



AMERICAN  
SCIENTIFIC  
PUBLISHERS

Copyright © 2018 American Scientific Publishers  
All rights reserved  
Printed in the United States of America

*Journal of Medical Imaging and  
Health Informatics*  
Vol. 8, 1368–1373, 2018

# An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection

Feifei Liu<sup>1,2</sup>, Chengyu Liu<sup>1,\*</sup>, Lina Zhao<sup>1,4</sup>, Xiangyu Zhang<sup>1</sup>, Xiaoling Wu<sup>3</sup>, Xiaoyan Xu<sup>4</sup>, Yulin Liu<sup>4</sup>, Caiyun Ma<sup>4</sup>, Shoushui Wei<sup>4</sup>, Zhiqiang He<sup>5</sup>, Jianqing Li<sup>1,3</sup>, and Eddie Ng Yin Kwee<sup>6</sup>

<sup>1</sup>The State Key Laboratory of Bioelectronics, Jiangsu Key Lab of Remote Measurement and Control, School of Instrument Science and Engineering, Southeast University, Nanjing, 210096, China

<sup>2</sup>Shandong Zhong Yang Software Limited Company, Jinan, 250101, China

<sup>3</sup>School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, 211166, China

<sup>4</sup>School of Control Science and Engineering, Shandong University, Jinan, 250061, China

<sup>5</sup>Lenovo Research, Beijing, 100085, China

<sup>6</sup>School of Mechanical and Aerospace Engineering, Nanyang Technological University, 639798, Singapore

Over the past few decades, methods for classification and detection of rhythm or morphology abnormalities in ECG signals have been widely studied. However, it lacks the comprehensive performance evaluation on an open database. This paper presents a detailed introduction for the database used for the 1st China Physiological Signal Challenge 2018 (CPSC 2018), which will be run as a special section during the ICBE 2018. CPSC 2018 aims to encourage the development of algorithms to identify the rhythm/morphology abnormalities from 12-lead ECGs. The data used in CPSC 2018 include one normal ECG type and eight abnormal types. This paper details the data source, recording information, patients' clinical baseline parameters as age, gender and so on. Meanwhile, it also presents the commonly used detection/classification methods for the abovementioned abnormal ECG types. We hope this paper could be a guide reference for the CPSC 2018, to facilitate the researchers familiar with the data and the related research advances.

**Keywords:** Electrocardiogram (ECG), Database, Rhythm and Morphology Abnormal, CPSC.

## 1. INTRODUCTION

Cardiovascular diseases (CVDs) continue to be the leading cause of morbidity and mortality worldwide.<sup>1</sup> An estimated 17.5 million people died from CVDs in 2012, representing 31% of all global deaths (WHO 2015).<sup>2</sup> Electrocardiogram (ECG) signal is the expression of the myocardium electrical activity on the body's surface. It can provide important information about the status of cardiac activity<sup>3</sup> and is commonly used to detect rhythm/morphology abnormalities. The standard 12 lead ECGs has been an important tool for clinicians to diagnose heart diseases.

Over the past few decades, methods for classification and detection of rhythm or morphology abnormalities in ECG signals have been widely studied. Many methods have demonstrated potential to accurately detect pathologies in clinical applications.<sup>1</sup> Unfortunately, the current works lack comprehensive comparisons performed on as many as heart abnormal types. Existing

works focus on single or a few combination, such as atrial fibrillation,<sup>4,5</sup> while some studied ST changes.<sup>6,7</sup>

The China Physiological Signal Challenge (CPSC) 2018 (<http://www.icbeb.org/Challenge.html>) attempts to address this issue by contributing a more comprehensive database. The signals of this database (9,831 records from 9458 patients with a time length of 7–60 min) came from 11 hospitals, containing nine types: one normal ECG type and eight abnormal types. The purpose of this paper is to provide a detailed description for the challenge data and a concise introduction for the existing methods, to help researchers familiar with the data and the related research advances.

## 2. BRIEF REVIEW ON THE RHYTHM/ MORPHOLOGY ABNORMALITIES CLASSIFICATION METHODS

### 2.1. Atrial Fibrillation

Atrial fibrillation (AF) is a serious cardiovascular disease with the phenomenon of beating irregularly. In 2003, the European

\*Author to whom correspondence should be addressed.

Society of Cardiovascular Diseases and North American Association of Pacing and Electrophysiology Arrhythmia divided atrial fibrillation into paroxysmal atrial fibrillation, persistent atrial fibrillation and permanent atrial fibrillation based on the time of atrial fibrillation episode.<sup>8</sup> There are many well-known schools and institutions that have established large-capacity, multi-category, and well-annotated databases for AF studying work, such as the MIT-BIH database,<sup>9</sup> the AHA database,<sup>10</sup> and the ST-T ECG database.<sup>11</sup> The automatic AF detection is divided into two major categories: one based on the characteristics of atrial activity; one atrial fibrillation based on the characteristics of the RR interval. The success of the PhysioNet/Computing in Cardiology Challenge 2017 of “AF Classification from a short single lead ECG recording” significantly promotes the process of AF detection research.<sup>12</sup>

In 2000, Andrikopoulos et al. defined the square of the *P*-wave standard deviation in ECG signals as *P*-wave variability. This value could represent the changes of atrial activity characteristics in atrial fibrillation process. Using 12-lead ECG data from 60 normal subjects and 50 AF subjects to verify the performance, 88% sensitivity and 75% specificity were obtained.<sup>13</sup> Lepage et al.<sup>14</sup> used wavelet transform and Hidden Markov Models (HMM) to extract the *P* waves from 63 normal and 82 AF patient records. Specificity and sensitivity obtained were 65% and 70%, respectively. Maji et al.<sup>4</sup> presented a method of automatic detection of AF by using higher order statistical moments of ECG signal in Empirical Mode Decomposition (EMD) domain. The performance of this method was tested on the MIT-BIH arrhythmia database with Sensitivity of 96% and Specificity of 93%. Teijeiro et al.<sup>15</sup> employed a Convolutional Neural Network (CNN) with one fully connected layer to learn the time-domain features. Then the proposed CNN-based feature learning mechanism was integrated with the other standard classifiers to improve the accuracy. The results demonstrated that the integration of CNN structure as a feature extractor with the other conventional classifier can improve the resulting classification performance.

## 2.2. Block

In the process of cardiac electrical activation, abnormal electrical conduction occurs, so that the heart cannot normally contract and pump blood, which is called conduction block. It mainly divided into intro-atrial block, atrioventricular block (AVB) and bundle branch block (BBB). At present, automatic detection methods focused on BBB.

In 2007, Ilic<sup>16</sup> employed the Continuous Wavelet Transform (CWT) to detect the normal and left BBB (LBBB). In 2011, Ceylan et al.<sup>17</sup> proposed wavelet neural network (WNN) with Morlet and Mexican hat wavelet functions as activation function in hidden layer to classify LBBB, right BBB (RBBB) and normal types, achieving accuracies of 97.9% and 99.2% respectively. In 2014, Kora and Kalva<sup>18</sup> proposed a hybrid technique: Bacterial Forging-Particle Swarm Optimization (BFPSO) for the feature selection and used the Levenberg-Marquardt neural network to classify LBBB, RBBB and normal types. This work achieved the results with 98.2% (LBBB), 98.15 (RBBB) and 98.1% (Normal). Huang et al.<sup>19</sup> proposed a heartbeat classification method through a combination of three different types of classifiers: a minimum distance classifier, a weighted linear discriminant classifier and support vector machine (SVM). The results showed a sensitivity of 91.4% and a positive predictive value of 37.3% for LBBB and

a sensitivity of 92.8% and a positive predictive value of 88.8% for RBBB. In 2005, Yuksel and Bekir<sup>20</sup> proposed a new neural network structure based on fuzzy clustering (FCNN) for BBB classification the results showed that FCNN method gave better recognition rates for mixed classification.

## 2.3. Ventricular Premature Contraction

Before the sinus node impulse has reached the ventricle, an electrical impulse is caused by an electrical impulse in advance of any site in the ventricle or the ectopic rhythm of the inter-ventricular septum, causing ventricular depolarization, known as premature ventricular contraction (PVC). It has a high incidence in adults and can cause other diseases. Frequent PVC can be complicated by syncope, angina, and heart failure, which poses a great health threat to the elderly. There are many researchers dedicated to the automatic detection of PVC. SVM, neural network, wavelet analysis, Gaussian mixture model, decision trees, multi-domain feature extraction, etc., were commonly used.

Benali et al.<sup>21</sup> used the supervised classifiers and efficient features to detect PVC automatically, and the accuracy achieved about 97.14%. Saibal et al.<sup>22</sup> proposed a model with a feature extractor based on cross-correlation approach and an Artificial Neural Network (ANN) classifier to detect PVC with an accuracy of 95.24%. Vessela et al.<sup>23</sup> developed a three processing stages-based method, including preprocessing filtration, heartbeat detection and heartbeat classification by estimation of the inter beat differences of the RR intervals and the QRS morphological descriptors, and obtained sensitivity of 92.2% and specificity of 96%. Using fuzzy C-means clustering algorithm, Sutar et al.<sup>24</sup> develop an classifier to use the features of informational entropy and mean Teager energy, and achieved sensitivity of 97.78% and PPV of 97.28%.

## 2.4. Premature Atrial Contraction

Premature atrial contraction (PAC) occurs at a high rate, commonly followed by PVC. Its main manifestations are palpitation, chest distress, uncomfortable dizziness in the precordial area and intermittent pulse.

Vessela et al.<sup>23</sup> proposed a method for PAC detection based on QRS morphological descriptors and the RR-interval information, achieving a sensitivity of 88%. Khazaei et al.<sup>25</sup> proposed a power spectral-based hybrid genetic algorithm-SVM (SVMGA) method and achieved sensitivity of 98.93% sensitivity, specificity of 99.42% and PPV of 97.56%. In Ref. [26], two-layered Hidden Markov Models (HMMs) were employed to classify normal, PVC and PAC, obtaining a sensitivity of 99.21% and PPV of 95.57% for PAC beats.

## 2.5. Heartbeat Classification

Sections 2.1–2.4 mainly described the detection methods for one arrhythmia type. In fact, more papers have been focused on the multiple classification.

Yeh et al.<sup>27</sup> proposed a cluster analysis (CA) method and obtained sensitivity as 95.59%, 91.32%, 90.50%, 94.51% and 93.77% for normal, LBBB, RBBB, PVC, and PAC respectively, with a total classification accuracy of 94.30%. Teijeiro et al.<sup>15</sup> proposed a novel knowledge-based approach to classify five beat classes in the MIT-BIH arrhythmia and obtained a sensitivity of 94.63% and PPV of 96.79% for multiple classification. In 2017, Rajpurkar et al.<sup>28</sup> built a dataset with of 64,121 ECG

records from 29,163 patients and trained 34-layer CNN to detect 14 rhythm classes. The database was much larger than other datasets, such as MIT/BIH. The model outperformed the average cardiologist performance on most rhythms, noticeably outperforming the cardiologists in the AV block set of arrhythmias which includes Mobitz I (Wenckebach), Mobitz II (AVB Type2) and complete heart block (CHB).

## 2.6. Ischemic Disease

Ischemic disease is an important cause of death worldwide, responsible for more than 8 million deaths globally every year.<sup>2</sup> If persistent, ischemia will lead to cell death and permanent damage to the heart muscle causing a myocardial infarction (MI). The main characteristics in ECGs are ST segment elevation or depression, *T* wave abnormality, the development of pathological *Q* waves.<sup>6</sup> ST-segment deviation is the most widely used feature for ischemic disease and MI detection.<sup>29</sup>

In 1984, Gallino et al.<sup>30</sup> analyzed the ST-segment deviation of 2-lead ECGs on 24 hour Holter monitor tapes. Concurrently, Mitchell<sup>31</sup> and Sun<sup>32</sup> used Karhunen-Loeve transform (KLT) to analyze the ST-segment changes. Then, principal component analysis (PCA) was applied. Diamantaras<sup>33</sup> proposed a nonlinear PCA method to the extracted ST segment. Recently, various advanced methods are used to identify ischemic disease and MI detection. In 2008, Gharaviri<sup>34</sup> proposed an adaptive neuro-fuzzy interface system (ANFIS) classifier to classify the ST segment elevation and depression, which obtained a sensitivity of 88.62% and a specificity of 99.65%. In 2012, Arif et al.<sup>35</sup> used *K*-nearest neighbor (KNN) to develop an automatic method for MI localization, using the features of *T* wave amplitude, *Q* wave amplitude and ST deviation measure and achieving overall classification accuracy of 98.3%. Sun et al.<sup>36</sup> proposed the latent topic multiple instance learning method to build a totally automatic ECG classification procedure, which achieved a good performance on recognition of ECGs related to MI. In 2015, Murthy and Meenakshi<sup>37</sup> provided the comparison of performance of ANN, SVM and KNN models for cardiac ischemia classification. In 2016, Padhy and Dandapat<sup>38</sup> employed a third-order tensor

structure to represent the multi-lead ECG signals and used the modern singular values and normalized multiscale wavelet energy as the input features. This work achieved the detection accuracy of 95.30%, with sensitivity and specificity of 94.6% and 96.0%, respectively. There are many published papers on automatic diagnosis of myocardial ischemic and myocardial infarction. Literature<sup>6</sup> gives a detailed overview and the readers can refer to this article.

## 3. DESCRIPTION OF THE 2018 CHINA PHYSIOLOGICAL SIGNAL CHALLENGE

### 3.1. Main Aim

The China Physiological Signal Challenge (CPSC) 2018 is the 1st China Physiological Signal Challenge, aiming to provide a platform for the open-source data and algorithms for the physiological signal analysis, and thus to promote the open-source research pattern for the cardiovascular disease detection and prediction in China. The practical objective is to identify the rhythm/morphology abnormalities from 12-lead ECGs, lasting several seconds to tens of seconds.

A number of studies have investigated the performances of different detection/classification methods for the abnormal ECG types. However, many studies are generally limited in applicability because (1) the classification of normal and only one single abnormality was performed; (2) the data were not sufficient without the use of a separate out of sample test dataset and only a small number of subjects were used, almost certainly resulting in over-fitting of the model and inflated statistics; (3) failure to post the data (and any code to process the data) publicly so others may compare their results directly. Therefore, this challenge contributes a more comprehensive database to address these issues, including one normal type and eight abnormal types, which are detailed as:

- (1) Atrial fibrillation (AF)
- (2) First-degree atrioventricular block (I-AVB)
- (3) Left bundle brunch block (LBBB)
- (4) Right bundle brunch block (RBBB)

Table I. Data profile for the training and testing sets.

Challenge set	Type	# recording of each hospital											# total recording	Time length (s)				
		A	B	C	D	E	F	G	H	I	J	K		Mean	SD	Min	Median	Max
Training	Normal	73	90	74	73	70	90	77	76	78	77	140	918	15.43	7.61	10.00	13.00	60.00
	AF	45	184	28	26	138	155	280	104	27	47	64	1098	15.01	8.39	9.00	11.00	60.00
	I-AVB	39	71	0	8	150	129	212	19	15	27	34	704	14.32	7.21	10.00	11.27	60.00
	LBBB	7	12	7	2	41	30	51	13	9	18	17	207	14.92	8.09	9.00	12.00	60.00
	RBBB	79	272	29	30	423	205	401	140	26	30	60	1695	14.42	7.60	10.00	11.19	60.00
	PAC	57	147	4	8	105	24	111	50	2	57	9	574	19.46	12.36	9.00	14.00	60.00
	PVC	76	111	3	21	150	80	132	13	18	32	17	653	20.21	12.85	6.00	15.00	60.00
	STD	117		22	32	525	36	52	1	26		15	826	15.13	6.82	8.00	12.78	60.00
	STE	3	90	6	7	38	12	16	0	22	0	8	202	17.15	10.72	10.00	11.89	60.00
Total/Average		490	980	169	202	1645	762	1338	416	223	288	364	6877	15.79	9.04	6.00	12.00	60.00
Test	Normal	31	38	32	32	30	38	33	33	34	33	60	394	15.91	–	–	–	–
	AF	19	79	12	11	59	66	120	44	11	20	28	469	17.31	–	–	–	–
	I-AVB	17	30	0	3	64	56	91	8	6	11	15	301	15.34	–	–	–	–
	LBBB	3	5	3	1	18	13	22	5	4	8	8	90	16.51	–	–	–	–
	RBBB	34	116	13	13	181	88	173	61	12	13	25	729	16.53	–	–	–	–
	PAC	23	65	1	3	45	12	48	22	3	24	4	250	23.06	–	–	–	–
	PVC	30	49	2	10	65	33	56	6	8	15	7	281	21.28	–	–	–	–
	STD	50	0	9	14	225	16	23	0	11	0	6	354	14.93	–	–	–	–
	STE	2	38	2	3	16	5	7	0	10	0	3	86	22.46	–	–	–	–
Total/Average		209	420	74	90	703	327	573	179	99	124	156	2954	18.15	–	–	–	–

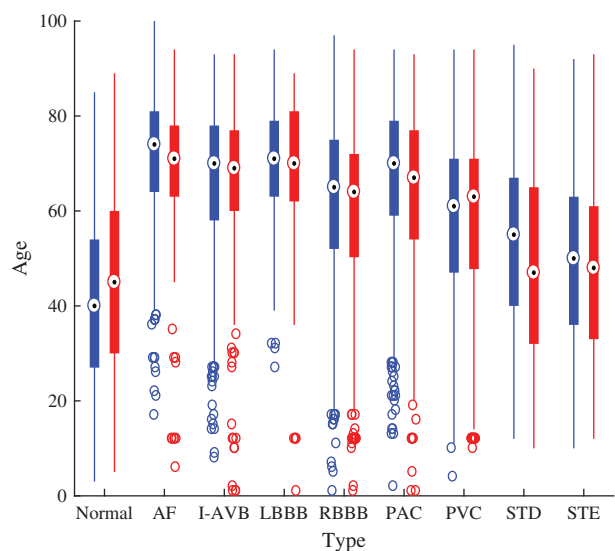


Fig. 1. The age distribution of training and test set.

- (5) Premature atrial contraction (PAC)
- (6) Premature ventricular contraction (PVC)
- (7) ST-segment depression (STD)
- (8) ST-segment elevated (STE).

**3.2. Challenge Data**

The challenge ECG recordings were collected from 11 hospitals and were generously donated for this Challenge. In order to protect the privacy of patients, the hospital names are not disclosed here. 11 hospitals are represented by A–K 11 letters. All data from 11 hospitals were combined to constitute the challenge database. The database was divided into both training and test sets with a random 70–30 training-test split. The training set contains 6,877 (female: 3178; male: 3699) 12 leads ECG recordings lasting from 6 s to just 60 s and the test set contains 2,954 (female: 1416; male: 1538) ECG recordings with the similar lengths. Table I shows the details of these two data sets, including the data source, recording information. Figure 1 illustrates the patients’ age distribution and so on. The test set is unavailable to the public and will remain private

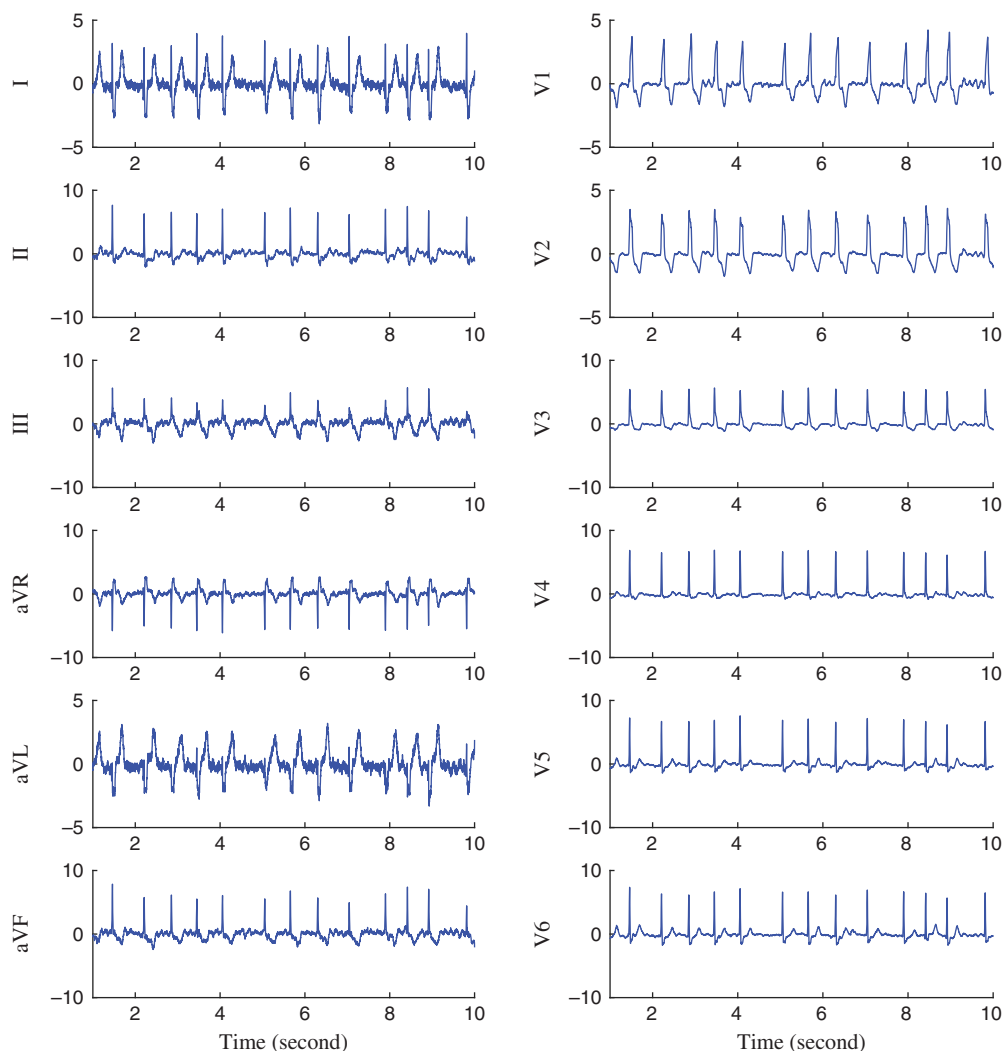


Fig. 2. Example of a 12-lead ECG waveforms (AF and RBBB).

for the purpose of scoring. ECG recordings were sampled as 500 Hz. All data are provided in MATLAB format (each recording is a .mat file containing the ECG data, as well as the patient sex and age information).

In both training and test sets, ECG signal recordings were divided into 9 types. For the most data, each recording has only one label. While, some recordings have two or three labels because the patient who provided the signals suffers from multiple diseases simultaneously. There are 477 and 203 subjects of this multi-label type in the training and test sets respectively. Figure 2 illustrates an example of the 12-lead ECG waveforms with AF and RBBB. For this multi-label situation, as long as the result provided by the contestant is consistent with one of the labels, we consider this to be a correct result.

### 3.3. Scoring Mechanism

The overall score is computed based on the number of recordings classified as nine types. The scoring for CPSC2018 uses a  $F_1$  measure, which is an average of the nine  $F_1$  values from each classification type. The counting rules for the numbers of the variables are defined in Table II.

For each of the nine types,  $F_1$  is defined as:

$$\text{Normal: } F_{11} = \frac{2 \times N_{11}}{N_{1X} + N_{X1}} \quad (1)$$

$$\text{AF: } F_{12} = \frac{2 \times N_{22}}{N_{2X} + N_{X2}} \quad (2)$$

$$\text{I-AVF: } F_{13} = \frac{2 \times N_{33}}{N_{3X} + N_{X3}} \quad (3)$$

$$\text{LBBB: } F_{14} = \frac{2 \times N_{44}}{N_{4X} + N_{X4}} \quad (4)$$

$$\text{RBBB: } F_{15} = \frac{2 \times N_{55}}{N_{5X} + N_{X5}} \quad (5)$$

$$\text{PAC: } F_{16} = \frac{2 \times N_{66}}{N_{6X} + N_{X6}} \quad (6)$$

$$\text{PVC: } F_{17} = \frac{2 \times N_{77}}{N_{7X} + N_{X7}} \quad (7)$$

$$\text{STD: } F_{18} = \frac{2 \times N_{88}}{N_{8X} + N_{X8}} \quad (8)$$

$$\text{STE: } F_{19} = \frac{2 \times N_{99}}{N_{9X} + N_{X9}} \quad (9)$$

Table II. Counting rules for the numbers of the variables.

	Predicted									Total
	Normal	AF	I-AVB	LBBB	RBBB	PAC	PVC	STD	STE	
Reference										
Normal	$N_{11}$	$N_{12}$	$N_{13}$	$N_{14}$	$N_{15}$	$N_{16}$	$N_{17}$	$N_{18}$	$N_{19}$	$N_{1X}$
AF	$N_{21}$	$N_{22}$	$N_{23}$	$N_{24}$	$N_{25}$	$N_{26}$	$N_{27}$	$N_{28}$	$N_{29}$	$N_{2X}$
I-AVB	$N_{31}$	$N_{32}$	$N_{33}$	$N_{34}$	$N_{35}$	$N_{36}$	$N_{37}$	$N_{38}$	$N_{39}$	$N_{3X}$
LBBB	$N_{41}$	$N_{42}$	$N_{43}$	$N_{44}$	$N_{45}$	$N_{46}$	$N_{47}$	$N_{48}$	$N_{49}$	$N_{4X}$
RBBB	$N_{51}$	$N_{52}$	$N_{53}$	$N_{54}$	$N_{55}$	$N_{56}$	$N_{57}$	$N_{58}$	$N_{59}$	$N_{5X}$
PAC	$N_{61}$	$N_{62}$	$N_{63}$	$N_{64}$	$N_{65}$	$N_{66}$	$N_{67}$	$N_{68}$	$N_{69}$	$N_{6X}$
PVC	$N_{71}$	$N_{72}$	$N_{73}$	$N_{74}$	$N_{75}$	$N_{76}$	$N_{77}$	$N_{78}$	$N_{79}$	$N_{7X}$
STD	$N_{81}$	$N_{82}$	$N_{83}$	$N_{84}$	$N_{85}$	$N_{86}$	$N_{87}$	$N_{88}$	$N_{89}$	$N_{8X}$
STE	$N_{91}$	$N_{92}$	$N_{93}$	$N_{94}$	$N_{95}$	$N_{96}$	$N_{97}$	$N_{98}$	$N_{99}$	$N_{9X}$
Total	$N_{X1}$	$N_{X2}$	$N_{X3}$	$N_{X4}$	$N_{X5}$	$N_{X6}$	$N_{X7}$	$N_{X8}$	$N_{X9}$	

The final challenge score is defined as follows:

$$F_1 = \frac{F_{11} + F_{12} + F_{13} + F_{14} + F_{15} + F_{16} + F_{17} + F_{18} + F_{19}}{9} \quad (10)$$

In addition, we also calculate the  $F_1$  measures for each of the four sub-abnormal types, i.e., the AF, block, premature contraction and ST-segment change, as follows:

$$\text{AF: } F_{AF} = \frac{2 \times N_{22}}{N_{2X} + N_{X2}} \quad (11)$$

$$\text{Block: } F_{Block} = \frac{2 \times (N_{33} + N_{44} + N_{55})}{N_{3X} + N_{X3} + N_{4X} + N_{X4} + N_{5X} + N_{X5}} \quad (12)$$

$$\text{Premature contraction: } F_{PC} = \frac{2 \times (N_{66} + N_{77})}{N_{6X} + N_{X6} + N_{7X} + N_{X7}} \quad (13)$$

$$\text{ST-segment change: } F_{ST} = \frac{2 \times (N_{88} + N_{99})}{N_{8X} + N_{X8} + N_{9X} + N_{X9}} \quad (14)$$

## 4. DISCUSSION

This paper presents an introduction for the database used in CPSC 2018, including the data source, recording information, patients' clinical baseline parameters (age and gender). Meanwhile, it also presents the common detection/classification methods for the abovementioned abnormal ECG types.

At present, there are various published ECG databases for evaluating the ECG abnormalities classification. MIT-BIH arrhythmic database<sup>39</sup> is the most frequently used one, consisting of 48 records at 360 Hz for approximately 30 min from 47 different patients.<sup>9</sup> This database contains 15 recommended classes of arrhythmia, which are further classified into five super-classes: normal ( $N$ ), supraventricular ectopic beat (SVEB), ventricular ectopic beat (VEB), fusion beat ( $F$ ) and unknown beat ( $Q$ ).<sup>40</sup> The European ST-T database was the standard for evaluating ST-T changes in ambulatory ECG,<sup>11</sup> which consists of 90 records acquired from 79 subjects suffering myocardial ischemia, sampled at 250 Hz. The MIT-BIH Atrial Fibrillation Database<sup>12</sup> includes 25 long-term ECG AF recordings (mostly paroxysmal), providing a database for evaluating AF detection method. These databases contained one disease or one type abnormal. Thus the evaluations focus on only one or a few combination of disease abnormalities, such as AF,<sup>4,5</sup> ST change,<sup>6,7</sup> etc.

To address this issue, CPSC 2018 contains more abnormal types (normal type and eight abnormal types). The public release of such a database has many potential benefits to a wide range of users. The data of 9,831 records from 9,458 patients were sufficient to use a separate out of sample test dataset. The availability of these data can encourage various researchers to develop innovative algorithms.

### Conflicts of Interest Statement

There is no conflict of interest to this work.

**Acknowledgments:** This work was supported by the National Natural Science Foundation of China (61571113), Key Research and Development Programs of Jiangsu Province (BE2017735), Shandong Provincial Natural Science Foundation in China (ZR2014EEM003), Fundamental Research Funds for the Central Universities in Southeast University (2242018k1G010).

## References and Notes

- C. Y. Liu, D. Springer, Q. Li, B. Moody, R. A. Juan, F. J. Chorro, F. Castells, J. M. Roig, I. Silva, A. E. W. Johnson, Z. Syed, S. E. Schmidt, C. D. Papadaniil, L. Hadjileontiadis, H. Naseri, A. Moukadem, A. Dieterlen, C. Brandt, H. Tang, M. Samieinasab, M. R. Samieinasab, R. Sameni, R. G. Mark, and G. D. Clifford, An open access database for the evaluation of heart sound algorithms. *Physiological Measurement* 37, 2181 (2016).
- W. H. Organization, Global health estimates: Deaths by cause, age, sex and country, 2000–2015, WHO, Geneva, Switzerland (2017).
- Y. C. Yeh and W. J. Wang, QRS complexes detection for ECG signal: The difference operation method. *Computer Methods and Programs in Biomedicine* 91, 245 (2008).
- U. Maji, M. Mitra, and S. Pal, Automatic detection of atrial fibrillation using empirical mode decomposition and statistical approach. *Procedia Technology* 10, 45 (2013).
- B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, Feature leaning with deep convolutional neural networks for screening patients with paroxysmal atrial fibrillation, *International Joint Conference on Neural Networks*, Vancouver, BC, Canada (2016), pp. 5057–5064.
- S. Ansari, N. Farzaneh, M. Duda, K. Horan, and H. B. Andersson, A review of automated methods for detection of myocardial ischemia and infarction using electrocardiogram and electronic health record. *IEEE Reviews in Biomedical Engineering* 10, 264 (2017).
- ANSI/AAMI, Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms, *American National Standards Institute, Inc. (ANSI), Association for the Advancement of Medical Instrumentation (AAMI) 1998-(R)2008* (2008).
- S. Lévy, A. J. Camm, S. Saksena, E. Aliot, G. Breithardt, H. Crijns, W. Davies, N. Kay, E. Prystowsky, R. Sutton, A. Waldo, D. G. Wyse, Working Group on Arrhythmias, Working Group on Cardiac Pacing of the European Society of Cardiology, and North American Society of Pacing and Electrophysiology, International consensus on nomenclature and classification of atrial fibrillation; a collaborative project of the working group on arrhythmias and the working group on cardiac pacing of the european society of cardiology and the North American society of pacing and electrophysiology. *Journal of Cardiovascular Electrophysiology* 14, 443 (2003).
- G. B. Moody and R. G. Mark, The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine* 20, 45 (2002).
- R. E. Hermes, D. B. Geselowitz, and G. C. Oliver, Development, distribution, and use of the American heart association database for ventricular arrhythmia detector evaluation. *Computers in Cardiology* 20, 263 (1980).
- A. Taddei, G. Distante, M. Erndin, P. Pisani, G. B. Moody, C. Zeelenberg, and C. Marchesi, The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *European Heart Journal* 13, 1164 (1992).
- G. Moody and R. Mark, A new method for detecting atrial fibrillation using R–R intervals. *Computers in Cardiology* 10, 227 (1983).
- G. K. Andrikopoulos, P. E. Dilaveris, D. J. Richter, E. J. Gialafos, A. G. Synetos, and J. E. Gialafos, Increased variance of P wave duration on the electrocardiogram distinguishes patients with idiopathic paroxysmal atrial fibrillation. *Pacing and Clinical Electrophysiology* 23, 1127 (2000).
- R. Lepage, J. M. Boucher, J. J. Blanc, and J. C. Cornilly, ECG segmentation and P-wave feature extraction: Application to patients prone to atrial fibrillation, *Engineering in Medicine and Biology Society, 2001, Proceedings of the International Conference of the IEEE, Istanbul, Turkey* (2001), Vol. 1, pp. 298–301.
- T. Teijeiro, P. Felix, J. Presedo, and D. Castro, Heartbeat classification using abstract features from the abductive interpretation of the ECG. *IEEE Journal of Biomedical and Health Informatics* 22, 409 (2016).
- S. S. Ilic, Detection of the left bundle branch block in continuous wavelet transform of ECG signal. *Elektronika Ir Elektrotehnika* 74, 33 (2007).
- R. Ceylan and Y. Ozbay, Wavelet neural network for classification of bundle branch blocks. *Lecture Notes in Engineering and Computer Science* 2191, 1003 (2011).
- P. Kora and S. R. Kalva, Hybrid bacterial foraging and particle swarm optimization for detecting bundle branch block, *Springerplus* 4, 48 (2015).
- H. Huang, J. Liu, Q. Zhu, R. Wang, and G. Hu, Detection of inter-patient left and right bundle branch block heartbeats in ECG using ensemble classifiers. *Biomedical Engineering Online* 13, 72 (2014).
- Y. Özbay, R. Ceylan, and B. Karlik, A fuzzy clustering neural network architecture for classification of ECG arrhythmias, *Computers in Biology and Medicine* 36, 376 (2006).
- R. Benali, F. B. Reguig, and Z. H. Slimane, Automatic classification of heartbeats using wavelet neural network. *J. Med Syst* 36, 883 (2012).
- A. C. Saibal Dutta and Sugata Munshi, Identification of ECG beats from cross-spectrum information aided learning vector quantization. *Measurement* 2, 2020 (2011).
- I. I. J. Vessela, T. Krasteva, and I. I. Christov, Automatic detection of premature atrial contractions in the electrocardiogram. *Electrotechniques and Electronics* 2, 49 (2006).
- R. G. Sutar and A. G. Kothari, Intelligent electrocardiogram pattern classification and recognition using low-cost cardio-care system. *IET Science, Measurement and Technology* 9, 134 (2015).
- A. Khazaei and A. Ebrahimzadeh, Classification of electrocardiogram signals with support vector machines and genetic algorithms using power spectral features. *Biomedical Signal Processing and Control* 5, 252 (2010).
- W. Liang, Y. Zhang, J. Tan, and Y. Li, A novel approach to ECG classification based upon two-layered HMMs in body sensor networks. *Sensors (Basel)* 14, 5994 (2014).
- Y. C. Yeh, C. W. Chiou, and H. J. Lin, Analyzing ECG for cardiac arrhythmia using cluster analysis. *Expert Systems with Applications* 39, 1000 (2012).
- P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, Cardiologist-level arrhythmia detection with convolutional neural networks. *Computer Vision and Pattern Recognition* arXiv:1707.01836v1, 1 (2017).
- H. L. Lu, K. Ong, and P. Chia, An automated ECG classification system based on a neuro-fuzzy system, *Computers in Cardiology, Cambridge, MA, USA* (2000), pp. 387–390.
- A. Gallino, S. Chierchia, G. Smith, M. Croom, M. Morgan, C. Marchesi, and A. Maseri, Computer system for analysis of ST segment changes on 24 hour Holter monitor tapes: Comparison with other available systems. *Journal of the American College of Cardiology* 4, 245 (1984).
- M. W. Krucoff, M. A. Croll, L. P. Pendley, and D. L. Burdette, Continuous computer-assisted electrocardiographic monitoring in patients with acute myocardial infarction: Early experience, *Computers in Cardiology, Jerusalem, Israel* (1989), pp. 197–200.
- G. Sun, C. W. Thomas, J. Liebman, and Y. Rudy, Classification of normal and ischemia from BSPM by neural network approach, *Engineering in Medicine and Biology Society, 1988, Proceedings of the International Conference of the IEEE, New Orleans, LA, USA* (1988), Vol. 3, pp. 1504–1505.
- K. I. Diamantaras and S. Y. Kung, *Principal Component Neural Networks: Theory And Applications*, John Wiley & Sons, Inc., New York, NY, USA (1996).
- A. Gharaviri, M. Teshnehlab, and H. A. Moghaddam, Ischemia detection via ECG using ANFIS, *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada* (2008), pp. 1163–1166.
- M. Arif, I. A. Malagore, and F. A. Afsar, Detection and localization of myocardial infarction using K-nearest neighbor classifier. *Journal of Medical Systems* 36, 279 (2012).
- L. Sun, Y. Lu, K. Yang, and S. Li, ECG analysis using multiple instance learning for myocardial infarction detection. *IEEE Trans. Biomed. Eng.* 59, 3348 (2012).
- H. S. N. Murthy and D. M. Meenakshi, ANN, SVM and KNN classifiers for prognosis of cardiac ischemia—A comparison. *Bonfring International Journal of Research in Communication Engineering* 5, 7 (2015).
- S. Padhy and S. Dandapat, Third-order tensor based analysis of multilead ECG for classification of myocardial infarction. *Biomedical Signal Processing and Control* 31, 71 (2017).
- A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals. *Circulation* 101, e215 (2000).
- E. J. D. Luz, W. R. Schwartz, G. Camara-Chavez, and D. Menotti, ECG-based heartbeat classification for arrhythmia detection: A survey. *Computer Methods and Programs in Biomedicine* 127, 144 (2016).

Received: 3 April 2018. Revised/Accepted: 16 April 2018.