Journal of Medical Imaging and Health Informatics
Vol. 10, 2663–2667, 2020

An Open-Access Long-Term Wearable ECG Database for Premature Ventricular Contractions and Supraventricular Premature Beat Detection

Zhipeng Cai[1], Chengyu Liu[1,*], Hongxiang Gao[1], Xingyao Wang[1], Lina Zhao[1], Qin Shen[2], E. Y. K. Ng[3], and Jianqing Li[1,4,*]

[1]*The State Key Laboratory of Bioelectronics, School of Instrument Science and Engineering, Southeast University, Nanjing, 210096, China*
[2]*The First Affiliated Hospital of Nanjing Medical University, Nanjing, 210029, China*
[3]*School of Mechanical and Aerospace Engineering, Nanyang Technological University, 639798, Singapore*
[4]*School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, 211166, China*

Wearable electrocardiogram (ECG) devices can provide real-time, long-term, non-invasive and comfortable ECG monitoring for premature beats (PB) assessment (typically presenting as premature ventricular contractions (PVC) and supraventricular premature beat (SPB)), which may foreshadow stroke or sudden cardiac death. However, the poor quality, introduced by the dry electrode in wearable ECG monitoring system, leads to the inefficient recognition of the existing PB detection technologies. Although many methods can achieve high recognition rate on current widely-used open-access clinical ECG databases, they still fail to work properly on dynamic ECG signals. This study presents an open-access ECG database comprises of 24-hour wearable ECG recordings. The database is used for the 3rd China Physiological Signal Challenge (CPSC 2020), where participants are expected to recognize PVC and SPB from these recordings. All the approved algorithms are evaluated by scoring standards and regulations defined in terms of PVC detection and SPB detection, respectively.

Keywords: Electrocardiogram (ECG), Premature Ventricular Contractions (PVC), Supraventricular Premature Beat (SPB), Database, CPSC.

1. INTRODUCTION

Electrocardiogram (ECG) is widely used by cardiologists as a standard tool for non-invasively monitoring and clinical diagnosis of cardiovascular disease. Arrhythmias are deadly heart diseases that cause more than 7 million deaths each year [1]. Premature beats (PBs), including supraventricular premature beat (SPB) or premature ventricular contractions (PVC), are the most common arrhythmias, manifested by alteration in RR interval duration and/or the QRS waveform, and the increased prevalence of these beats may be a precursor to stroke or sudden cardiac death [2]. Due to the randomness and uncertainty in the occurrence of PBs, the diagnosis of PBs needs observation of dynamic ECG during a long period [3]. Thanks to advances in wearable technology and the rising popularity of mobile healthcare, several remote, continuous and wearable ECG monitoring systems have been proposed to achieve comfortable, non-invasive ambulatory ECG signals in daily life [4]. The increase volume of ECG data has driven the need for automatic PB detection algorithms that can be used for early diagnosis and prevention of heart disease and reduce the manual interpretation work for physicians.

Although numerous ECG heart beat detection and classification algorithms have been developed based on different techniques, including ECG morphology and heartbeat interval [5, 6], principal component analysis (PCA) methods [7], redundant discrete wavelet transform (RDWT) [8] for feature extraction and convolutional neural network (CNN) [9], self-constructing neural fuzzy inference network (SoNFIN) [10], learning vector quantization (LVQ) neural network [11] and support vector machines (SVM) [12] for heartbeat classification. It is reported that the combination of these algorithms can achieve satisfactory performance in terms of classification accuracy. However, these algorithms are based on clinical ECG signals (such as MIT-BIH Arrhythmia Database or AHA Database [13, 14]), which are stationary without severe noises or artifact. It is still challenging to realize accurate classification in dynamic signals with severe noise and artifacts, especially for PB recognition in wearable ECG signals [11].

*Authors to whom correspondence should be addressed.

To improve the PB detection algorithms in dynamic ECG signals, the open-access databases play an important role. Although well-thought-out databases, such as the developed ECG database [15], heart sound database [16], electroencephalogram (EEG) database [17], 1st and 2nd CPSC [18, 19], etc., have been developed, there still lacks specialized dynamic database for testing PB detection algorithms. Thus, a well-designed wearable ECG database can adequately test the performance of PB detection algorithms is highly needed. This study provides a new ECG database containing long-term wearable ECG recordings from clinical arrhythmia patterns, to encourage the participants to develop more efficient and robust algorithms for PVC and SPB detection, like a game of 'Gold Rush.'

2. CHALLENGE DATA

Training data consists of 10 single-lead ECG recordings collected from patients with cardiovascular disease, each of the recording last for about 24 hours (shown in Table I). Test set contains similar ECG recordings of same lengths, which is unavailable to public and will remain private for the purpose of scoring for the duration of the Challenge and for some period afterwards. All data were collected by a unified wearable ECG device with a sampling frequency of 400 Hz, and provided in MATLAB format (each including two *.mat file: one is ECG data and another one is the corresponding SPB and PVC annotation file).

All the 24-hour wearable ECG recordings are challenging for PB detection. In general, there are two signal types of the challenging ECG recordings. We summarize the signal types as follows.

2.1. Type A: Pathological Arrhythmias

Usually, SPBs possess a normal QRS complex with irregular RR intervals, while PVCs present bizarre QRS complex with irregular RR intervals [20]. Similar to SPB, AF is characterized by normal QRS complex with irregular RR intervals and fibrillatory waves [21]. Then, traditional rule-based algorithms exhibit poor performance when distinguish SPB from AF [22]. In addition, the 24-hour wearable ECG recordings come from different channels, which indicates that the polarity of the PVC may be consistent with normal beats in some recordings and opposite to normal beats in other recordings. It is also worthy to note that the PBs occur in repeating patterns (bigeminy, trigeminy, quadrigeminy, couplet, etc.) in some recordings and there also exist multi-source PVCs in some recordings (see Fig. 1).

Table I. The detailed information of training data.

Recordings	AF (*Y*/*N*)*	Length (*h*)	# *N* beats	# *S* beats	# *V* beats	# Total beats
A01	*N*	25.89	109062	24	0	109086
A02	*Y*	22.83	98936	0	4554	103490
A03	*Y*	24.70	137249	0	382	137631
A04	*N*	24.51	77812	3466	19024	100302
A05	*N*	23.57	94614	25	1	94640
A06	*N*	24.59	77621	6	0	77627
A07	*N*	23.11	73325	3481	15150	91956
A08	*Y*	25.46	115518	0	2793	118311
A09	*N*	25.84	88229	1462	2	89693
A10	*N*	23.64	72821	9071	169	82061

Notes: * *Y* represents the current recording contains atrial fibrillation (AF) signals; *N* represents the current recording does not contain AF signal; *S* represents for supraventricular premature beat; *V* represents for premature ventricular contractions.

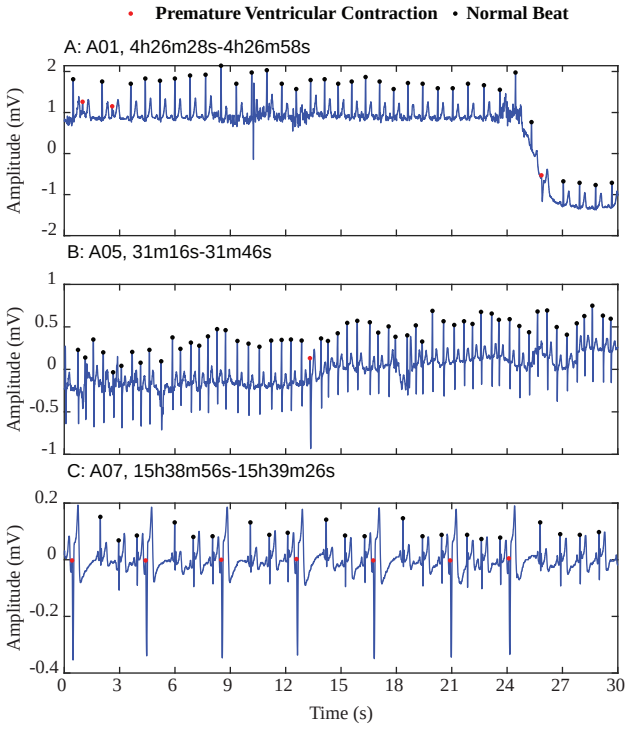

Fig. 1. Examples of ECG waveforms with different pathological arrhythmias. Red circles denote the reference location of PVC and black ones denote the reference location of normal beats. From top to bottom: (A) Multi-source PVCs; (B) PVC in AF and (C) ventricular trigeminy.

2.2. Type B: Poor Signal Quality Due to Artifact and Noise

Due to the poor contact and high skin-electrode impedance between dry electrodes and skin, wearable ECG signals collected by dry electrodes are easily contaminated by artifacts and noises [23, 24]. Unfortunately, whether using time-domain or frequency-domain denoising methods, these noises are difficult to remove. The reason is that their frequency content is overlapped with the frequency band of signal interest and they have similar morphologies to QRS complex [25]. The typical artifacts and noises (Fig. 2) are from:

(1) Electrode contact noise: Loss of contact between the electrode and skin manifesting as sharp changes with saturation on the ECGs (usually due to an electrode being nearly or completely pulled off).

(2) Electrode movement artifacts: Electrode movement away from the contact area on the skin, leading to variations in the impedance between the electrode and skin, which will cause potential variations in the ECG and usually manifest themselves as rapid (but continuous) baseline jumps or complete saturation.

(3) Device noise: Noises generated by the hardware of the device.

It is worthy to note that wearable ECG is often contaminated by noise in similar morphologies caused the interest signal nearly invisible by human eyes. To remove all noises completely is impossible, so it is important to quantify the nature of noises and choose an appropriate algorithm to evaluate the signal quality.

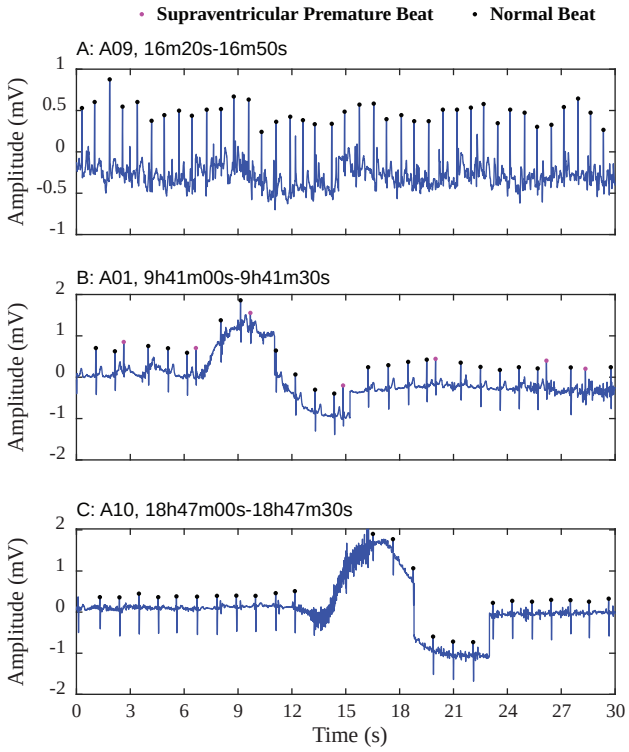

Fig. 2. Examples of poor signal quality ECG episodes. Pink circles denote the reference location of SPB and black ones denote the reference location of normal beats. The challenges for accurate PVC and SPB detection are from: (A) Device noise and electrode contact noise; (B) baseline wander noise and electrode movement artifacts interference and C) electrode contact noise and electrode sliding interference.

2.3. Annotation

The annotation process for all raw data consisted of four steps: signal quality detection, QRS complex location, QRS type determination, and manual review. Firstly, all raw ECG recordings were quickly scanned and the poor quality episodes (where the QRS complex could not be recognized by visual inspection) were labeled. Then, the ECG recordings, except for poor quality episodes, were beat-by-beat annotated first by the P&T QRS detector [26] and then manually hand-corrected to the peak of each QRS complex by visual inspection. Subsequently, each heartbeat marked with QRS complex position was determined to different beat type, and the positions of PVC and SPB were marked manually. Subsequently, manual review was performed by a single individual to correct any obvious mistake.

3. EVALUATION METHOD

CPSC 2020 is comprised of two events related to scoring: PVC detection and SPB detection. Only test set will be used for the event scoring. PVC and SPB annotations in the training and test sets are labeled and initially confirmed by cardiologists and trained volunteers. Score is calculated according to the following rules.

3.1. Event 1: PVC Detection

In this event, the goal is to generate a set of PVC annotations for each recording that can match the reference PVC annotations. For each reference PVC annotation, a matched PVC annotation should lie in 150 ms duration centered by the reference PVC annotation [27]. Noted that the reference PVC annotations appear in the first and last 0.2 seconds are ignored. Detected PVC should be within 150 ms from the reference ones. The scoring rules are:

- a false positive (*FP*) detection deduct 1 point.
- a false negative (*FN*) detection deduct 5 points, since from a clinical perspective, missed diagnosis is more serious than misdiagnosis, thus we penalize *FN* detection.

The final score for Event 1 (PVC_{err}) is the sum of all deducted points.

3.2. Event 2: SPB Detection

In this event, the goal is to generate a set of SPB annotations for each recording that can match the reference SPB annotations.

Table II. Summary of the major databases used for ECG signal analysis.

Database	# Recordings	Time	# ECG channels	Sampling frequency (Hz)	# QRS types	Other Information
American heart association ventricular arrhythmia database[a]	80	35-min/3-h	2	250	8	Beat-by-beat annotations
European ST-T database[a]	90	2-h	2	250	–	Beat-by-beat annotations
Fantasia database[a]	40	120-min	3	250	–	Beat-by-beat annotations
INCART[a]	75	30-min	12	275	10	Beat-by-beat annotations
Long-term-ST[a]	86	21∼24 h	2/3	250	–	Beat-by-beat annotations
MGH/MF waveform database	250	Varying lengths	3	360	–	Beat-by-beat annotations
MIT-BIH arrhythmia [a]	48	30-min	2	360	15	Beat-by-beat annotations
MIT ST change database[a]	28	Varying lengths	2	360	–	Beat-by-beat annotations
Noise stress test database[a]	15	30-min	2	360	–	Beat-by-beat annotations
PTB[a]	268	Varying lengths	14	1000	9	Beat-by-beat annotations
QT database[a]	105	15-min	2	250	–	Beat-by-beat annotations
Supraventricular arrhythmia database[a]	78	30-min	2	128	–	Beat-by-beat annotations
T-wave alternans database[a]	100	2-min	12/2/3	500	–	Beat-by-beat annotations
UCI machine learning: Arrhythmia dataset	452	24-h	12	–	16	Diagnostic labeling
1st CPSC	6877	Varying lengths	12	500	9	Diagnostic labeling
2st CPSC	2000	10-s	12	500	–	Beat-by-beat annotations

Notes: [a]From PhysioBank datasets [15] available at https://physionet.org/. # represents for the number of specific item.

For each reference SPB annotation, a matched SPB annotation should lie in 150 ms duration centered by the reference SPB annotation [27]. Noted that the reference SPB annotations appear in the first and last 0.2 seconds are also ignored. Detected SPB should be within 150 ms from the reference ones. The scoring rules are:

- a false positive (*FP*) detection deduct 1 point.
- a false negative (*FN*) detection deduct 5 points, since from a clinical perspective, missed diagnosis is more serious than misdiagnosis, thus we penalize *FN* detection.

The final score for Event 2 (SPB_{err}) is the sum of all deducted points.

4. DISCUSSION

In this paper, a brand-new database is proposed to facilitate the development of algorithms for searching PVC and SPB from 24-hour wearable single-lead ECG recordings. To data, there existed several well-annotated and validated ECG databases (see Table II) [15, 28–30] that can be used for the evaluation of PVC and SPB detection algorithms and test on these well-annotated and validated databases provide reproducible and comparable results. In addition, the scoring performances of some methods on these databases are too high due to the relatively good signal quality of ECG waveforms in these databases. However, the performances of these PVC and SPB detection algorithms are not tested on the strict setting even for rigorous conditions, such as wearable signals. We emphasize this by developing a new wearable ECG database for challenging PVC and SPB detection tasks, and by proposing new evaluation rules for algorithmic testing in the developed challenging ECG databases. CPSC 2020 contains 10 challenging 24-hour ECG recordings with manually annotated QRS types. This database includes signals from both pathological rhythm and artifacts, and is a real-world collected ECG database from the wearable device. We hope this strictly manual annotated database can benefit the study for dynamic ECG processing.

Conflicts of Interest Statement

The authors declare no conflict of interest.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (81871444), Key Research and Development Programs of Jiangsu Province (BE2017735). We thank the supports from ICBEB, 1st CPSC, 2nd CPSC, the State Key Laboratory of Bioelectronics in China and the Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent monitoring Lab. For more information on the 3rd CPSC Challenge and download the data, please visit the URL (http://www.icbeb.org/Challenge.html).

References and Notes

1. Ibrahim, N.E., Desai, A.S., Singh, J.P. and Januzzi, Jr. J.L., **2019**. Heart failure university 2015: Presentation summaries. *Reviews in Cardiovascular Medicine, 17*(S1), pp.1–8.
2. Alfaro-Ponce, M., Chairez, I. and Etienne-Cummings, R., **2019**. Automatic detection of electrocardiographic arrhythmias by parallel continuous neural networks implemented in FPGA. *Neural Computing and Applications, 31*(2), pp.363–375.
3. Al Rahhal, M.M., Al Ajlan, N., Bazi, Y., Al Hichri, H. and Rabczuk, T., **2018**. Automatic Premature Ventricular Contractions Detection for Multi-Lead Electrocardiogram Signal. *2018 IEEE International Conference on Electro/Information Technology (EIT)*, pp.169–173.
4. Liu, C., Zhang, X., Zhao, L., Liu, F., Chen, X., Yao, Y. and Li, J., **2018**. Signal quality assessment and lightweight QRS detection for wearable ECG SmartVest system. *IEEE Internet of Things Journal, 6*(2), pp.1363–1374.
5. Afkhami, R.G., Azarnia, G. and Tinati, M.A., **2016**. Cardiac arrhythmia classification using statistical and mixture modeling features of ECG signals. *Pattern Recognition Letters, 70*(1), pp.45–51.
6. Chen, S., Hua, W., Li, Z., Li, J. and Gao, X., **2017**. Heartbeat classification using projected and dynamic features of ECG signal. *Biomedical Signal Processing and Control, 31*(1), pp.165–173.
7. Zarei, R., He, J., Huang, G. and Zhang, Y., **2016**. Effective and efficient detection of premature ventricular contractions based on variation of principal directions. *Digital Signal Processing, 50*(1), pp.93–102.
8. Arrais Junior, E., Valentim, R.A.D.M. and Brandão, G.B., **2018**. Real-time premature ventricular contractions detection based on redundant discrete wavelet transform. *Research on Biomedical Engineering, 34*(3), pp.187–197.
9. Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M., Gertych, A. and Tan, R.S., **2017**. A deep convolutional neural network model to classify heartbeats. *Computers in Biology and Medicine, 89*(1), pp.389–396.
10. Liu, S., Cheng, D. and Lin, C., **2013**. Arrhythmia identification with two-lead electrocardiograms using artificial neural networks and support vector machines for a portable ECG monitor system. *Sensors, 13*(1), pp.813–828.
11. Liu, X., Du, H., Wang, G., Zhou, S. and Zhang, H., **2015**. Automatic diagnosis of premature ventricular contraction based on lyapunov exponents and LVQ neural network. *Computer Methods and Programs in Biomedicine, 122*(1), pp.47–55.
12. Khazaee, A. and Zadeh, A.E., **2014**. ECG beat classification using particle swarm optimization and support vector machine. *Frontiers of Computer Science, 8*(2), pp.217–231.
13. Hermes, R.E., Geselowitz, D.B. and Oliver, G., **1980**. Development, distribution, and use of the American heart association database for ventricular arrhythmia detector evaluation. *Computers in Cardiology, 7*(1), pp.263–266.
14. Moody, G.B. and Mark, R.G., **2001**. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine, 20*(3), pp.45–50.
15. Goldberger, A.L., Amaral, L.A., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C. and Stanley, H.E., **2000**. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation, 101*(23), pp.e215–e220.
16. Liu, C., Springer, D., Li, Q., Moody, B., Juan, R.A., Chorro, F.J., Castells, F., Roig, J.M., Silva, I. and Johnson, A.E., **2016**. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement, 37*(12), pp.2181–2183.
17. Hunter, M., Smith, R.L., Hyslop, W., Rosso, O.A., Gerlach, R., Rostas, J., Williams, D.B. and Henskens, F., **2005**. The Australian EEG database. *Clinical EEG and Neuroscience, 36*(2), pp.76–81.
18. Liu, F., Liu, C., Zhao, L., Zhang, X., Wu, X., Xu, X., Liu, Y., Ma, C., Wei, S., He, Z., Li, J. and Ng, E.Y.K., **2018**. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. *Journal of Medical Imaging and Health Informatics, 8*(7), pp.1368–1373.
19. Gao, H., Liu, C., Wang, X., Zhao, L., Shen, Q., Ng, E.Y.K. and Li, J., **2019**. An open-access ECG database for algorithm evaluation of QRS detection and heart rate estimation. *Journal of Medical Imaging and Health Informatics, 9*(9), pp.1853–1858.
20. Ye, C., Kumar, B.V. and Coimbra, M.T., **2012**. Heartbeat classification using morphological and dynamic features of ECG signals. *IEEE Transactions on Biomedical Engineering, 59*(10), pp.2930–2941.
21. Henriksson, M., García-Alberola, A. and Sörnmo, L., **2020**. Short-term reproducibility of parameters characterizing atrial fibrillatory waves. *Computers in Biology and Medicine, 117*(1), p.103613.
22. Jalali, A. and Lee, M., **2019**. Atrial fibrillation prediction with residual network using sensitivity and orthogonality constraints. *IEEE Journal of Biomedical and Health Informatics, 24*(2), pp.407–413.
23. Xiong, F., Chen, D., Chen, Z. and Dai, S., **2019**. Cancellation of motion artifacts in ambulatory ECG signals using TD-LMS adaptive filtering techniques. *Journal of Visual Communication and Image Representation, 58*(1), pp.606–618.
24. Berwal, D., Vandana, C.R., Dewan, S., Jiji, C.V. and Baghini, M.S., **2019**. Motion artifact removal in ambulatory ECG Signal for heart rate variability analysis. *IEEE Sensors Journal, 19*(24), pp.12432–12442.
25. Clifford, G.D., Azuaje, F. and Mcsharry, P., **2006**. ECG statistics, noise, artifacts, and missing data. *Advanced Methods and Tools for ECG Data Analysis, 6*(1), p.18.

26. Pan, J. and Tompkins, W.J., **1985**. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering, 32*(3), pp.230–236.
27. Association for the Advancement of Medical Instrumentation, **2013**. *ANSI/AAMI EC57: 2012-Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms*. American National Standard.
28. Taddei, A., Distante, G., Emdin, M., Pisani, P., Moody, G.B., Zeelenberg, C. and Marchesi, C., **1992**. The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *European Heart Journal, 13*(9), pp.1164–1172.
29. Van Bemmel, J.H. and Willems, J.L., **1990**. Standardization and validation of medical decision-support systems: The CSE project. *Methods of Information in Medicine, 29*(4), pp.261–262.
30. Laguna, P., Mark, R.G., Goldberg, A. and Moody, G.B., **1997**. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. *Computers in Cardiology, 1997*, pp.673–676.

Received: 24 April 2020. Revised/Accepted: 27 April 2020.