S & M 2852

Approximate Model for Stress Assessment Using Electroencephalogram Signal

Ying Lin, Hai-Feng Chen, Hui-Hong Chen,
Zhen-Lun Yang, Ting-Cheng Chang,* and Zeng-Rong Zhan

School of Information Engineering, Guangzhou Panyu Polytechnic, Guangzhou 511483, China

(Received September 6, 2021; accepted December 2, 2021)

Keywords: EEG, ECG, stress, wearable, approximate model

Mental stress is a problem that people may often face. Although there are some psychobiological stress measurement methods based on electroencephalogram (EEG) signals, these methods use expensive medical equipment to gather multichannel signals and cannot measure stress in real time in daily life. Many novel wearable devices now have a sensor for physiological signal detection, which helps people manage health conditions. Several studies have recently investigated the application of wearable devices to stress detection. In this paper, an approximate model based on an EEG signal is designed for measuring stress. To establish the connection between the EEG and electrocardiogram signals, we use wearable devices to simultaneously collect the two types of data from volunteers. The exponentially weighted moving average is used to smooth out the EEG power spectrum features (α, β, etc.). An EEG-based feature vector is constructed to predict stress scores, with polynomial regression used to build the model. The experimental results show that the proposed method achieves a symmetric mean absolute percentage error of 17.44 and a root mean square error of 11.26.

1. Introduction

Modern life is full of stress. Anxiety, insomnia, and depression have become problems that plague many people. Appropriate stress helps us accomplish our goals, but prolonged stress is not conducive to our health.(1) Previous works have proposed methods to detect stress that use physiological signals, such as electroencephalogram (EEG) signals, electrocardiogram (ECG) signals, and galvanic skin response (GSR), to measure psychological stress more objectively. Stress level identification was a common task in these studies. However, signals obtained in the laboratory are different from those resulting from stress in daily life.(2) Nowadays, some people use wearable devices almost every day. Many of these devices have a sensor to collect physiological data, making them ideal instruments for stress detection studies.(3)

The functioning of the heart can reveal health-related information. Since the autonomic nervous system affects the heart's activity, heart-related features can be used to measure stress. Stress detection research based on ECG signals is usually a classification task, including two classifications (stressed, relaxed) or three classifications (low, medium, high stress). The support

*Corresponding author: e-mail: 18250922163@qq.com
https://doi.org/10.18494/SAM3637

vector machine (SVM), k-nearest neighbors (kNN), neural networks, and decision trees are mostly used in these tasks.[4–6] Some researchers have studied the problem of stress detection in a daily life environment. Sriramprakash *et al.* used GSR and ECG sensors to identify stress in working individuals.[7] With the development of wearable devices, many smart bands now have real-time stress assessment functions, such as the Apple Watch, Huawei Watch, and Fitbit. These commercial devices record how stress changes with time. However, the electrodes of these devices are far from the heart, causing the severe decay of the ECG signals.

Brain activity is also affected by stress. Researchers have extracted and selected features from EEG signals and used a machine learning algorithm to build a model of brain activity. Some works reviewed stress level recognition.[8–10] Sandulescu *et al.* studied the application of wearable devices in stress detection.[11] Minguillon *et al.* collected the prefrontal relative gamma power to evaluate the level of stress. Since EEG signals have a higher temporal resolution than the heart rate (HR), they are more suitable for real-time stress detection implementation.[12,13] These methods all use multichannel signals. However, wearable devices are usually low-power devices and are not capable of providing multichannel data. Moreover, to the best of our knowledge, recent studies have mainly focused on classifying stress into two or three levels, not stress scores, in other words, a continuous variable indicating the stress. Therefore, it is desirable to develop a stress detection method that uses only a few electrodes to reduce the amount of computation and determine stress scores using EEG features.

For stress detection, if the same person's EEG and ECG data are collected simultaneously, there must be a correlation between the two signals. From previous studies, we also know that the prefrontal cortex is affected by stress, and the frequency domain features of EEG signals can detect this effect. Therefore, a well-trained ECG stress assessment model could be used to generate labels for the EEG data, that is, to build an approximate model based on EEG features. In this work, we investigated a method of using biosensors to collect EEG and ECG signals, and built a model that can present mental stress as a score.

2. Method

In this study, we propose a stress assessment method based on the frequency domain features of frontal EEG data that can be employed in a wearable device. Our approach uses commercial wearable sensors that can collect data in a daily life environment. In this section, the architecture and algorithm of the method are described.

Figure 1 shows the architecture of the proposed method. The entire training process is divided into four stages: data collection, preprocessing, feature selection, and training.

2.1 Data collection

To obtain a more precise ECG signal, the user should place the sensor closer to the heart. Consequently, the sensor should be mounted on a chest strap. In this work, a wearable chest strap and brain band [Fig. 2(a)] were used to collect biological signals, and an Android application [Fig. 2(b)] was developed to gather data. Both wearable devices use dry-electrode and CR2032

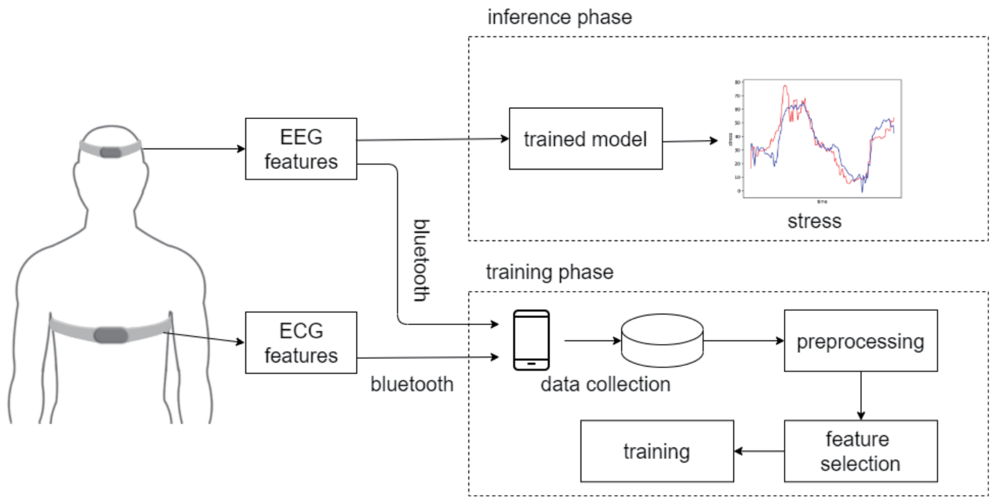

Fig. 1. (Color online) System architecture.

(a) (b)

Fig. 2. (Color online) (a) EEG and ECG sensors. (b) Android application used to collect and store data.

batteries. The sensors themselves extract the features, and the feature information is broadcast via Bluetooth every 6 s. When the sensors are placed on the body, the sensors start sampling and processing the signals. The amplitudes of the power spectrum features α, β, δ, θ, and γ can be acquired from the EEG sensor without any preprocessing of the raw data. These features are sufficient for the stress assessment task. Similarly, the HR, low frequency (LF, 0.04–0.15 Hz) component, high frequency (HF, 0.15–0.4 Hz) component, standard deviation of normal to normal (SDNN), total power (TP), and stress score can be acquired from the ECG sensor.

As shown in Fig. 1, EEG and heart rate variability (HRV) data are used to train the model during the training phase. EEG feature data must be paired with a stress score to establish the relationship between EEG and ECG data. Although the two sensors broadcast data at the same rate, they are not synchronized. The Android application stores the stress score in a cache to deal with this problem and updates the stored variables every time new data is received. When EEG data is received, the newest stress score is paired with the data. In the inference stage, only the EEG feature is needed to predict the stress.

2.2 Preprocessing

The EEG signal is susceptible to noise. Most methods preprocess the raw signal to remove blinking, eye movement, and the electromyogram (EMG) to obtain clear data. In this study, these tasks are performed by a sensor. However, the amplitude of the frequency domain features still fluctuates greatly. It is necessary to preprocess the features data to make it smoother. The moving average (MA) is a common algorithm used to filter out noise, leading to a lag. As the sliding window increases, the lag increases, hindering real-time stress calculation. Therefore, the exponentially weighted moving average (EWMA) algorithm is used to smooth out the amplitude of the features data. Two preprocessing steps are employed: (i) remove outliers to reduce the impact on the average and (ii) use the EWMA algorithm to smooth the signal and reduce the lag.

(i) Calculate the sample variance s and average $\bar{x}$ on the training set, and limit the maximum and minimum values of the data to $\bar{x} \pm s$, as shown in Eqs. (1)–(3).

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{1}$$

$$x_i = \max\{\bar{x} + s, x_i\} \tag{2}$$

$$x_i = \min\{\bar{x} - s, x_i\} \tag{3}$$

(ii) Smooth with EWMA. The EWMA algorithm is a widely used smoothing algorithm in the fields of finance and engineering. The weight of the data can be changed by adjusting the hyperparameter α to improve the robustness of the model. A smaller α takes more recent data into consideration. Equation (4) shows the EWMA.

$$v_t = \frac{x_t + (1-\alpha)x_{t-1} + (1-\alpha)^2 x_{t-2} + \ldots + (1-\alpha)^t x_0}{1 + (1-\alpha) + (1-\alpha)^2 + \ldots + (1-\alpha)^t} \tag{4}$$

Here, α is the decay coefficient, x_t is the amplitude of the feature at time t, and v_t is the average.

2.3 Construction of feature vector

The EEG signal contains rich features. From previous studies, we know that brain activity is affected by mental stress. α waves are related to a relaxed and calm state, and β waves to stress and excitement. Therefore, α and β are selected to construct the feature vector as follows:

$$\boldsymbol{F} = \begin{bmatrix} 1 & \alpha & \beta & \alpha\beta & \alpha^2 & \beta^2 \end{bmatrix}. \tag{5}$$

2.4 Model

Algorithms running on wearable devices must not be computation-intensive in order to extend their battery life. In some studies, deep neural networks have been used to detect stress.(14,15) These models are complex and have a too heavy computational load for a wearable device with a battery. Moreover, our aim is to predict a continuous stress value rather than a stress level, and therefore classification algorithms such as SVM, kNN, and decision trees are excluded. For these reasons, polynomial regression is selected to build the model. Equation (6) shows our model:

$$Stress = \boldsymbol{F}\boldsymbol{w}^{\boldsymbol{T}}, \tag{6}$$

where $\boldsymbol{w} = [\omega_0 \quad \omega_1 \quad \dots \quad \omega_5]$.

3. Experiment and Results

3.1 Dataset

A total of 23 healthy subjects between the ages of 20 and 40 voluntarily participated, and they were asked not to exercise for 24 h before the experiment. Each volunteer was assigned a headband and a chest strap, which they had to wear continuously for no less than 2 h in the office during the experiment. The app starts recording after receiving the data from both sensors. Daily activities were allowed, such as working, talking, and resting. We did not conduct the Stroop color-word test or Trier social stress test (TSST) to induce stress. After data was collected, it was exported through DB Browser (SQLite). Table 1 shows eight records in the database. The collected data was ordered by time. Each record contains a recording time, subject name, EEG features, stress value, and hardware information. For privacy reasons, Table 1 only shows some

Table 1
EEG features data and stress scores.

ID	α	β	δ	θ	γ	Stress
1	390	520	30228	203	137	52.65
2	117	374	9404	301	17	41.19
3	172	564	1741	253	34	42.29
4	172	564	1741	253	34	40.76
5	1155	1249	57513	3332	266	41.37
6	1155	1249	57513	3332	266	41.93
7	1270	919	18925	2939	565	46.46
8	137	530	3136	160	22	42.4

of these features. Note that the stress score was obtained directly from the chest strap and served as the target value in the experiment.

Figure 3 shows the amplitude of β for a volunteer. In the graph, the gray line shows the raw data and the blue line shows the result after preprocessing. In this experiment, the smoothing factor α was set to 0.1. It can be seen that after EWMA filtering, the amplitude is much smoother and there is no notable lag.

The features data was sorted (in ascending order by stress score) and smoothed to find the correlation. Figure 4 shows that the amplitude of β and the stress are positively correlated.

3.2 Training model

Polynomial regression is one of the most commonly used machine learning algorithms, and its coefficients can be estimated through the stochastic gradient descent (SGD) method or the

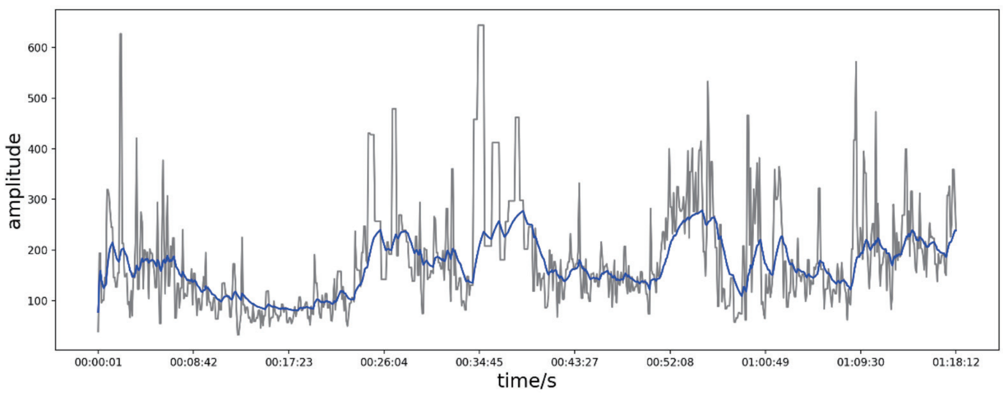

Fig. 3. (Color online) Amplitude of β after smoothing.

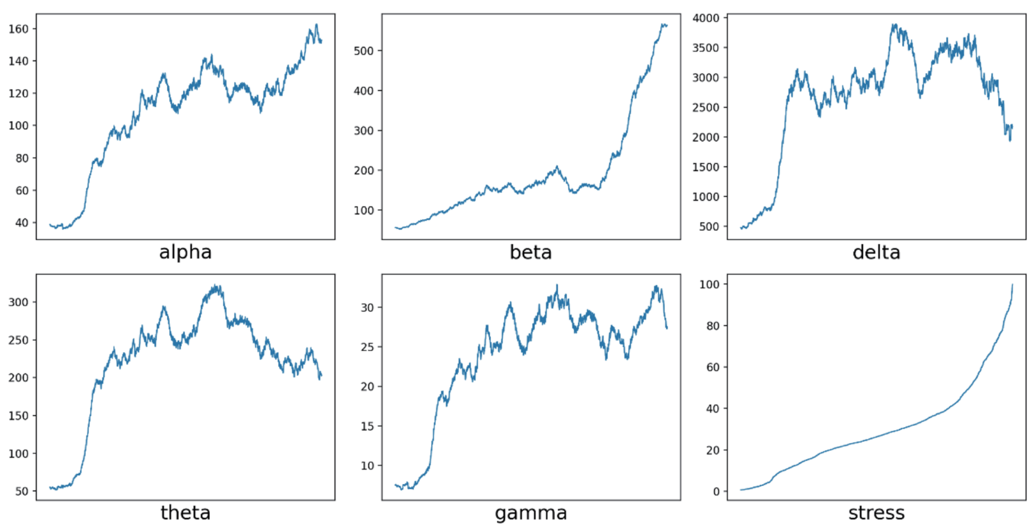

Fig. 4. (Color online) EEG features data after sorting and MA smoothing (window = 200).

normal equation. Since the dimensionality of the feature vector of the dataset is small, the normal equation was used to obtain the coefficients, as follows:

$$\theta = (X^T X)^{-1} X^T y, \tag{7}$$

where X is the training set, y is the stress value, and θ denotes the coefficients of the model. To obtain a more robust model and better testing performance, the dataset is usually shuffled during training. In our experiment, the samples have an ordered relationship. Hence, no shuffling was performed.

3.3 Results and discussion

Figure 5 shows part of the training and testing results of the model. The red line shows the stress scores of the ECG sensor and the blue line shows the inference result of the trained model. It can be seen that the proposed model can maintain a consistent trend with the output of the ECG model, which indicates that the feature vector in Eq. (5) is sufficient for stress assessment. Figure 5 suggests that mental stress can rapidly fluctuate during daily activities, increasing the difficulty of fitting.

Various statistical metrics can be used to measure the performance of a regression model. In this experiment, root mean square error (RMSE) and symmetric mean absolute percentage error (SMAPE) were used to evaluate the model. The definition of SMAPE was slightly changed to obtain a result between 0 and 100. For our model, we obtained a mean RMSE of 11.26 and a mean SMAPE of 17.44.

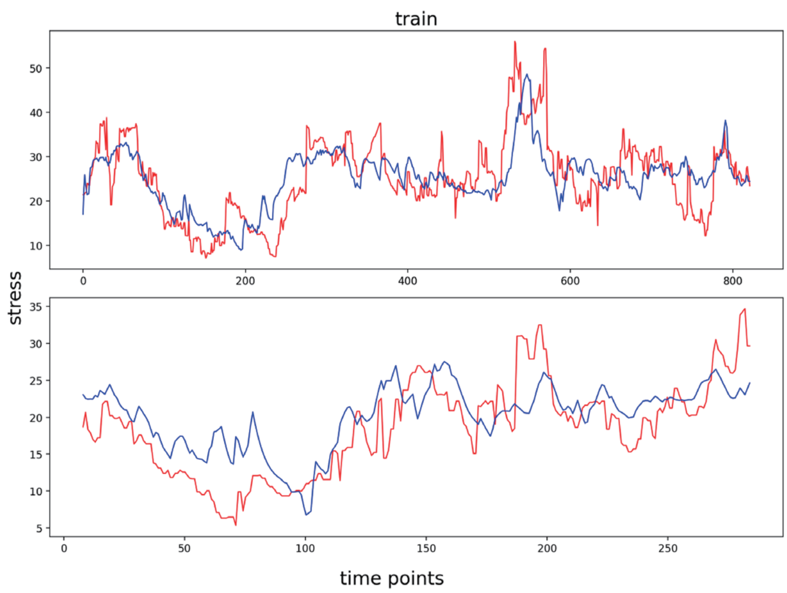

Fig. 5. (Color online) Training and testing results obtained from α and β.

Table 2
Tenfold cross-validation results of the model with different features.

Selected feature(s)	SMAPE (%)			RMSE		
	Mean	Max	Min	Mean	Max	Min
A	18.34	22.69	10.45	12.1	16.68	7.56
B	17.74	23.07	8.33	11.5	16.28	7.23
$\alpha\ \beta$	17.44	22.35	9.04	11.26	16.1	6.77
$\alpha\ \beta\ \delta\ \gamma\ \theta$	17.68	22.19	14.63	11.57	16.33	7.5

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{8}$$

$$SMAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|\widehat{y_i} - y_i|}{|\widehat{y_i}| + |y_i|} \tag{9}$$

Here, y_i is the stress score obtained from the ECG-based model and $\hat{y}_i$ is the value predicted from the EEG-based model.

Some researchers have used a single channel and dry-electrode instrument to record the prefrontal signal and implement a stress assessment system.(16–18) Nevertheless, these studies only distinguished stressed subjects from non-stressed subjects, and no studies using a continuous metric have been reported. In this experiment, by utilizing a pre-trained ECG model, our method measures the stress as a continuous value, giving a more fine-grained result.

3.4 Cross-validation

Owing to the small dataset, tenfold cross-validation was performed to obtain a more robust model. Table 2 shows the results of the model with different features. The model achieves the minimum error when using α and β. A single component and irrelevant features increase the error.

4. Conclusions

In this study, we proposed a real-time EEG-based stress assessment method for wearable devices. We performed an experiment to collect data and validate the process. The model used is efficient, straightforward, and suitable for low-power devices. The accuracy of the result is sufficient for stress assessment in daily life. In machine learning tasks, labeling data can be very time-consuming and expensive. However, the introduction of an ECG-based model facilitates the preparation of labels of EEG data, making the training procedure much more manageable.

Moreover, our method is capable of predicting stress scores, not just stress levels, which allows users to monitor the trend of their stress. The present method also has some limitations.

The performance of the procedure depends on the selected ECG model, and if the stress changes rapidly in a short period, then the output of the EEG model has a lag. We believe that by improving the ECG-based stress detection and increasing the sampling rate of the sensors, the accuracy of the proposed model will be improved. In the future, we plan to increase the number of participants to build a more robust model and design a more effective filtering algorithm and feature vector to reduce errors.

Acknowledgments

This research was funded by the Department of Education of Guangdong Province, China, under Grant Nos. 2020KQNCX192 and 2019GKTSCX069, Guangzhou Panyu Polytechnic Innovation Team under Grant No. 2020CXTD003 (2011/210113263), Guangzhou Panyu Polytechnic Research Project under Grant No. 2021KJ04 (2011/210113263), Guangzhou Municipal Education Bureau, China, under Grant No. 201831785, and the Character and Innovation Project of Education of Guangdong Province under Grant No. 2018GKTSCX081. We would like to thank all volunteers involved in our experiment.

References

1 S. C. Segerstrom and G. E. Miller: Psychol. Bull. **130** (2004) 601. https://doi.org/10.1037/0033-2909.130.4.601
2 R. W. Picard: IEEE MultiMedia **23** (2016) 3. https://doi.org/10.1109/MMUL.2016.38
3 Y. S. Can, B. Arnrich, and C. Ersoy: J. Biomed. Inf. **92** (2019) 103. https://doi.org/10.1016/j.jbi.2019.103139
4 P. Karthikeyan, M. Murugappan, and S. Yaacob: J. Mech. Med. Biol. **13** (2013) 1350038. https://doi.org/10.1142/s0219519413500383
5 N. Munla, M. Khalil, A. Shahin, and A. Mourad: 2015 Int. Conf. Advances in Biomedical Engineering (IEEE, 2015) 61.
6 S. Pourmohammadi and A. Maleki: Comput. Methods Programs Biomed. **193** (2020) 105482. https://doi.org/10.1016/j.cmpb.2020.105482
7 S. Sriramprakash, V. D. Prasanna, and O. V. R. Murthy: Procedia Comput. Sci. **115** (2017) 359. https://doi.org/10.1016/j.procs.2017.09.090
8 X. Hou, Y. Liu, O. Sourina, Y. R. E. Tan, L. Wang, and W. Mueller-Wittig: 2015 IEEE Int. Conf. Systems, Man, and Cybernetics (IEEE, 2015) 3110.
9 F. M. Al-Shargie, T. B. Tang, N. Badruddin, and M. Kiguchi: Int. Conf. Innovation in Biomedical Engineering and Life Sciences (Springer, 2015) 15.
10 G. Jun and K. G. Smitha: 2016 IEEE Int. Conf. Systems, Man, and Cybernetics (IEEE, 2016) 003270.
11 V. Sandulescu, S. Andrews, D. Ellis, N. Bellotto, and O. M. Mozos: Int. Work-Conf. Interplay between Natural and Artificial Computation (Springer, 2015) 526.
12 J. Minguillon, E. Perez, M. A. Lopez-Gordo, F. Pelayo, and M. J. Sanchez-Carrion: Sensors **18** (2018) 2504. https://doi.org/10.3390/s18082504
13 J. Minguillon, M. A. Lopez-Gordo, and F. Pelayo: Front. Comput. Neurosci. **10** (2016) 101. https://doi.org/10.3389/fncom.2016.00101
14 C. Y. Liao, R. C. Chen, and S. K. Tai: 2018 IEEE Int. Conf. Applied System Invention (IEEE, 2018) 90.
15 J. H. Lee, H. Gamper, I. Tashev, S. Dong, S. Ma, J. Remaley, J. D. Holbery, S. H. Yoon: Extended Abstracts of the 2020 CHI Conf. Human Factors in Computing Systems (ACM, 2020) 1.
16 P. Arpaia, N. Moccaldi, R. Prevete, I. Sannino, and A. Tedesco: IEEE Trans. Instrum. Meas. **69** (2020) 8335. https://doi.org/10.1109/TIM.2020.2988744
17 S. M. U. Saeed, S. M. Anwar, and M. Majid: IEICE Trans. Inf. Syst. **100** (2017) 2241. https://doi.org/10.1587/transinf.2016EDL8248
18 S. M. U. Saeed, S. M. Anwar, M. Majid, M. Awais, and M. Alnowami: Biomed Res. Int. (Hindawi, 2018). https://doi.org/10.1155/2018/1049257

About the Authors

Ying Lin received his B.S. degree from Foshan University, China, in 2016 and his M.S. degree from Nanhua University, China, in 2019. Since 2019, he has been a lecturer at Guangzhou Panyu Polytechnic. His research interests are in bioengineering, sensors, and machine learning. (ying.lins@outlook.com)

Hai-Feng Chen received his B.S. degree from Sun Yat-sen University, China, in 2011 and his M.S. degree from Sun Yat-sen University in 2013. Since 2019, he has been a lecturer at Guangzhou Panyu Polytechnic. His research interests are in bioengineering and software engineering. (chenhf@gzpyp.edu.cn)

Hui-Hong Chen received her B.S. degree from Central South University, China, in 2004 and her M.S. degree from Central South University in 2009. From 2009 to 2012, she was a software engineer at Guangzhou CongXing Electronics Co., Ltd., China. From 2012 to 2014, she was a software engineer at Oracle (Shenzhen) R&D Center, China. Since 2014, she has been an associate professor at Guangzhou Panyu Polytechnic. Her research interests are in image recognition, software development, and big data. (chenjin_2002@163.com)

Zhen-Lun Yang received his B.S and M.S degrees from South Normal China University, China, in 2001 and 2004. He received his Ph.D. degree from South China University of Technology, China, in 2016. Since 2011, he has been an associate professor at Guangzhou Panyu Polytechnic. His research interests are in evolutionary computing and machine learning. (yangzl@gzpyp.edu.cn)

Ting-Cheng Chang received his M.S. and Ph.D. degrees in process control and mechanical engineering in 1992 and 1996 from University of Houston and University of Texas at Arlington, Texas, USA, respectively. He is now a professor of information engineering at Guangzhou Panyu Polytechnic, China. His research interests lie in the Internet of Things, data mining, big data, and optimization theory. (18250922163@qq.com)

Zeng-Rong Zhan received his M.S. degree from Chalmers University of Technology, Sweden, in 2007 and his Ph.D. degree from Sun Yat-sen University, China, in 2018. He is now an associate professor of Guangzhou Panyu Polytechnic. His research interests are in machine learning and big data. (zhanpost@gmail.com)