Modern Management based on Big Data II and Machine Learning and Intelligent Systems III
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA210301

Unreadable Segment Recognition of Single-Lead ECG Signals Based on XGBoost: Fuision of Shannon Energy Envelope and Empirical Mode Decomposition

Hanshuang XIE[a,1], Jiayi YAN[a], Huaiyu ZHU[b], Qineng CAO[a], Yamin LIU[a], Yun PAN[b] and Fan WU[a]
[a] *Research and Development Department, Hangzhou Proton Technology, Co., Ltd., Hangzhou, China*
[b] *College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, China*

Abstract. The quality of ECG signals is commonly affected by severe noise, especially for the single-lead ECG signals acquired from long-term wearable devices. Recognizing and ignoring these interfered signals can reduce the error rate of automatic ECG analysis system, and in addition, improve the efficiency of cardiologists. Based on XGBoost classifier, we propose an unreadable ECG segment recognition method using features extracted through Shannon Energy Envelope (SEE) and Empirical Mode Decomposition (EMD). An unreadable CarePatch™ ECG patch database is established, containing 8169 readable segments and 6114 unreadable segments with a length of 10 seconds. The XGBoost with 5-fold cross-validation is applied and obtained an accuracy of 99.51+/-0.15%. In conclusion, SSE and EMD features contribute to the unreadable segments recognition and alleviate the misdiagnosis of abnormal rhythms.

Keywords. Unreadable segment, single-lead ECG, wearable devices, XGBoost

1. Introduction

With the development of m-Health technology, wearable ECG monitoring devices of various physical forms such as card-type, patch-type, watch-type, etc. have appeared, which are widely applied to various clinical diagnosis scenes. By extending the monitoring time and scenarios, wearable ECG devices with limited lead channels compensate for diseases diagnose that are easily missed by resting 12-lead ECG and traditional 24-hour Holter, such as cryptogenic syncope, palpitation, paroxysmal atrial fibrillation, transient arrhythmia [1]. However, as daily activities are inevitable in long-term wearing, ECG signals collected by wearable devices are severely interfered by

[1] Corresponding Author, Hanshuang XIE, Research and Development Department, Hangzhou Proton Technology, Co., Ltd., Room 405, Building #8A, East Software Park, No. 90 Wensan Road, Xihu District, Hangzhou, China; E-mail: hanshuang.xie@protontek.com.

various noise, which will reduce the quality of signals and the work efficiency of cardiologists. Although shielding noisy segments can promote the performance of classification algorithm and reduce the false alarm rate [2], deleting such noise brutally may lead to false recognition of some diseases such as arrhythmia. Hence, to retain the effective signals as much as possible, only segments with R peaks completely overwhelmed by noise are regarded as unreadable segments.

Some early published works have proposed several generally applied ECG signal quality evaluation and noise estimation methods. Lee *et al.* [3] applied empirical mode decomposition (EMD) for extracting the first-order intrinsic mode function (F-IMF). However, the data they used is obtained from healthy persons, which cannot reflect the influence of noise on the misdiagnosis of arrhythmia. Li *et al.* [4] experimented with 13 signal quality indexes (SQIs) and their combinations from 12-lead ECG data based on support vector machine (SVM). Zhang *et al.* [5] adopted pSQI, kSQI and basSQI as indicators and confirmed that the single-lead ECG signal evaluation algorithm is capable of signal acquisition, denoising and QRS extraction. Zhao *et al.* [6] executed simple heuristic fusion to extract four SQIs and established the fuzzy vector for classification and assessment. Satija *et al.* [7] took autocorrelation function features into consideration. Moeyersons *et al.* [8] extracted three features from the autocorrelation function (ACF) and fed them into a RUSBoost classifier. The autocorrelation function features above can improve the assessment results effectively, but this time-consuming method presents challenges in long-term real-time ECG signal quality evaluation task. Zaunseder *et al.* [9] adopted 35 simple spectrum features, and obtained balance between accuracy and computational complexity with random forest (RF) classifier on 12-lead ECG data. They also stated that a more accurate classification is not feasible using their feature space. Zhang *et al.* [10] presented encoding Lempel-Ziv complexity (ELZC) method for feature extraction, combining other linear and nonlinear features and have proved that PE and ELZC features contribute most to noise recognition. Orphanidou *et al.* [11] utilized wavelet entropy features based on heart rate variability and SVM for signal quality classification, and verified the generalization of their model on different sing-lead ECG sensors.

In this paper, we present an unreadable segments recognition method of single-lead ECG signals based on XGBoost classifier, propose a new feature set extracted from Shannon Entropy Envelope, combined the features with that those extracted from time-domain and EMD. We also established an unreadable CarePatch$^{\text{TM}}$ ECG patch database using the data collected by the single-lead wearable patch devices. We conduct simulation experiments on the constructed dataset and achieve an accuracy of 99.51+/-0.15%.

2. Method

The proposed method as shown in Figure 1 contains the following steps: 1) divide ECG signals into 10 seconds' fragments; 2) filter signals with finite impulse response (FIR) band-pass filter with a cut-off frequency range between 0.67 Hz and 40 Hz; 3) for filtered signals SigFir, apply morphological opening and closing operations and thereafter, obtain Shannon Energy Envelope (SEE); On the other hand, apply EMD and obtain the first IMF component; 4) extract relevant features from SEE and EMD respectively; 5) train the XGBoost classifier; 6) detect the unreadable segments of ECG signals.

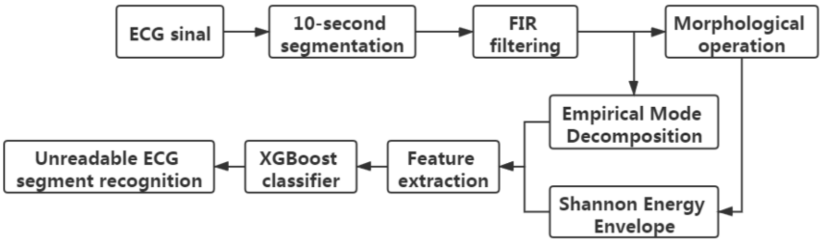

Figure 1. The block diagram of the proposed method.

2.1. Pre-processing

In his paper, we adopt FIR band-pass filter (0.67Hz ~ 40Hz) as a basic filter to remove the noise [12]. It is confirmed that the FIR band-pass filter with a cut-off frequency of 0.67 Hz can handle baseline wander well, and signals with a frequency lower than 40 Hz retain most of the useful information [13].

2.2. Features

2.2.1. Feature Selection

By expanding earlier published works [3-6], we extract 28 signal quality indicators for unreadable segments recognition, as shown in Table 1.

Table 1. Features extracted from the proposed methods.

Methods	Feature Name	Dimension	Methods	Feature Name	Dimension
sigFir[4-6]	Kurtosis	1	SEE	meanAmp	1
	Skewness	1		stdAmp	1
	Entropy	1		ratioMeanSdt	1
	validAmplitude	1		first8PeaksMeanstdRatio	1
	invalidRpeaksRatio	1		first8PeaksMean	1
EMD[3]	Mean of IMF1	3		first8PeaksStd	1
	Std of IMF1	3		first5PeaksMeanstdRatio	1
	Zero crossing rate	3		first5PeaksMean	1
	Shannon entropy	3		first5PeaksStd	1
				histCountRatio	1
				Large2mvRatio	1

2.2.2. Shannon Energy Envelope (SSE)

In this paper, Shannon Energy Envelope related features are adopted innovatively. The filtered signal SigFir is put through imopen and imclose operations to obtain sigMor, and then the Shannon envelope SigShannon is computed. Finally, 11 relevant features are extracted for the classification of unreadable segments. The steps are as follows:

1) Apply band-pass filter with a cut-off frequency of 6-18 Hz, retain the energy of the QRS complex and obtain sigFilter;
2) Calculate the difference of the signal and normalize it to get sigDiffNor;
3) Compute Shannon entropy and obtain sigShannonEntropy;
4) Apply mean filtering with a window length of 0.18s twice and obtain sigAvg;
5) Calculate the difference of sigAvg and normalize it, obtain the final Shannon envelope sigShannon;

6) Extract relevant features from sigShannon.

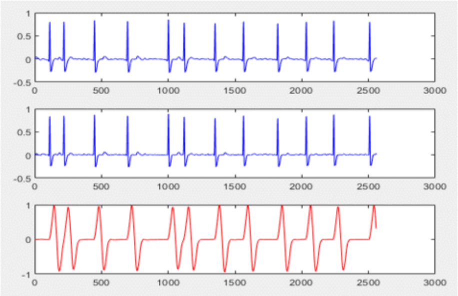

Figure 2. The sigShannon of readable segments (original, filtered and Shannon envelope)

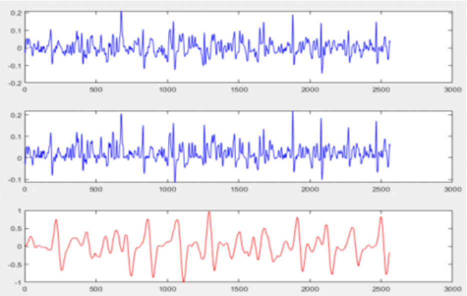

Figure 3. The sigShannon of unreadable segments (original, filtered and Shannon envelope)

Examples of the sigShannon of readable and unreadable segments are shown in Figure 2 and Figure 3.

We further extracted various features from the Shannon envelope sigShannon and finally selected 11 from them (as shown in Table 1) which have significant differences in the distribution of readable and unreadable segments respectively. The intuitively differences in feature distribution are shown in Figure 4 and Figure 5.

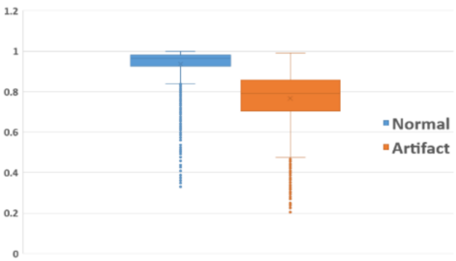

Figure 4. The distribution of first5PeaksMean feature on readable and unreadable segments

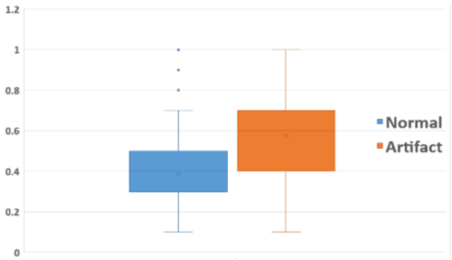

Figure 5. The distribution of histCountRatio feature on readable and unreadable segments

2.3. Modelling

In this paper, XGBoost is applied for classification, with boosting rounds n_estimators = 101, learning rate = 0.5, maximum tree depth max_depth=10, minimum sum of instance weight needed in a child min_child_weight = 8, nthread = 1, subsample = 0.85.

3. The Unreadable CarePatch^{TM} ECG patch database

Due to the users' daily activities, single-lead ECG data collected by long-term wearable devices usually contain more forms of heartbeats and more complex noises, which poses a great challenge to automatic ECG signal analysis. However, it is difficult to make a performance comparison of unreadable segments recognition algorithms on single-lead ECG because of the lack of a long-term single-lead ECG database. Therefore, constructing a long-term unreadable segments of single-lead ECG database is of great significance for the algorithm to assist in unreadable segments recognition tasks.

In this study, we firstly construct a long-term single-lead ECG data set. Then we conduct the proposed unreadable segments recognition algorithm and the other recognition algorithms on the dataset for a comparison. This self-build data set is collected and established by NMPA-certified CarePatch™ patch devices (NMPA#ZJ20202070050), containing 14283 single-lead ECG data segments, with 8169 readable ones and 6114 unreadable ones respectively. Each segment is of 10 seconds long, and the distribution of the training set and test set is 12854 (5492 unreadable ones, 7362 readable ones) and 1429 (622 unreadable ones, 807 readable ones) respectively. This unreadable CarePatch™ ECG patch database is collected from 499 patients' single-lead ECG recordings from July, 2017 to April, 2020. These patients were requested to wear CarePatch™ patch devices for 7 days. The patch device collected non-standard lead ECG signals from up left of the chest with a 256 Hz sampling rate and 12-bit resolution [14].

4. Results

4.1. Results on Test Set

We divide the total data set into training set (90%) and test set (10%), and with 10-fold cross-validation on the training set, we obtained an average accuracy of 99.51+/-0.15%. The generalization performance on the test set is presented with ROC curve depicted in Figure 6 (AUC=0.9999), the confusion matrix presented in Figure 7.

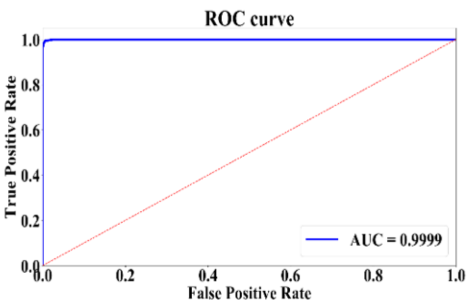

Figure 6. ROC curve on the test set.

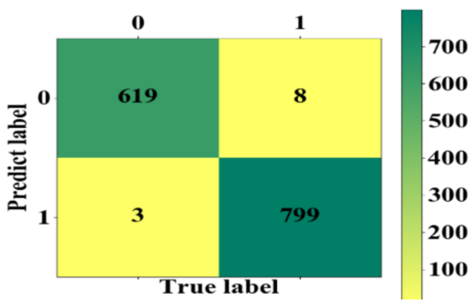

Figure 7. Confusion matrix on the test set

4.2. Results for Different Feature Combinations

To further identify the effectiveness of the proposed methods, we compared our features and the other feature set by combining SigFir features, EMD features and SEE features as illustrated in Table 2. The results are all obtained by XGBoost classifier, and we can see that the feature combination adopted by this paper contributes to the best performance on the unreadable segment recognition task.

Table 2. The performance of XGBoost on different feature combinations.

Feature set	Precision%	Recall%	F1
SigFir	97.27	97.27	0.973
EMD	98.74	98.74	0.987
Shannon envelope	97.62	97.62	0.976
SigFir+EMD	98.88	98.88	0.989

SigFir+Shannon envelope	99.09	99.09	0.991
EMD+Shannon envelope	98.96	98.95	0.990
This paper	99.23	99.23	0.992

5. Conclusion

The contributes of this paper can be summarized as followings:

(1) Proposed an unreadable segment recognition of single-lead ECG signals based on XGBoost with the fusion of Shannon energy envelope features, empirical mode decomposition features and other time-domain features, which outperforms the state-of-the-art methods.

(2) Constructed an unreadable CarePatch^{TM} ECG patch database based on the medical-level single-lead ECG monitoring devices, which contains 7-day ECG signals and along with various interference and noise. The construction of such database is of significance to evaluate the unreadable segment recognition algorithm.

The training and testing code developed in this work is publicly available in the repository[2].

References

[1] Chen XL, Liu M, Yang J. Clinical application of single lead long-term ambulatory electrocardiography. J. Practical Electrocardiol., 2018; 27(04): 269-272+278.

[2] Abdelazez M, Quesnel PX, Chan AD, Yang H. Signal quality analysis of ambulatory electrocardiograms to gate false myocardial ischemia alarms. IEEE Trans. Biomed. Eng. 2016 Aug 25; 64(6):1318-25.

[3] Lee J, McManus DD, Merchant S, Chon KH. Automatic motion and noise artifact detection in Holter ECG data using empirical mode decomposition and statistical approaches. IEEE Trans. Biomed. Eng. 2011 Nov 10;59(6):1499-506.

[4] Li Q, Rajagopalan C, Clifford GD. A machine learning approach to multi-level ECG signal quality classification. Comput. Meth. Programs Biomed. 2014 Dec 1;117(3):435-47.

[5] Zhang Y, Zhao Z. Evaluation of single-lead ECG signal quality with different states of motion. In2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI) 2017 Oct 14 (pp. 1-7). IEEE.

[6] Zhao Z, Zhang Y. SQI quality evaluation mechanism of single-lead ECG signal based on simple heuristic fusion and fuzzy comprehensive evaluation. Front. Physiol. 2018 Jun 14; 9:727.

[7] Satija U, Ramkumar B, Manikandan MS. An automated ECG signal quality assessment method for unsupervised diagnostic systems. Biocybern. Biomed. Eng. 2018 Jan 1;38(1):54-70.

[8] Moeyersons J, Smets E, Morales J, Villa A, De Raedt W, Testelmans D, Buyse B, Van Hoof C, Willems R, Van Huffel S, Varon C. Artefact detection and quality assessment of ambulatory ECG signals. Comput. Meth. Programs Biomed. 2019 Dec 1; 182: 105050.

[9] Zaunseder S, Huhle R, Malberg H. CinC challenge—Assessing the usability of ECG by ensemble decision trees. In2011 Computing in Cardiology 2011 Sep 18 (pp. 277-280). IEEE.

[10] Zhang Y, Wei S, Zhang L, Liu C. Comparing the performance of random forest, SVM and their variants for ECG quality assessment combined with nonlinear features. J. Med. Biol. Eng. 2019 Jun;39(3):381-92.

[11] Orphanidou C, Drobnjak I. Quality assessment of ambulatory ECG using wavelet entropy of the HRV signal. IEEE J. Biomed. Health Inform. 2016 Oct 5;21(5):1216-23.

[12] Venkatachalam KL, Herbrandson JE, Asirvatham SJ. Signals and signal processing for the electrophysiologist: part I: electrogram acquisition. Circ.-Arrhythmia Electrophysiol. 2011 Dec;4(6):965-73.

[2] https://gitee.com/xieHanshuang/unreadable-segment-recognition.git

[13] Romero FP, Romaguera LV, Vázquez-Seisdedos CR, Costa MG, Neto JE. Baseline wander removal methods for ECG signals: A comparative study. arXiv preprint arXiv:1807.11359. 2018 Jul 30.
[14] Zhu H, Pan Y, Wu F, Huan R. Optimized electrode locations for wearable single-lead ECG monitoring devices: A case study using WFEES modules based on the LANS method. Sensors. 2019 Jan;19(20):4458.