiological Signals Obtained by Wearable Sensors

Cheng He, Yun-jin Yao and Xue-song Ye

Abstract Automatic emotion recognition is a major topic in the area of human–robot interaction. This paper presents an emotion recognition system based on physiological signals. Emotion induction experiments which induced joy, sadness, anger, and pleasure were conducted on 11 subjects. The subjects' electrocardiogram (ECG) and respiration (RSP) signals were recorded simultaneously by a physiological monitoring device based on wearable sensors. Compared to the non-wearable physiological monitoring devices often used in other emotion recognition systems, the wearable physiological monitoring device does not restrict the subjects' movement. From the acquired physiological signals, one hundred and forty-five signal features were extracted. A feature selection method based on genetic algorithm was developed to minimize errors resulting from useless signal features as well as reduce computation complexity. To recognize emotions from the selected physiological signal features, a support vector machine (SVM) method was applied, which achieved a recognition accuracy of 81.82, 63.64, 54.55, and 30.00 % for joy, sadness, anger, and pleasure, respectively. The results showed that it is feasible to recognize emotions from physiological signals.

Keywords Emotion recognition · Physiological signals · Wearable sensors · Genetic algorithm · Support vector machine

1 Introduction

Automatic emotion recognition is a major topic in the area of human–robot interaction. People express emotions through facial expressions, tone of voice, body postures, and gestures which are accompanied with physiological changes. Facial expressions, tone of voice, body postures, and gestures are controlled by the

C. He · Y. Yao · X. Ye (✉)

Department of Biomedical Engineering and Instrument Science,
Zhejiang University, 310027 Hangzhou, China
e-mail: yexuesong@zju.edu.cn

somatic nervous system while physiological signals, such as electroencephalogram (EEG), heart rate (HR), electrocardiogram (ECG), respiration (RSP), blood pressure (BP), electromyogram (EMG), skin conductance (SC), blood volume pulse (BVP), and skin temperature (ST) are mainly controlled by the autonomous nervous system. That means facial expressions, tone of voice, body postures, and gestures can be suppressed or masked intentionally while physiological signals can hardly be masked. Using physiological signals to recognize emotions is also helpful to those people who suffer from physical or mental illness thus exhibit problems with facial expressions, tone of voice, body postures or gestures.

Researches have shown a strong correlation between emotions and physiological signals. However, whether it is reliable to recognize emotions from physiological signals is still problematic. Numerous researches were investigating the problem (Picard et al. 2001; Lisetti and Nasoz 2004; Kim and André 2008; Rattanyu et al. 2010; Verma and Tiwary 2014).

This paper presents an emotion recognition system based on physiological signals obtained by wearable sensors. Some common emotion models and emotion induction methods are described briefly. The data collection procedure during which a physiological monitoring device based on wearable sensors was used is introduced. The strategy for feature extraction from the acquired physiological signals and the feature selection method based on genetic algorithm are illustrated. The support vector machine (SVM) method which was used to classify the physiological features into four kinds of emotions is demonstrated. The experiment implementation procedure is presented as well. Finally, the results of the experiments are discussed, which contribute to a conclusion.

2 Method

2.1 Emotion

In discrete emotion theory, all humans are thought to have an innate set of basic emotions that are cross-culturally recognizable (Ekman and Friesen 1971). In dimensional emotion theory, however, emotions are defined according to multiple dimensions (Schlosberg 1954). Although it is problematic which emotions are basic in discrete emotion theory (Gendron and Barrett 2009) and in which dimensions emotions should be defined in dimensional theory (Rubin and Talarico 2009), it's no doubt that joy, sadness, anger, and pleasure are four different common emotions in humans. Those four emotions were chosen as the classification categories in our study.

To obtain the physiological signals associated with the specific emotions, an effective emotion induction procedure is of significance. Numerous emotion or mood induction procedures (MIPs) have been reported including presenting subjects with emotional stimuli (pictures, film clips, etc.), and letting subjects play

games (van't Wout et al. 2010) or interact with human confederate (Kučera and Haviger 2012). Several picture, audio, or video databases for emotion induction have also been created (Biehl et al. 1997; Bai et al. 2005; Bradley and Lang 2008).

In our study, we did not use the emotion induction materials from those databases above because those materials did not induce the expected emotions effectively in our experiments. Instead, we selected several contagious video clips which performed better in our emotion induction experiments.

2.2 Physiological Signals Processing

2.2.1 Data Collection

Several kinds of physiological signals including ECG and RSP signals have been revealed to be correlated with emotions. To collect ECG and RSP signals, a physiological monitoring device based on wearable sensors which monitors multiple physiological signals simultaneously in real time (Zhou et al. 2015) was used. The ECG signals were sampled at 250 Hz and the RSP signals were sampled at 10 Hz. The schematic representation of a normal ECG waveform is shown in Fig. 1 and the ECG and RSP waveforms obtained by the physiological monitoring device are shown in Figs. 2 and 3, respectively.

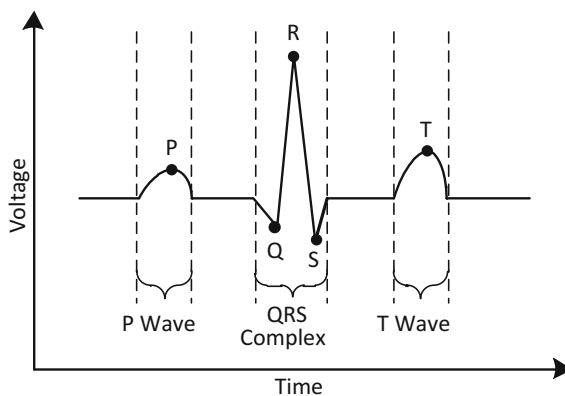


Fig. 1 Schematic representation of a normal electrocardiogram (ECG) waveform. An ECG waveform consists of a P wave, a QRS complex and a T wave. The QRS complex usually has much larger amplitude than the P wave and the T wave. P is the peak of a P wave. Q is the start of a QRS complex. R is the peak of a QRS complex. S is the end of a QRS complex. T is the peak of a T wave

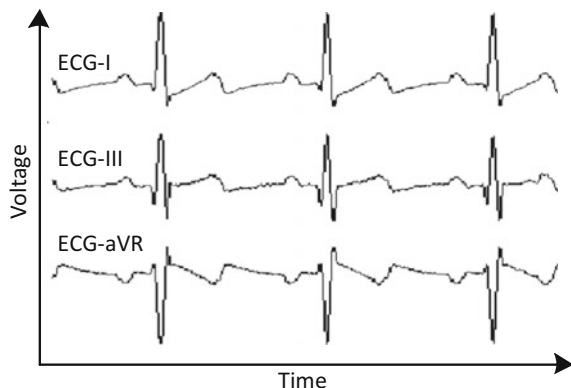
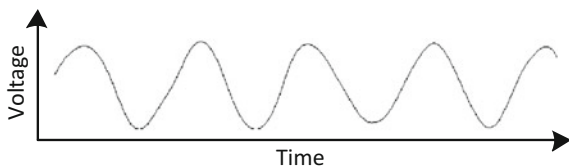


Fig. 2 Electrocardiogram (ECG) signals obtained by the physiological monitoring device. ECG-I is the voltage between the left arm electrode and right arm electrode. ECG-III is the voltage between the left leg electrode and the right leg electrode. ECG-aVR is the voltage between the right arm electrode and the combination of the left arm electrode and the left leg electrode

Fig. 3 Respiration (RSP) signals obtained by the physiological monitoring device



2.2.2 Feature Extraction

After the P-waves, the QRS complexes, and the T waves of the ECG signals were determined, a total of 78 ECG signal features were extracted as follows:

1. The mean value, median value, standard variance, minimum value, maximum value, and value range of R-R, P-P, Q-Q, S-S, T-T, P-Q, Q-S, and S-T time intervals;
2. The mean value, median value, standard variance, minimum value, maximum value, and value range of the amplitudes of P waves, QRS complexes, and T waves divided by the mean value of the corresponding ECG waveforms;
3. The mean value, median value, standard variance, minimum value, maximum value, and value range of HRD (the histogram distribution of R-R time intervals);
4. HR50 (the number of pairs of adjacent R-R time intervals differing by more than 50 ms divided by the total number of R-R time intervals);
5. HRDV (sum of HRD divided by the maximum value of HRD)
6. Each spectrum power of ECG signals in four frequency band (0–0.2 Hz, 0.2–0.4 Hz, 0.4–0.6 Hz, and 0.6–0.8 Hz).

Before RSP features were extracted, a low-pass filter was applied to the raw RSP signals. After that, a total of 67 RSP signal features were extracted as follows:

1. The mean value, median value, standard variance, minimum value, maximum value, value range, and peak ratio (the number of peaks divided by the length of data) of the following signals:
 - (a) RSP waves, RSP peak–peak intervals, and RSP peak amplitudes;
 - (b) The first difference of RSP waves, RSP peak–peak intervals, and RSP peak amplitudes
 - (c) The second difference of RSP waves, RSP peak–peak intervals, and RSP peak amplitudes
2. Each spectrum power of RSP signals in four frequency band (0–0.1 Hz, 0.1–0.2 Hz, 0.3–0.3 Hz, and 0.3–0.4 Hz).

Considering the seventy-eight ECG signal features and the sixty-seven RSP signal features, a total of one hundred and forty-five features were extracted.

2.2.3 Feature Selection

More features usually provide more information about the original signals, but also lead to an increase in computational complexity. Besides, the random noise in those signal features which make little contribution to identify different emotions might leads to overfitting in supervised machine learning such as SVM. Therefore, an effective feature selection method to select only a key subset of measured features to create a classification model is needed. Emotion recognition can be looked as a pattern recognition issue. For a pattern recognition issue, the selection criterion usually involves the minimization of a specific measure of predictive error for models which fit to different subsets. A common method is sequential feature selection (SFS) (Cover and Van Campenhout 1977), which adds features from a candidate subset while evaluating the criterion. Another novel method is using genetic algorithm (Deb et al. 2002) to select features, which will be described here.

The genetic algorithm (GA) is a method based on natural selection which drives biological evolution. The GA repeatedly modifies a population of individual solutions. At each step, the GA selects individuals at random from the current population to be parents and uses them to produce the children for the next generation. There are some rules like crossover at each step to create the next generation from the current population. At each step, the individual selection is random, but the survival opportunity of each individual is not equal. The individuals who have higher survival opportunity are more likely to be selected and keep evolving till the optimization goal is reached. In our study, the survival opportunity was evaluated by the emotion recognition error.

Through the GA algorithm described above, fourteen features were selected from the original one hundred and forty-five features.

2.3 Emotion Recognition

To recognize emotions from the key features selected by GA, a modified support vector machine (SVM) method was used. An SVM classifies data by finding the optimal hyperplane that separates all data points of one class from those of another class (Cortes and Vapnik 1995). The optimal hyperplane for an SVM means the one with the maximum margin between the two classes. A margin is the maximal width of two slabs parallel to the hyperplane that have no interior data points. A larger margin assures the hyperplane is more likely to classify new data correctly. The data points that are on the boundary of the slab are called support vectors. The complexity of the classifier is characterized by the number of support vectors rather than the dimensionality of the transformed hyperspace. An example of SVM is shown in Fig. 4.

Sometimes the data might not allow for a separating hyperplane. As shown in Fig. 5, the outliers caused by error such as artifact during data collection make it difficult to find a proper separating hyperplane. Even if a separating hyperplane is found, the margin is small. In that case, a soft margin method is proposed which chooses a hyperplane that splits the examples as cleanly as possible while still maximizing the distance to the nearest cleanly split (Cortes and Vapnik 1995).

Some binary classification problems do not have an effective linear separating hyperplane, so-called nonlinear classification, as shown in Fig. 6a. In this case, the initial hyperspace S is transformed to a higher dimensional hyperspace S' , as shown in Fig. 6b. In the higher dimensional hyperspace S' , there is a linear hyperplane to successfully separate the two classes. Usually, the analysis formula of the

Fig. 4 Linear Support vector machine. The optimal linear hyperplane separates all samples into two classes with a maximum margin

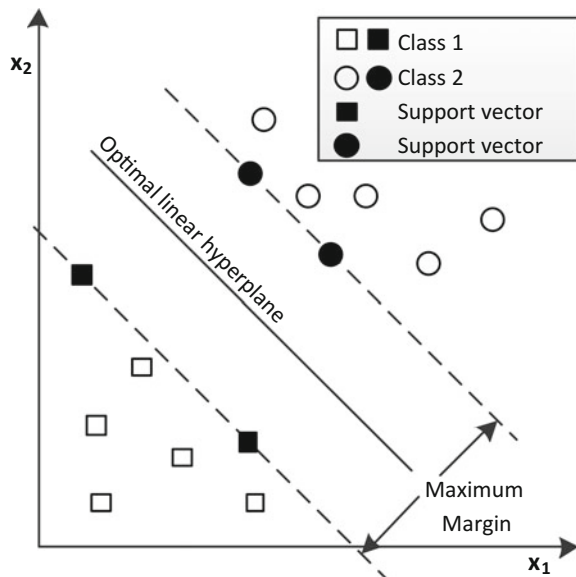
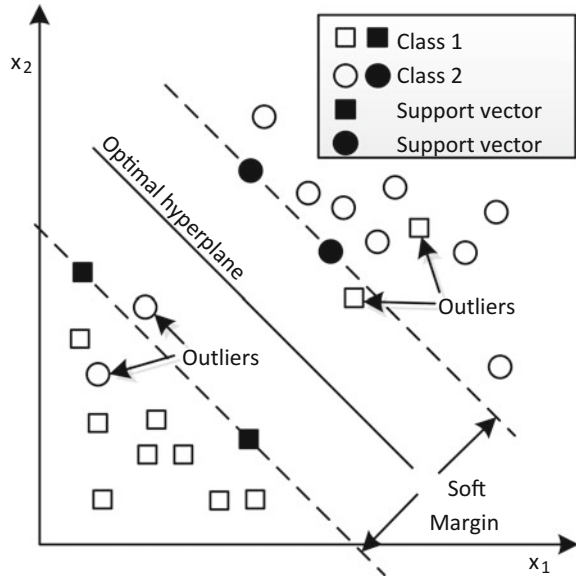


Fig. 5 Support vector machine with a soft margin. The soft margin allows for some mislabeled samples to maximize the margin



transformation is difficult to get. However, It is found that all the calculations for hyperplane classification use nothing more than dot products, then a nonlinear kernel function in linear hyperspace S is developed to replace the dot products in the higher dimensional hyperspace S' (Boser et al. 1992). Some common kernels are listed here: polynomial function, Gaussian radial basis function (RBF) and multi-layer perceptron (neural network, NN).

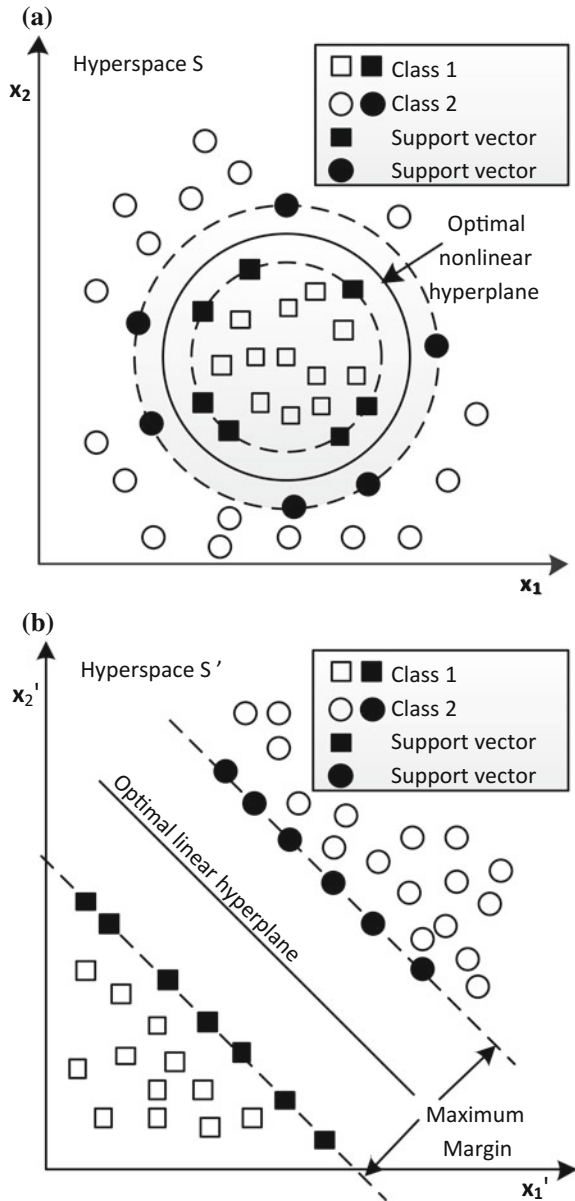
Support vector machines (SVMs) are originally designed for binary classification. But there have been some extensions for multiclass classification (Hsu and Lin 2002). One of the multiclass classification approaches using SVM is building binary classifiers which distinguished between every pair of classes, so-called one-against-one (Knerr et al. 1990). For a one-against-one approach, classification is done by a max-wins voting strategy that every classifier assigns the instance to one of the two classes, then the vote for the assigned class is increased by one vote, and finally, the class with the most votes determines the instance classification.

In our multi-emotion recognition system, we applied linear and nonlinear SVMs with a soft margin as the classifiers and one-against-one method as the multiclass classification approach.

2.4 Experiment Implementation

First, we used Beck Depression Inventory-II (BDI-II) (Beck et al. 1996) and Toronto Alexithymia Scale (TAS) to select 11 subjects who had no depression (BDI-II score < 4) and were capable of expressing emotions clearly (TAS < 60)

Fig. 6 Transform a nonlinear SVM to a linear SVM. Kernels are used to fit the maximum margin nonlinear hyperplane in a transformed linear hyperplane without knowing the analysis formula the transformation function



from the experiment volunteers. This study was approved by the Institutional Review Board of Zhejiang University. Informed written consent was obtained from all experiment volunteers. Then, we prepared a quiet multimedia experiment room equipped with an air conditioner, a computer, a 17 inch LCD screen, and a pair of stereo headphones. In each experiment, one subject wearing the physiological

sensors sat alone in the multimedia room with air temperature setting to 25 °C and watched the video clips or listened to the music we had prepared. Before and after each experiment, the subject was asked to fill out a questionnaire about his or her experience and emotion. The signals of those subjects who did not report expected emotion were labeled invalid and discarded. For the valid signals, the time slot when the subjects were most likely in an expected emotion state was determined.

3 Results and Discussion

The SVM classifiers were tested with leave-one-out cross validation. Leave-one-out cross validation involves using one observation as the validation set and the remaining observations as the training set. This is repeated on all ways to cut the original sample on a validation set of one observations and a training set. The linear SVM and the nonlinear SVMs with polynomial function, Gaussian radial basis function, and multilayer perceptron as the kernel function were all tested. The results are shown in Table 1.

As shown in Table 1, the linear SVM classifier achieved the highest recognition accuracy for four emotions in total while the nonlinear SVM classifier with a polynomial function as the kernel achieved the lowest recognition accuracy for four emotions in total. The linear SVM classifier performed better than the nonlinear SVM classifiers in total. That is probably because the number of physiological features was reduced from one hundred and forty-five to fourteen during the feature selection procedure and the linear SVM was able to provide a relatively good classifier.

As to the recognition accuracy for each emotion, the linear SVM and nonlinear SVMs with different functions as the kernel performed differently. However, each SVM classifier achieved higher recognition accuracy for joy than for the other three emotions. There might be two reasons for that. One is that joy causes greater physiological changes than the other three emotions. The other one is that the induction for joy was more effective than the other three emotions in the emotion induction experiments. In addition, although the physiological data from the subjects who did not report expected emotions were labeled as invalid and discarded,

Table 1 Emotion recognition accuracy using the linear SVM and the nonlinear SVMs with different functions as the kernel. RBF is Gaussian radial basis function. Poly is polynomial function. NN is multilayer perceptron (neural network) function

	Linear (%)	RBF (%)	Poly (%)	NN (%)
Joy	81.82	63.64	45.45	90.91
Sadness	63.64	18.18	18.18	63.64
Anger	54.55	18.18	27.27	36.36
Pleasure	30.00	20.00	20.00	10.00
Total	58.14	30.23	27.91	51.16

there exists the possibility that the subjects reported their emotions inaccurately. Another possible reason is that pleasure is close to joy and the SVM classifier failed to distinguish them.

To improve the performance of the presented emotion recognition system, the following methods could be taken into account in further work. Precisely designed emotion induction experiments could be conducted on more subjects. Some other supervised machine learning algorithms could be developed. And some other physiological signals like EMG signals might be obtained together with ECG and RSP signals, but it should be noted that the acquisition process of physiological signals should not make the subjects feel uncomfortable.

4 Conclusion

As physiological signals cannot be masked intentionally, recognizing emotions from physiological signals has advantages over from facial expressions, tone of voice, body postures, and gestures. Based on a combination of a feature selection method and a support vector machine method, it is feasible to recognize emotions from physiological signals.

Acknowledgments This work is supported by National Science and Technology Major Project of the Ministry of Science and Technology of China (No. 2013ZX03005008). And the authors would like to thank Congcong Zhou, Chunlong Tu, Jian Tian, Jingjie Feng, and Yun Gao for their physiological monitoring device and advice.

References

- Bai L, Ma H, Huang YX (2005) The development of native chinese affective picture system—a pretest in 46 college students. *Chin Ment Health J* 19(11):719–722
- Beck AT, Steer RA, Brown GK (1996) Manual for the beck depression inventory-II. Technical report. Psychological Corporation, San Antonio, Texas
- Biehl M, Matsumoto D, Ekman P et al (1997) Matsumoto and Ekman’s Japanese and Caucasian facial expressions of emotion (JACFEE): reliability data and cross-national differences. *J Nonverbal Behav* 21(1):3–21. doi:[10.1023/A:1024902500935](https://doi.org/10.1023/A:1024902500935)
- Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: *Proceedings of the fifth annual workshop on Computational learning theory*, pp 144–152
- Bradley MM, Lang PJ (2008) International affective picture system (IAPS): affective ratings of pictures and instruction manual. Technical report, The Center for Research in Psychophysiology, University of Florida, Gainesville, Florida
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. doi:[10.1007/BF00994018](https://doi.org/10.1007/BF00994018)
- Cover TM, Van Campenhout JM (1977) On the possible orderings in the measurement selection problem. *IEEE Trans Syst Man Cybern* 7(9):657–661. doi:[10.1109/TSMC.1977.4309803](https://doi.org/10.1109/TSMC.1977.4309803)
- Deb K, Pratap A, Agarwal S et al (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput* 6(2):182–197. doi:[10.1109/4235.996017](https://doi.org/10.1109/4235.996017)

- Ekman P, Friesen WV (1971) Constants across cultures in the face and emotion. *J Pers Soc Psychol* 17(2):124. doi:[10.1037/h0030377](https://doi.org/10.1037/h0030377)
- Gendron M, Barrett LF (2009) Reconstructing the past: a century of ideas about emotion in psychology. *Emot Rev* 1(4):316–339. doi:[10.1177/1754073909338877](https://doi.org/10.1177/1754073909338877)
- Hsu CW, Lin CJ (2002) A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Networks* 13(2):415–425. doi:[10.1109/72.991427](https://doi.org/10.1109/72.991427)
- Kim J, André E (2008) Emotion recognition based on physiological changes in music listening. *IEEE Trans Pattern Anal Mach Intell* 30(12):2067–2083. doi:[10.1109/TPAMI.2008.26](https://doi.org/10.1109/TPAMI.2008.26)
- Knerr S, Personnaz L, Dreyfus G (1990) Single-layer learning revisited: a stepwise procedure for building and training a neural network. *Neurocomputing*. Springer, Heidelberg, pp 41–50. doi:[10.1007/978-3-642-76153-9_5](https://doi.org/10.1007/978-3-642-76153-9_5)
- Kučera D, Haviger J (2012) Using mood induction procedures in psychological research. *Procedia Soc Behav Sci* 69:31–40. doi:[10.1016/j.sbspro.2012.11.380](https://doi.org/10.1016/j.sbspro.2012.11.380)
- Lisetti CL, Nasoz F (2004) Using noninvasive wearable computers to recognize human emotions from physiological signals. *EURASIP J Appl Sig Process* 2004(11):1672–1687. doi:[10.1155/S1110865704406192](https://doi.org/10.1155/S1110865704406192)
- Picard RW, Vyzas E, Healey J (2001) Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Trans Pattern Anal Mach Intell* 23(10):1175–1191. doi:[10.1109/34.954607](https://doi.org/10.1109/34.954607)
- Rattanyu K, Ohkura M, Mizukawa M (2010) Emotion monitoring from physiological signals for service robots in the living space. In: 2010 International conference on control automation and systems. Gyeonggi-do, pp 580–583
- Rubin DC, Talarico JM (2009) A comparison of dimensional models of emotion: evidence from emotions, prototypical events, autobiographical memories, and words. *Memory* 17(8):802–808
- Schlosberg H (1954) Three dimensions of emotion. *Psychol Rev* 61(2):81. doi:[10.1037/h0054570](https://doi.org/10.1037/h0054570)
- van't Wout M, Chang LJ, Sanfey AG (2010) The influence of emotion regulation on social interactive decision-making. *Emotion* 10(6):815–821. doi:[10.1037/a0020069](https://doi.org/10.1037/a0020069)
- Verma GK, Tiwary US (2014) Multimodal fusion framework: a multiresolution approach for emotion classification and recognition from physiological signals. *Neuroimage* 102(P1):162–172. doi:[10.1016/j.neuroimage.2013.11.007](https://doi.org/10.1016/j.neuroimage.2013.11.007)
- Zhou CC, Tu CL, Tian J et al (2015) A low power miniaturized monitoring system of six human physiological parameters based on wearable body sensor network. *Sens Rev* 35(2):210–218. doi:[10.1108/SR-08-2014-687](https://doi.org/10.1108/SR-08-2014-687)