

# Article

# A Spatial Pyramid Pooling-Based Deep Convolutional Neural Network for the Classification of **Electrocardiogram Beats**

# Jia Li <sup>1,2</sup>, Yujuan Si <sup>1,2,\*</sup>, Liuqi Lang <sup>2</sup>, Lixun Liu <sup>2</sup> and Tao Xu <sup>3</sup>

- 1 College of Instrument Science and Electrical Engineering, Jilin University, Changchun 130061, China; lijia14@mails.jlu.edu.cn
- 2 Department of Electronic Information Engineering, Zhuhai College of Jilin University, Zhuhai 519041, China; liuqilang2005@163.com (L.La.); emtrix@163.com (L.Li.)
- 3 Department of Biomechanical Engineering, City University of Hong Kong, Hong Kong SAR 999077, China; taoxu8-c@my.cityu.edu.hk
- Corresponding author: siyj@jlu.edu.cn

Received: 16 July 2018; Accepted: 31 August 2018; Published: 8 September 2018



**Abstract:** An accurate electrocardiogram (ECG) beat classification can benefit the diagnosis of the cardiovascular disease. Deep convolutional neural networks (CNN) can automatically extract valid features from data, which is an effective way for the classification of the ECG beats. However, the fully-connected layer in CNNs requires a fixed input dimension, which limits the CNNs to receive fixed-scale inputs. Signals of different scales are generally processed into the same size by segmentation and downsampling. If information loss occurs during a uniformly-sized process, the classification accuracy will ultimately be affected. To solve this problem, this paper constructs a new CNN framework spatial pyramid pooling (SPP) method, which solves the deficiency caused by the size of input data. The Massachusetts Institute of Technology-Biotechnology (MIT-BIH) arrhythmia database is employed as the training and testing data for the classification of heartbeat signals into six categories. Compared with the traditional method, which may lose a large amount of important information and easy to be over-fitted, the robustness of the proposed method can be guaranteed by extracting data features from different sizes. Experimental results show that the proposed architecture network can extract more high-quality features and exhibits higher classification accuracy (94%) than the traditional deep CNNs (90.4%).

Keywords: ECG beats; classification; feature extraction; convolutional neural networks; spatial pyramid pooling

# 1. Introduction

# 1.1. Present Situation for Electrocardiogram Pattern Recognition

An electrocardiogram (ECG) is a pattern in which various forms of potential changes are extracted from the body surface via an electrocardiograph. Moreover, the ECG also has an important reference value for basic cardiac functions and related pathological research, and an experienced cardiologist can easily tell the arrhythmia according to the morphological pattern of the ECG signals. However, the computer-aided approaches to the morphological pattern recognition of the ECG signal are difficult to realize. It is due to the time-varying dynamics and various profiles of the ECG signals that make the precision of the classification vary from patient to patient [1]. Nevertheless, computer-aided approaches can improve the efficiency of diagnosis, and thus freeing physicians from cumbersome pattern recognition tasks. Additionally, the development of pattern recognition of an ECG signal and



real-time diagnosis of cardiovascular [1] requires further exploration for the E-home health monitoring device [2] in the future.

#### 1.2. Computer-Aided Method for Pattern Recognition and Preprocessing of Heartbeat Signals

Artificial intelligence and machine learning have been widely used in heartbeat recognition and classification. Current methods include the support vector machine (SVM) [3], least squares support vector machine (LS-SVM) [4], particle swarm optimization support vector machine (PSO-SVM) [5], particle swarm optimization radius basis function (PSO-RBF) [6], and neural networks (NN) [7]. In addition, the pre-processing method like the Fourier transform (FT) [8] and the principle component analysis (PCA) [9] have also been explored for the accurate identification of the ECG signals. In Ref. [10], a Stationary Wavelet Transform (SWT) algorithm was deemed suitable for de-noising of the ECG signals judging from a comparison of three de-noising algorithms based on wavelet packet transform (WPT), lifting wavelet (LW), and an SWT.

#### 1.3. Feature Extraction Method for an ECG

However, the ECG signal identification technology is limited by noise reduction and feature extraction, which complicates the improvement of effective ECG signal recognition. The ECG feature extraction is a key technique for heartbeat recognition. Feature extraction selects a representative feature subset from the raw ECG signal. These feature subsets have better generalization capabilities and can improve the accuracy of the ECG heartbeat classification. Underlying feature extraction mainly revolves around the extraction of the time-domain features, frequency-domain features, or morphological features of the signal, such as by FT [11], discrete cosine transform (DCT) [12], and wavelet transform (WT) [13]. Some high-level feature extraction methods are also available, including dictionary learning [14] and CNNs [15,16]. With the increase of the number of patients, the accuracy of the classification will be decreased due to the large pattern variations of the ECG signals among different patients, and the preprocessing methods like PCA [9] and Fourier transform [8] may increase the complexity and the time of the computing as well. To enhance the heartbeat classification performance, selecting a suitable feature is of paramount importance.

#### 1.4. CNN and Spatial Pyramid Pooling (SPP)-Net for Pattern Recognition

In recent years, CNN algorithms have proven particularly effective in language and image recognition [17]. The network structure of the CNN algorithm includes many hidden layers, and it also has an unmatched feature-learning level compared with the traditional machine learning methods. A traditional classification method like SVM needs to conduct the feature extraction alone before feeding the data into the classifier. For example, Khorrami and Moavenian employed three feature extraction methods (i.e., DCT, continuous WT, and discrete WT) to realize the feature extraction before the classification [12]. It is noteworthy that the selection of the mother wavelet is very important to the feature selection. In addition, it is better to pre-compute the basic function of the DCT offline to improve the computational efficiency. As mentioned in Ref. [12], the selection of the best feature extraction method depends on the substantial value considered for the training time, and the training and testing performance. Therefore, the feature extraction is generated automatically, and such a feature extraction method has a better effect on classification for complex tasks. The CNN itself is a feature extractor, and its convolutional layer works as a series of filters that are deployed for feature extraction. Moreover, the other layers, such as the pooling layer and the fully-connected layer, are used to reduce the number of the parameters to be learned and retains the most useful information for a classifier.

However, the existing CNNs mandate that the input data should have the same size and such a fixed-size constraint comes from the fully-connected layer and requires a fixed-length vector for the input. This artificial operation may result in loss of image information, which also affects the classification accuracy. A new structure of the CNNs called SPP-net [18] has solved these problems for pattern recognition by adding an SPP layer on top of the last convolutional layer. The SPP layer pools the features and generates fixed-length outputs, which are subsequently fed into the fully-connected layers (or other classifiers). The SPP-net allows CNN to accept inputs of any scale, which increases the scale invariance of the model, suppresses overfitting, and enables extraction of local features of the data at multiple scales [18–20]. The SPP-net is implemented by switching from one network size (224  $\times$  224) to another (180  $\times$  180) and training each full epoch on one network. After that, the network size should be switched to the other (while retaining all weights) for the next full epoch. Accordingly, most fixed-size pictures are trained on a single network, whereas different-sized pictures are trained on a separate network. The weights of different networks cannot be shared under such network switching.

#### 1.5. Goal and Arrangement of This Paper

In this study, the size of heartbeats divided from the ECG is not equal and thus unsuitable for SPP-net training. To avoid the network switching, a new SPP-net based CNNs model has been constructed in this paper for the heartbeats classification. This model retains the advantages of the SPP and allows different-sized heartbeats to be sent to the same network for training, thus reducing the complexity of the network. This approach also avoids the complexity of data reconstruction during feature extraction and classification. The SPP structure is employed for the classifications, and it also enables the adaption of the CNN structure without cropping or warping the original heartbeat signal. In addition, the input of the SPP structure is simplified into one-dimension (1-D), which is suitable for the heartbeat classification with less of a computational burden.

Additionally, due to the non-stationary nature of the ECG signal, frequency domain filters may distort a transient interval of the signal and important biomedical information may get lost [21–23]. However, a wavelet is simply a small wave, which enables the analyzing of the transient, non-stationary or time-varying signals easily [24]. Moreover, due to the sparsity, locality, and multi-resolution nature [25] of the WT, WT is therefore employed as the pre-processing method for the ECG signal. The rest of the paper is arranged as follows: Section 2 introduces the methods and procedure adopted in this study, including the SPP, ECG-SPP-net, pre-processing of input data to the ECG-SPP-net, feature extraction, and classification. To validate the performance of the proposed method, the accuracies of different network structures are analyzed in Section 3, and a conclusion is finally provided in Section 4.

### 2. Method

#### 2.1. Spatial Pyramid Pooling Method

As mentioned above, SPP guarantees fixed eigenvector output by using multiple different-sized pool operations to achieve input at any scale. Specific pooling operations include max pooling, average pooling, and stochastic pooling [26]. Ref. [27] found that stochastic pooling and max-pooling were more robust than average pooling. In this paper, the SPP method is combined with a deep CNN [18]. An SPP is placed as a layer in the network between the convolutional layer and the fully-connected layer (Figure 1). The input of the SPP layer is the total number of the feature maps of the last convolutional operation, which is denoted as  $M_{con_2}$ , and each feature vector is denoted as  $N_{con_2}$ . The pyramid level can be expressed as  $1 \times n$  bins. It is assumed that one feature vector has a size of  $1 \times a$  (e.g.,  $1 \times 13$ ), and a pooling level with  $1 \times n$  bins can be implemented with a sliding window size  $\lceil a/n \rceil$  and stride  $\lfloor a/n \rfloor$ , where  $\lceil \cdot \rceil$  and  $\lfloor \cdot \rfloor$  denote the ceiling and flooring operations, respectively [18]. A three-level pooling  $(1 \times 1, 1 \times 2, \text{ and } 1 \times 4)$  for one feature vector (a size of  $1 \times 13$ ) is shown in Figure 2. Then, a fixed feature-vector output can be achieved as the input of the fully-connected layer regardless of the size of the feature maps.



**Figure 2.** A three-level pooling  $(1 \times 1, 1 \times 2, \text{ and } 1 \times 4)$  for one feature vector (a size of  $1 \times 13$ ).

# 2.2. Electrocardiogram-Spatial Pyramid Pooling-Net Method

In this study, an ECG-SPP-net for the classification of heartbeats is developed, and such a network consists of alternate convolutional layers and subsampling layers. The detailed structure of the ECG-SPP-net is shown in Table 1. Each convolutional layer can be considered a fuzzy filter, which enhances the original signal characteristics and reduces noise. In the convolution layer, the feature vector of the upper layer is convoluted with the convolutional kernel of the current layer. The result of the convolution operation passes through the activation function and then forms the feature map of this layer. The convolution output can be expressed as

$$x_j^l = f(z) = f\left(\sum_{M_j} W_{ij}^l \times x_i^{l-1} + b_j^l\right) i \in M_j$$
(1)

where  $x_j^l$  denotes the feature-vector corresponding to the first convolution kernel of the *j* convolutional layer, and  $M_j$  represents the accepted domain of the current neuron and denotes the *i*-th weighting coefficient of the *j*-th convolutional kernel of the first layer.  $b_j^l$  denotes the offset coefficient corresponding to the *j*-th product of the first layer. The activation function is

$$f(z) = \frac{1}{1 + e^{-z}}$$
(2)

The pooling can be considered as a special kind of convolution. The pooling layer subsamples data using the principle of local correlation and retains useful information while reducing data dimensions.

The pooled operation is used to maintain features, so they possess displacement and zoom invariance. The pooling layer serves the function of secondary feature extraction, and its calculation formula is

$$x_j^l = f\left(W_j^l \times down\left(x_j^{l-1}\right) + b_j^l\right) \tag{3}$$

where  $down(\bullet)$  is the subsampling method,  $W_j^l$  is the weight coefficient, and  $b_j^l$  is the bias coefficient.

Layers	Туре	No. of Neurons (Output Layers)	Kernel Size for Each Output Feature Map	Stride
1	Convolution_1	$N_{con 1} \times 6$	5	1
2	Max pooling	$N_{con 1}/2 \times 6$	2	2
3	Convolution_2	$N_{con 1} \times 12$	5	1
4	SPP	$7 \times 12$	$[N_{con_2}/2, N_{con_2}/4, N_{con_2}/8]$	$[N_{con_2}/2, N_{con_2}/4, N_{con_2}/8]$
5	Fully connected	84	_	-

Table 1. The detailed overview of ECG-SPP-net structure.

#### 2.3. Pre-Processing

A classification system which is composed of pre-processing, feature extraction and classification, is constructed based on the Electrocardiogram-SPP-Net, as shown in Figure 3. In the pre-processing stage, 46 records of the MIT-BIH arrhythmia database containing 100,300 heartbeats were selected. In this database, the ECG first marked the category label of each heartbeat. Then, the ECG signal was cut off into segments according to the label [28]. The label was located at the R peak, which was denoted as  $R_1$ ,  $R_2$ , and  $R_3$  for the three peaks of an ECG signal (Figure 4). The segments, which are segment 1 and segment 2 in Figure 4, were the ECG signals between the two peaks. Then, each segment was broken through its middle section. The anterior of a segment was connected to the posterior of a segment that emerged earlier (Figure 4). The resultant heartbeat contained all the information from the P-wave to the T-wave. Then, each heartbeat was normalized into the range of values between 0 and 1 before sending the preprocessed heartbeat signal into the ECG-SPP-net. Such a large population of heartbeats were classified into six categories, which were normal beat (N), paced beat (/), atrial premature beat (A), premature ventricular contraction (V), left ventricular bundle branch block (L), and right bundle branch block (R). Due to the proportion of the normal heartbeats accounts 73.3% (n = 73,542) of the total samples of the heartbeats, 6000 normal heartbeats were randomly chosen for the classification. The sample set that contains the six kinds of heartbeats is shown in Table 2. Moreover, 70% of heartbeats were also selected from the sample set as the training dataset of the classifier, and the other 30% of beats were used as the test pattern for performance evaluation.

Table 2. Number of beats in the sample set.

Class	Ν	1	Α	V	L	R	Total
Beats	6000	3616	2480	6676	8069	5916	32,757

The WT was utilized as the de-noising method by using a db5 decomposition [25] in three scales with Stein's unbiased likelihood threshold estimator. Subsequently, the baseline drift and noise were moved. Figure 5 displays a comparison of ECG signals between the original one and the de-noised one and such a sample set was taken from the MIT/BIH arrhythmia database. Before sending the pre-processed data into the CNN network, the normalization for all the heartbeat signals was conducted first. The heartbeat signals were bandpass filtered at 0.1–100 Hz and digitized at 360 Hz. The function mapminmax in MATLAB was employed as the method for the amplitude normalization, which puts the amplitude of the sampling point into the interval of [0,1]. Then, such normalized data was fed into the CNN network.





**Figure 4.** The segmentation of the ECG signal into a heartbeat. Three peaks of the ECG signal are represented as  $R_1$ ,  $R_2$ , and  $R_3$ . The peak to peak of the signal was regarded as one segment. The anterior of a segment was connected to the posterior of a segment which emerged earlier. The grey area was a complete heartbeat signal.



Figure 5. Comparison between the original signal and the de-noising signal.

## 2.4. Feature Extraction

CNNs can automatically generate high-level features (i.e., weights and thresholds) through training. First, the sample was sent to the network for training, the input vector was obtained, and the loss function was compared with the given target vector:

$$L = \frac{1}{2} \sum_{k=0}^{n-1} (d_k - y_k)^2 \tag{4}$$

where *L* is the loss function (standard deviation),  $y_k$  is the output vector, and  $d_k$  is the target vector. The weight and threshold values are updated according to *L*, and the update step can be expressed as follows:

$$\Delta W_{jk}(n) = \frac{\alpha}{1+l} \times \left( \Delta W_{jk}(n-1) + 1 \right) \times \delta_k \times h_j \tag{5}$$

where  $\alpha$  represents the learning rate, *j* represents the neural units of the hidden layer, k represents the output layer unit, *M* represents the number of output neuron units, *h<sub>j</sub>* represents the output vector of the hidden layer, *W* represents the adjusted weight, and  $\delta$  is the threshold to be adjusted.

$$\delta k = h_j (1 - h_j) \sum_{k=1}^{M-1} \delta_k W_{jk}$$
(6)

The feature extraction process is shown below.

Step 1: The ECG-SPP-net was initialized by setting the weight *W* as a random number within [0,1]. The threshold value  $\delta$  was set to be 0 and the learning rate  $\alpha$  was defined as 0.1. Finally, the training epochs was set to be 60.

Step 2: The heartbeat from the training set was sent into ECG-SPP-net. The network was trained with one sample for each round due to various sizes of different heartbeats, and the target output vector was set to be  $d_k$ .

Step 3: Calculate the actual output vector  $y_k$  with Equations (1)–(3) and conducted the pooling with the proposed SPP algorithm in Figure 2. Then, the cost function was calculated with Equation (4).

Step 4: The weight W and threshold value  $\delta$  were updated according to Equations (5) and (6).

Steps 2–4 were repeated 60 times and the values of *W* and  $\delta$  were obtained as the high-level features extracted automatically by the ECG-SPP-net. Such high-level features and heartbeat signals from the test sets were then sent to the ECG-SPP-net for testing before sending the results to the classifier.

#### 2.5. Classifier

Softmax regression can solve multiple classification problems relative to the binary classification problem solved by logistic regression. According to a different test input *x*, the probability value *p* was estimated as the result of the classification. The hypothesis function output a *k*-dimensional vector (the sum of vector elements is 1) to represent the estimated probability of *k* categories. The function  $h_{\theta}(x)$  is shown below:

$$h_{\theta}\left(x^{i}\right) = \begin{bmatrix} p\left(y^{i} = 1 \mid x^{i}; \theta\right) \\ p\left(y^{i} = 2 \mid x^{i}; \theta\right) \\ \vdots \\ p\left(y^{i} = 3 \mid x^{i}; \theta\right) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_{j}^{T} x^{i}}} \begin{bmatrix} e^{\theta_{1}^{T} x^{i}} \\ e^{\theta_{2}^{T} x^{i}} \\ \vdots \\ e^{\theta_{k}^{T} x^{i}} \end{bmatrix}$$
(7)

where  $\theta_1, \theta_2, \ldots, \theta_n \in \mathbb{R}^{n+1}$  denote the model parameters, and  $\sum_{j=1}^k e^{\theta_j^T x^i}$  normalized the probability distribution so that the summation of all probabilities is 1. The one with the highest probability was used as the classification result of the test.

## 2.6. Experimental Setting

The ECG-SPP-net was evaluated by comparing the overall accuracies of the proposed method and the other two methods. We used the same denoising method for the three network structures and the same classifier (Softmax). Three network structures are shown in Table 3, where "Y" indicates adoption and "N" indicates none.

Main Operating	Method 1	Method 2	Proposed Method
Fixed-size heartbeat	Y	Y	Ν
CNNs	Y	Y	Y
SPP layer	Ν	Y	Y

Table 3. Three network structures.

Method 1: Heartbeats with different sizes were unified into 300 sampling points [14]. In the MIT-BIH arrhythmia database, the ECG signals were bandpass filtered at 0.1–100 Hz and digitized at 360 Hz. Then, a beat of ECG was resampled to 300 sample points by downsampling or upsampling according to the duration of a heartbeat. The processed heartbeat was sent to the CNN for training. The parameters of layers 1, 2, 3, and 5 of the network were the same as in Table 1. The SPP in layer 4 was removed and alternated to the largest pooling strategy, setting both the pooling size and the step size to be 2.

Method 2: The unified size of this method was the same as in Method 1. The processed heartbeat was sent to ECG-SPP-net. The parameters in each layer were the same as in Table 1.

Proposed method: The number of sample point for each heartbeat was not considered, and as such, the heartbeat was sent to the ECG-SPP-net directly after pre-processing.

## 3. Results and Analysis

As aforesaid, 70% of heartbeats were randomly chosen from the sample set as the training dataset of the classifier and the other 30% of heartbeats were used as the test pattern for performance evaluation. Table 4 shows the confusion matrix for the testing beats of a one-time simulation. Regarding the test dataset, the accuracy of the normal beat reached 99.7%. However, the accuracy of the atrial premature beat was only 71.24%. The average of the accuracy of the six type of beats for one-time simulation was calculated, and the accuracy of the classification for the proposed method reached up to 94%.

	Classification Result						
Ground Truth	N	\	Α	V	L	R	Accuracy
Ν	1794	1	1	0	0	4	99.7%
\	0	1077	0	0	7	1	99.26%
А	43	2	530	63	37	69	71.24%
V	0	6	13	1921	31	32	95.9%
L	0	1	1	5	2402	12	99.2%
R	0	0	4	1	6	1764	99.38%

**Table 4.** Confusion matrix of the heartbeat classification results for the testing beats for the proposed method.

The classification performance was influenced by the training dataset, which was randomly chosen for the classification. To avoid the influence of randomness, the simulation for each network structure was repeated 10 times, and the comparison of the accuracies for three network structures is shown in Figure 6. The accuracies of Method 1 and Method 2 were reduced by nearly 3.6% and 1.5% resulting from the data loss during the sampling process. In addition, the relatively lower accuracy of Method 1 was also derived from the removal of the SPP network structure as compared with the accuracy of Method 2. The two-sided Wilcoxon rank sum test [29] was also employed to evaluate whether the results between different methods had a significant difference, and the *p*-value shown in Table 5 manifests that the result of the proposed method had a significant difference when compared with the other two methods. Therefore, building an SPP structure into the traditional CNN allowed the input of different-sized heartbeats and could extract better features and improve the classification performance.



**Figure 6.** The comparison of accuracies among the three network structures. For each method, we repeated the simulation for each network structure 10 times. The bar is the mean of the accuracy of each method for 10 times. The error is the standard deviation of the accuracy of each method for the experiments repeated 10 times.

Table 5. Wilcoxon rank sum test results of each method against the proposed method.

	Against to	Probability of Accept
Method 1	Proposed method	$1.15  imes 10^{-5}$
Method 2	Proposed method	$1.72  imes 10^{-5}$

## 4. Discussion

An ECG-SPP-net was developed for the heartbeat classification in this work. Heartbeats were filtered between 0.1 to 100 Hz and digitized at 360 Hz. Then, the ECG signals were filtered with wavelet

denoising and segmented into heartbeats with the proposed segmentation method. Each heartbeat contained 200 to 400 sample points. After that, the heartbeat was normalized between 0 to 1 before entering the ECG-SPP-net. Such a preprocessing method could reserve all the information of heartbeat without any distortion, and this was beneficial for the classification accuracy [30–32]. In addition, the influence of the number of the feature image of each convolutional layer to the classification was also considered. The numbers of the feature image for the first convolution layer and the second convolution layer were set to be 6 and 12, respectively. Assuming that a heartbeat contained 300 sample points, and then a three-level pyramid pooling  $(1 \times 1, 1 \times 2, \text{ and } 1 \times 4)$  was adopted. Before sending the input to the fully-connected layer, 84 ( $12 \times 7 \times 1$ ) feature values were acquired, which accounted for 28% of the input data (300 sample points). In Ref. [33], 2400 ( $150 \times 4 \times 4$ ) feature values were achieved, accounting for 45% of the input data ( $73 \times 73$  sample points) before entering the fully-connected layer. Such a setting could achieve a high accuracy of classification of the heartbeats in a relatively short running time of training. The level of pyramid pooling was also considered in this work. For example, the four-level pyramid pooling  $(1 \times 1, 1 \times 2, 1 \times 4, \text{ and } 1 \times 8)$  have been tried in this study. Such a setting increased the training time dramatically with less improvement of classification accuracy.

There are two merits for the proposed design. First, the network allows the entry of different-sized data to the CNN based neural network. Such different-sized data can be trained over a single network to enable weight sharing and avoid the complex operations of multiple network switching [18]. Second, the proposed method is designed on the basis of the CNN, which acts as a feature extractor for simplifying the feature extraction procedure. For some traditional classification methods, the feature extraction of the effective signal should consider many factors including the training and testing performance [12]. In this work, the solely concerned work is the structure of the CNN based network.

Although ECG-SPP-net possesses an advantage of extracting quality features automatically, future research is necessary to address some shortcomings. First, as different-sized heartbeats are sent to the same network for training, the heartbeats can only be sent to the network in a single channel, which leads to a prolonged training time. Second, training deep neural networks requires a large amount of data while the sample set in this paper was limited and therefore not suitable for popular CNN models. In addition, the classification system based on the ECG-SPP-net structure must still be improved in terms of classification accuracy.

#### 5. Conclusions

In this paper, we build an ECG-SPP-net for the classification of heartbeats. Simulation results showed that ECG-SPP-net can extract more representative features than traditional CNNs and has a higher classification accuracy. In the future, more effective structures and optimized parameters based on ECG-SPP-net will be proposed to improve classification performance and reduce the training time.

Author Contributions: Conceptualization—Jia Li, Yujuan Si, Data curation—Liuqi Lang, Lixun Liu, Data analysis—Jia Li, Tao Xu, Writing—Original Draft Jia Li, Yujuan Si, Writing—Edit and Review, Jia Li, Tao Xu.

**Funding:** This work was supported by the Key Scientific and Technological Research Project of Jilin Province (Grant No. 20170414017GH), the Natural Science Foundation of Guangdong Province (Grant No. 2016A030313658), the Innovation and Strengthening School Project (provincial key platform and major scientific research project) supported by Guangdong Government (Grant No. 2015KTSCX175), the Premier Discipline Enhancement Scheme Supported by Zhuhai Government (Grant No. 2015YXXK02-2), and the Premier Key Discipline Enhancement Scheme supported by Guangdong Government Funds (Grant No. 2016GDYSZDXK036).

Conflicts of Interest: The authors declare no conflict of interest.

#### References

1. Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 664–675. [CrossRef] [PubMed]

- Ma, J.; Dong, M. R&D of versatile distributed e-home healthcare system for cardiovascular disease monitoring and diagnosis. In Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Valencia, Spain, 1–4 June 2014; pp. 444–447.
- 3. Shen, C.; Kao, W.; Yang, Y.; Hsu, M.; Wu, Y.; Lai, F. Detection of cardiac arrhythmia in electrocardiograms using adaptive feature extraction and modified support vector machines. *Expert Syst. Appl.* **2012**, *39*, 7845–7852. [CrossRef]
- Sharma, S.; Nagal, D. Identification of QRS complexes in single-lead ECG Using LS-SVM. In Proceedings of the International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), Jaipur, India, 25 September 2014; pp. 1–4.
- 5. Fei, S.W. Diagnostic study on arrhythmia cordis based on particle swarm optimization-based support vector machine. *Expert Syst. Appl.* **2010**, *37*, 6748–6752. [CrossRef]
- 6. Gui, X.; Han, L.; Guo, S. An ECG fuzzy classification method based on adaptive PSO-RBF algorithm. *J. Am. Coll. Cardiol.* **2016**, *68*, C111.
- El-Khafif, S.H.; El-Brawany, M.A. Artificial neural network-based automated ECG signal classifier. *ISRN Biomed. Eng.* 2013, 2013. Available online: http://dx.doi.org/10.1155/2013/261917 (accessed on 31 August 2018). [CrossRef]
- 8. Minami, K.; Nakajima, H.; Toyoshima, T. Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network. *IEEE Trans. Biomed. Eng.* **1999**, *46*, 179–185. [CrossRef] [PubMed]
- 9. Kim, J.; Shin, H.S.; Shin, K.; Lee, M. Robust algorithm for arrhythmia classification in ECG using extreme learning machine. *Biomed. Eng. Online* **2009**, *8*, 31. [CrossRef] [PubMed]
- Li, S.; Liu, G.; Lin, Z. Comparisons of wavelet packet, lifting wavelet and stationary wavelet transform for de-noising ECG. In Proceedings of the IEEE International Conference on Computer Science and Information Technology, Beijing, China, 8–14 August 2009; pp. 491–494.
- 11. Al-Nashash, H. A dynamic Fourier series for the compression of ECG using FFT and adaptive coefficient estimation. *Med. Eng. Phys.* **1995**, *17*, 197–203. [CrossRef]
- 12. Khorrami, H.; Moavenian, M. A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification. *Expert Syst. Appl.* **2010**, *37*, 5751–5757. [CrossRef]
- 13. Cvetkovic, D.; Übeyli, E.D.; Cosic, I. Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: A pilot study. *Digit. Signal Process.* **2008**, *18*, 861–874. [CrossRef]
- 14. Liu, T.; Si, Y.; Wen, D.; Zang, M.; Lang, L. Dictionary learning for VQ feature extraction in ECG beats classification. *Expert Syst. Appl.* **2016**, *53*, 129–137. [CrossRef]
- Acharya, U.R.; Fujita, H.; Lih, O.S.; Adam, M.; Tan, J.H.; Chua, C.K. Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network. *Knowl. Based Syst.* 2017, 132, 62–71. [CrossRef]
- Acharya, U.R.; Fujita, H.; Lih, O.S.; Hagiwara, Y.; Tan, J.H.; Adam, M. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Inf. Sci.* 2017, 405, 81–90. [CrossRef]
- 17. Liu, F.; Lin, G.; Shen, C. CRF learning with CNN features for image segmentation. *Pattern Recognit.* **2015**, *48*, 2983–2992. [CrossRef]
- 18. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef] [PubMed]
- 19. Qu, T.; Zhang, Q.; Sun, S. Vehicle detection from high-resolution aerial images using spatial pyramid pooling-based deep convolutional neural networks. *Multimedia Tools Appl.* **2017**, *76*, 21651–21663. [CrossRef]
- 20. Yue, J.; Mao, S.; Li, M. A deep learning framework for hyperspectral image classification using spatial pyramid pooling. *Remote Sens. Lett.* **2016**, *7*, 875–884. [CrossRef]
- 21. Üstündağ, M.; Gökbulut, M.; Şengür, A.; Ata, F. Denoising of weak ECG signals by using wavelet analysis and fuzzy thresholding. *Netw. Model. Anal. Health Inf. Bioinf.* **2012**, *1*, 135–140. [CrossRef]
- 22. Sayadi, O.; Shamsollahi, M.B. ECG denoising and compression using a modified extended Kalman filter structure. *IEEE Trans. Biomed. Eng.* **2008**, *55*, 2240–2248. [CrossRef] [PubMed]
- 23. Lu, G.; Brittain, J.S.; Holland, P.; Yianni, J.; Green, A.L.; Stein, J.F.; Aziz, T.Z.; Wang, S. Removing ECG noise from surface EMG signals using adaptive filtering. *Neurosci. Lett.* **2009**, *462*, 14–19. [CrossRef] [PubMed]
- 24. Alfaouri, M.; Daqrouq, K. ECG signal denoising by wavelet transform thresholding. *Am. J. Appl. Sci.* 2008, *5*, 276–281. [CrossRef]

- Wu, D.; Bai, Z. An improved method for ECG signal feature point detection based on wavelet transform. In Proceedings of the IEEE Conference on Industrial Electronics and Applications (ICIEA), Singapore, 26 November 2012; pp. 1836–1841.
- 26. Zeiler, M.D.; Fergus, R. Stochastic pooling for regularization of deep convolutional neural networks. In Proceedings of the International Conference on Learning Representation, Scottsdale, AZ, USA, 4 May 2013; pp. 1–9.
- Serre, T.; Wolf, L.; Poggio, T. Object recognition with features inspired by visual cortex. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; pp. 994–1000.
- 28. Liu, M.; Li, G.; Hao, H.; Hou, Z.; Liu, X.T. Wave Shape Classification Based on Convolutional Neural Network. *Acta Autom. Sin.* **2016**, *42*, 1339–1346.
- 29. Gibbons, J.D.; Chakraborti, S. *Nonparametric Statistical Inference. International Encyclopedia of Statistical Science,* 1st ed.; Springer: Berlin, Germany, 2011; pp. 977–979.
- 30. Zhao, J.; Wong, P.K.; Ma, X.; Xie, Z. Chassis integrated control for active suspension, active front steering and direct yaw moment systems using hierarchical strategy. *Veh. Syst. Dyn.* **2017**, *55*, 72–103. [CrossRef]
- 31. Zhao, J.; Wong, P.K.; Ma, X.; Xie, Z. Design and analysis of an integrated SMC-TPWP strategy for a semi-active air suspension with stepper motor-driven GFASA. *Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng.* **2018**. [CrossRef]
- 32. Ma, X.; Wong, P.K.; Zhao, J. Practical multi-objective control for automotive semi-active suspension system with nonlinear hydraulic adjustable damper. *Mech. Syst. Signal Process.* **2019**, *117*, 667–688. [CrossRef]
- 33. Zhai, X.; Tin, C. Automated ECG Classification using Dual Heartbeat Coupling based on Convolutional Neural Network. *IEEE Access* 2018, *6*, 27465–27472. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).