



# Ensemble Learning Based Gender Recognition from Physiological Signals

Huiling Zhang<sup>1</sup>, Ning Guo<sup>1</sup>, Guangyuan Liu<sup>2</sup>, Junhao Hu<sup>3</sup>,  
Jiaxiu Zhou<sup>4</sup>, Shengzhong Feng<sup>1</sup>, and Yanjie Wei<sup>1</sup>(✉)

- <sup>1</sup> Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences,  
Shenzhen 518055, China  
yj.wei@siaat.ac.cn
- <sup>2</sup> College of Electronic and Information Engineering, Southwest University,  
Chongqing 400715, China  
liugy@swu.edu.cn
- <sup>3</sup> Chongqing Optoelectronics Research Institute, Chongqing 400060, China
- <sup>4</sup> Shenzhen Children's Hospital, Shenzhen 518000, China

**Abstract.** Gender recognition based on facial image, body gesture and speech has been widely studied. In this paper, we propose a gender recognition approach based on four different types of physiological signals, namely, electrocardiogram (ECG), electromyogram (EMG), respiratory (RSP) and galvanic skin response (GSR). The core steps of the experiment consist of data collection, feature extraction and feature selection & classification. We developed a wrapper method based on Adaboost and sequential backward selection for feature selection and classification. Through the data analysis of 234 participants, we obtained a recognition accuracy of 91.1% with a subset of 12 features from ECG/EMG/RSP/GSR, 82.3% with 11 features from ECG only, 80.8% with 5 features from RSP only, indicating the effectiveness of the proposed method. The ECG, EMG, RSP, GSR signals are collected from human wrist, face, chest and fingers respectively, hence the method proposed in this paper can be easily applied to wearable devices.

**Keywords:** Gender recognition · Physiological signal · Feature selection  
Ensemble learning · Wearable devices

## 1 Introduction

Gender contains a wide range of information regarding the characteristics difference between male and female. Automated gender recognition has numerous applications, including gender medicine [1, 2], video surveillance [3, 4], human machine interaction [5, 6]. Recently, with the development of social networks and mobile devices such as smartphones, gender recognition applications become more and more important. The research contents include facial image [7–9], speech [10, 11], body gesture [12, 13], and physiological signal [14] based gender recognition, among which gender recognition using physiological signals is more reliable but more difficult for data acquisition and analysis.

In practice, automatic gender recognition is a two-class classification problem. With little prior knowledge, massive number of features will be extracted from the raw data. Searching for an optimal feature subset from a high dimensional feature space is known to be an NP-complete problem. As a key issue in machine learning and related fields, feature selection (FS) is used to select a better feature combination from many solutions, the essence of which is combinatorial optimization. Wrapper feature selection method, which utilizes the learning machine of interest as a black box to score subsets of feature according to their predictive power [15], has shown its superior performance in various machine learning applications.

In this paper, we propose a gender recognition method from multiple physiological signals, in particular, we developed a wrapper algorithm based on Adaboost.M1 [16] and sequential backward selection (SBS) for physiological feature selection. Through the data acquisition, feature extraction and feature selection & gender recognition procedure, we obtained a prediction accuracy of 91.1% on a dataset of 234 participants, and we also find a subset of 12 features which can best represent our gender recognition model.

## 2 Materials and Methods

The proposed physiological-signal based gender recognition system is composed of three core components: Data Collection module, Feature Extraction module and Feature Selection & Classification module.

### 2.1 Data Acquisition

234 students from Southwest University with no history of cardiac disease and mental disease voluntarily participated in the test. The electrocardiogram (ECG), electromyogram (EMG), respiratory (RSP) and galvanic skin response (GSR) signals are collected with BIOPAC System MP150 from the subject's wrist, facial muscle, chest and fingers, respectively. The sampling rates are 200 Hz for ECG, 1000 Hz for EMG, 100 Hz for RSP, and 20 Hz for GSR.

234 groups (154 female vs. 80 male samples) of valid data were obtained, and each signal record is an 80-s fragment. Figure 1 illustrates the raw signals of ECG, EMG, RSP and GSR from one participant.

### 2.2 Feature Extraction

The raw physiological signals are firstly preprocessed using wavelet transform. A bunch of statistical features such as maximum, minimum, mean and standard deviation are then extracted from the preprocessed signals as well as different transformations of the signals. The raw features are extracted mainly by the AuBT Biosignal Toolbox [17]. The details of the features can be found on our website [http://hpc.siat.ac.cn/~hlzhang/GR/193\\_features.html](http://hpc.siat.ac.cn/~hlzhang/GR/193_features.html).

We have 234 samples with 84 ECG features, 21 EMG features, 67 RSP features and 21 GSR features, resulting in a raw data matrix of size 234 \* 193. The value for

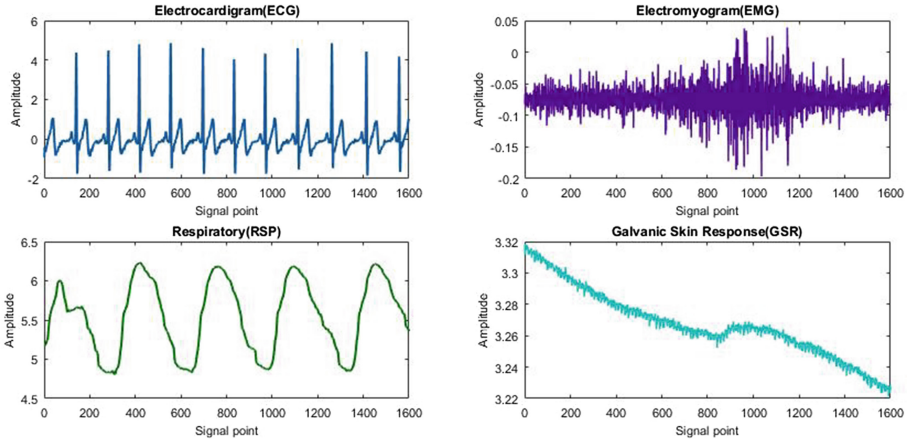


Fig. 1. Examples of raw ECG, EMG, RSP and GSR signals

each feature position is then normalized by Z-score to enable faster convergence of our feature selection algorithm.

### 2.3 Feature Selection and Gender Classification

The feature selection and classification algorithm, a wrapper method combining *ada*-boost and SBS, is outlined in Fig. 2.

In this paper, classification and regression trees (CART) [18] is used as weak classifier of *Ada*boost. Assuming that the data record number is  $m$  and the feature dimension is  $n$ , the time complexity of CART is  $O(nm \log m)$  ( $\log m$  is the depth of tree and  $O(nm)$  is the computational complexity of each layer), the iteration number of the SBS procedure (*while* loop in Algorithm Boost\_FS) is  $n-c$  ( $c$  is a constant determined by line 17 in the algorithm), and the time complexity of quick-sort (feature importance as indicated by line 12 in the algorithm) is  $O(n \log n)$ . For a  $v$ -fold ( $v$  is a constant, which is 20 in this paper) cross validation, the computational complexity of the proposed algorithm is:

$$O((n - c) * (v * O(knm \log m) + O(n \log n) + n - c)) \approx O(kn^2 m \log m) \quad (1)$$

where  $k$  is the number of trees used.

As seen from Eq. (1), the computational complexity of Algorithm Boost\_FS grows as a quadratic function of feature dimension and as a  $m \log m$  function of data record numbers, which demonstrates the scalability of the algorithm for big data applications.

In this paper, we call the feature subset found by Boost\_FS algorithm with highest prediction accuracy as the best feature subset.

**Algorithm Boost\_FS****Input:** Data Set  $S$  with  $N$  features**Output:** Optimal feature subset**Steps:**

```

01   $ACC_{Glob\_Max} = 0$ 
02   $ft\_num = N$ 
03  while ( $ft\_num > 0$ )
04       $S[1..v] \leftarrow \text{Divide}(S, v)$  /*Randomly divide  $S$  into  $v$  equal parts*/
05       $ACC_{Loc\_Max} = 0$ 
06       $ACC_{Loc}[1 : v] \leftarrow 0$ 
07      for ( $i$  in  $1 : v$ )
08           $Model \leftarrow \text{Adaboost.M1.train}(S[i], -i)$ 
09           $Acc\_Loc[i] \leftarrow \text{Adaboost.M1.test}(S[i])$ 
10          if ( $ACC_{Loc\_Max} \leq ACC_{Loc}[i]$ )
11               $ACC_{Loc\_Max} = ACC_{Loc}[i]$ 
12               $Feat\_Order \leftarrow \text{Feature\_Importance}(Model)$ 
13       $ACC_{Loc\_Mean} \leftarrow \text{Mean}(ACC_{Loc})$ 
14      if ( $ACC_{Glob\_Max} \leq ACC_{Loc\_Mean}$ )
15           $ACC_{Glob\_Max} = ACC_{Loc\_Mean}$ 
16           $Feat\_Order\_Glob = Feat\_Order$ 
17       $(S, n) \leftarrow \text{Remove\_feature}(S, Feat\_Order)$  /*Remove  $n$  least important feature(s) with the same score */
18       $ft\_num = ft\_num - n$ 
19  return  $Feat\_Order\_Glob$ 

```

**Fig. 2.** The wrapper algorithm Boost\_FS**2.4 Evaluation Metrics**

We use various metrics, such as accuracy, precision, recall (also known as sensitivity), specificity, F1 score and ROC curve [19], to measure the quality of the prediction results. Shown as below are the definitions of accuracy, precision, recall and specificity values:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (2)$$

$$Precision = TP / (TP + FP) \quad (3)$$

$$Recall = Sensitivity = TP / (TP + FN) \quad (4)$$

$$Specificity = TN / (TN + FP) \quad (5)$$

where  $TP$ ,  $TN$ ,  $FP$  and  $FN$  represent the number of true positives, true negatives, false positives and false negatives.

*F1* is the harmonic average of the precision and recall calculated as follows:

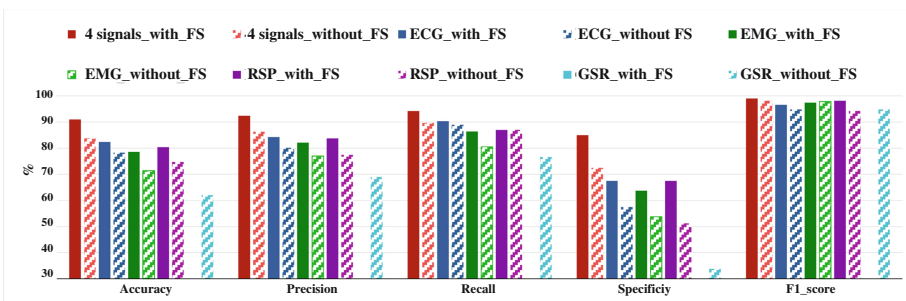
$$F1 = 2 * Precision / (Precision + Recall) \tag{6}$$

We also draw the Receiver Operating Characteristic (ROC) curve and calculate the area under this curve (AUC) for performance evaluation of our gender recognition models. The ROC curve, which is defined as a plot of test Sensitivity as the y coordinate versus its 1 - Specificity as the x coordinate, is an effective method of evaluating the performance of classification models. The AUC value, ranging from 0 to 1, shows the stability and performance of a model. An AUC value of 0 indicates a perfectly inaccurate test and a value of 1 reflects a perfectly accurate test.

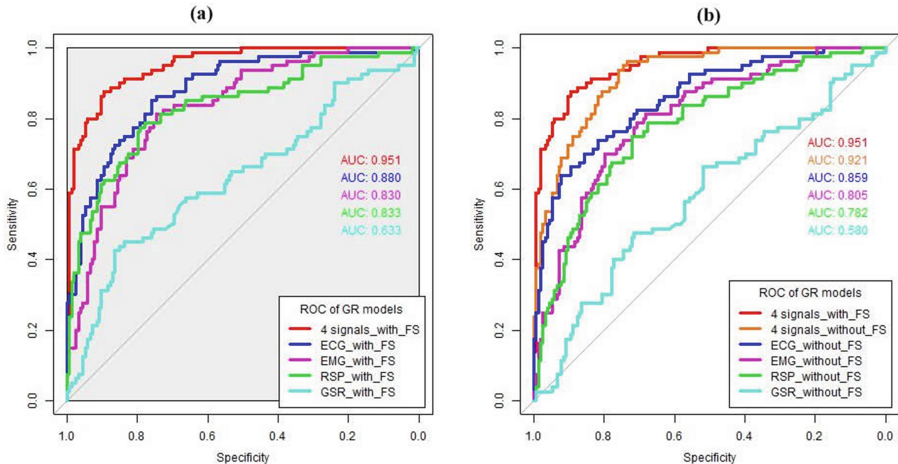
### 3 Results and Analysis

In order to evaluate the performance of the proposed method, we employ a 20-fold cross-validation scheme. The dataset is divided into 20 folds with approximately 11/12 samples in each fold. 19 folds are used for training the gender recognition model, and the remaining fold is used for testing. We have 5 individual runs of BOOST\_FS, with different input feature matrix extracted from different physiological signals. Each iteration (*while* loop in Algorithm BOOST\_FS in Fig. 2) in the run generates a feature subset and the corresponding evaluation metrics. The best feature subset in each run is the subset with the highest prediction accuracy.

Figures 3, 4 and Table 1 show the overall performance of different models with or without the feature selection. From left to right, Fig. 3 illustrates the prediction accuracies, precisions, recalls, specificities and F1\_scores using different gender recognition models. The model from 4 signals (ECG/EMG/RSP/GSR) with FS achieves the highest performance for all metrics: 91.1% accuracy, 92.4% precision, 94.2% recall, 85% precision and 99.0% F1\_socre. The GSR based model without FS shows the worst performance for all metrics except for the F1\_socre. Figure 4 shows the ROC curves of tests with and without using BOOST\_FS. 4\_signals\_with\_FS shows the highest AUC value of 0.951, followed by 4\_signals\_without\_FS of 0.921, ECG\_with\_FS of 0.880 and RSP\_with\_FS of 0.833, which are consistent with the results and analysis in Fig. 3.



**Fig. 3.** Performance of accuracy, precision, recall, specificity, F1\_socre for 10 different recognition models



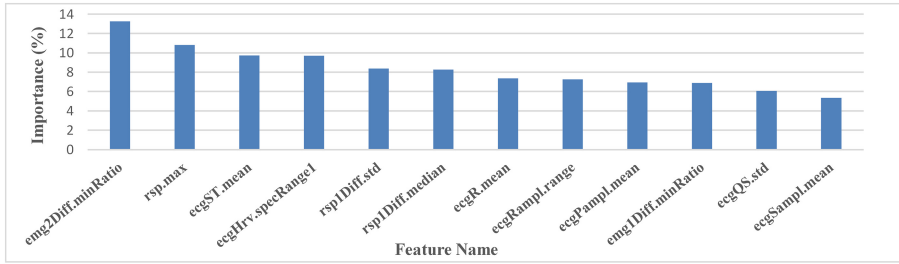
**Fig. 4.** ROC curve for (a) recognition models with FS; (b) recognition models without FS

**Table 1.** Prediction accuracies and feature numbers for different gender recognition models

Physiological signal	Accuracy without FS	Acc with FS	Total feature number	Best feature number
ECG/EMG/RSP/GSR	83.7%	91.1%	193	12
ECG	78.2%	82.3%	84	11
EMG	71.4%	78.8%	21	12
RSP	74.8%	80.8%	67	5
GSR	61.8%	69.1%	21	2

The feature number of the best subset for each type of signal and the corresponding prediction accuracy are tabulated in Table 1. For comparison, we also list the prediction accuracies and original feature numbers without FS. For 5 group of physiological signals shown in Table 1, the prediction accuracies using FS are increased by 7.4%/4.1%/7.4%/6%/7.3% compared with those without FS, correspondingly, feature numbers using FS are reduced by 181/73/9/62/19. The highest prediction accuracy from BOOST\_FS is 91.1% with 12 features from the combined ECG/EMG/RSP/GSR data.

Figure 5 shows the feature names and feature importances in the best subset determined by the BOOST\_FS algorithm. Detail information of the features can be found in Sect. 2.2 and on our website. 7 out of the 12 features selected from massive computational efforts are ECG features, and 3 are RSP features, which are reasonable since previous studies have reported the physiological difference between men and women in cardiac [20] and thoraco-abdominal [21] functions.



**Fig. 5.** Feature Importance of 12 features chosen by the BOOST\_FS algorithm

## 4 Conclusions

In this work, we introduced an automated physiological signal based gender recognition system. We observed that a model built from multiple physiological signals can outperform model based on a single physiological signal. We further showed that recognition performance can be improved obviously through using a wrapper feature selection procedure. Finally, we analyzed the best feature subset which can best represent gender differences. The future work would concentrate on developing more effective feature selection algorithm, taking the effect of human age into account and applying our gender recognition system to human machine interface.

**Acknowledgement.** This work is supported by National Science Foundation of China under grant no. U1435215 and 61433012, Guangdong Provincial Department of Science and Technology under grant No. 2016B090918122, the Science Technology and Innovation Committee of Shenzhen Municipality under grant No. JCYJ20160331190123578, and No. GJHZ20170314154722613, Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund under Grant No. U1501501, and Youth Innovation Promotion Association, CAS to Yanjie Wei.

## References

1. Ludwig, S., Oertelt-Prigione, S., Kurmeyer, C., Gross, M., Grüters-Kieslich, A., Regitz-Zagrosek, V., Peters, H.: A successful strategy to integrate sex and gender medicine into a newly developed medical curriculum. *J. Women's Health* **24**, 996–1005 (2015)
2. Canevelli, M., Quarata, F., Remiddi, F., Lucchini, F., Lacorte, E., Vanacore, N., Bruno, G., Cesari, M.: Sex and gender differences in the treatment of Alzheimer's disease: a systematic review of randomized controlled trials. *Pharmacol. Res.* **115**, 218–223 (2017)
3. Bonetto, M., Korshunov, P., Ramponi, G., Ebrahimi, T.: Privacy in mini-drone based video surveillance. In: 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Slovenia, pp. 1–6. IEEE (2015)
4. Venetianer, P.L., Lipton, A.J., Chosak, A.J., Frazier, M.F., Haering, N., Myers, G.W., Yin, W., Zhang, Z., Cutting, R.: Video surveillance system employing video primitives. Google Patents (2018)
5. Rukavina, S., Gruss, S., Hoffmann, H., Tan, J.-W., Walter, S., Traue, H.C.: Affective computing and the impact of gender and age. *PLoS ONE* **11**, e0150584 (2016)

6. Han, H., Otto, C., Liu, X., Jain, A.K.: Demographic estimation from face images: Human vs. machine performance. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 1148–1161 (2015)
7. Bekios-Calfa, J., Buenaposada, J.M., Baumela, L.: Robust gender recognition by exploiting facial attributes dependencies. *Pattern Recogn. Lett.* **36**, 228–234 (2014)
8. Levi, G., Hassner, T.: Age and gender classification using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, pp. 34–42. IEEE (2015)
9. Dantcheva, A., Brémond, F.: Gender estimation based on smile-dynamics. *IEEE Trans. Inf. Forensics Secur.* **12**, 719–729 (2017)
10. Pahwa, A., Aggarwal, G.: Speech feature extraction for gender recognition. *Int. J. Image Graph. Sig. Process.* **8**, 17 (2016)
11. Li, M., Han, K.J., Narayanan, S.: Automatic speaker age and gender recognition using acoustic and prosodic level information fusion. *Comput. Speech Lang.* **27**, 151–167 (2013)
12. Lu, J., Wang, G., Moulin, P.: Human identity and gender recognition from gait sequences with arbitrary walking directions. *IEEE Trans. Inf. Forensics Secur.* **9**, 51–61 (2014)
13. Cao, L., Dikmen, M., Fu, Y., Huang, T.S.: Gender recognition from body. In: Proceedings of the 16th ACM International Conference on Multimedia, Vancouver, Canada, pp. 725–728. ACM (2008)
14. Hu, J.: An approach to EEG-based gender recognition using entropy measurement methods. *Knowl.-Based Syst.* **140**, 134–141 (2018)
15. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182 (2003)
16. Schapire, R.E., Freund, Y.: *Boosting: Foundations and Algorithms*. Adaptive Computation and Machine Learning Series. The MIT Press, Cambridge (2012)
17. AuBT. <https://www.informatik.uni-augsburg.de/lehrstuehle/hcm/projects/tools/aubt/>. Accessed 25 Feb 2013
18. Steinberg, D., Colla, P.: CART: classification and regression trees. In: *The Top Ten Algorithms in Data Mining*, vol. 9, p. 179 (2009)
19. Powers, D.M.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Int. J. Mach. Learn. Technol.* **2**, 37–63 (2011)
20. Alfakih, K., Walters, K., Jones, T., Ridgway, J., Hall, A.S., Sivananthan, M.: New gender-specific partition values for ECG criteria of left ventricular hypertrophy: recalibration against cardiac MRI. *Hypertension* **44**, 175–179 (2004)
21. Romei, M., Mauro, A.L., D'angelo, M., Turconi, A., Bresolin, N., Pedotti, A., Aliverti, A.: Effects of gender and posture on thoraco-abdominal kinematics during quiet breathing in healthy adults. *Respir. Physiol. Neurobiol.* **172**, 184–191 (2010)