IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 31, 2019, accepted February 17, 2019, date of publication February 21, 2019, date of current version March 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2900719

Noise Rejection for Wearable ECGs Using Modified Frequency Slice Wavelet Transform and Convolutional Neural Networks

ZHONGYAO ZHAO[1], CHENGYU LIU[1], YAOWEI LI[1], YIXUAN LI[1], JINGYU WANG[2], BOR-SHYH LIN[3], (Senior Member, IEEE), AND JIANQING LI[1,4]

[1]School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China
[2]College of Engineering, Northeastern University, Boston, MA 02115, USA
[3]Institute of Imaging and Biomedical Photonics, National Chiao Tung University, Tainan 71150, Taiwan
[4]School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 210029, China

Corresponding author: Chengyu Liu (chengyu@seu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 81871444 and in part by the Primary Research and Development Plan of Jiangsu Province under Grant BE2017735.

ABSTRACT Progress in wearable techniques makes the long-term daily electrocardiogram (ECG) monitoring possible. However, the long-term wearable ECGs can be significantly contaminated by various noises, which affect the detection and diagnosis of cardiovascular diseases (CVDs). The situation becomes more serious for wearable ECG screening, where the data are huge, and doctors have no way to visually check the signal quality episode-by-episode. Therefore, automatic and accurate noise rejection for the wearable big-data ECGs is craving. This paper addressed this issue and proposed a noise rejection method for wearable ECGs based on the combination of modified frequency slice wavelet transform (MFSWT) and convolutional neural network (CNN). Wearable ECGs were recorded using the newly developed 12-lead Lenovo smart ECG vest with a sample rate of 500 Hz and a resolution of 16 bit. One thousand 10-s ECG segments were picked up and were manually labeled into three quality types: clinically useful segments with good signal quality (type A), clinically useful segments with poor signal quality (type B), and clinically useless segments (pure noises, type C). Each of the 1,000 10-s ECG segments were transformed into a 2-D time-frequency (T-F) image using the MFSWT, with a pixel size of 200×50. Then, the 2-D grayscale images from MFSWT were fed into a 13-layer CNN model for training the classification models. Results from the standard 5-folder cross-validation showed that the proposed combination method of MFSWT and CNN achieved a highest classification accuracy of 86.3%, which was higher than the comparable methods from continuous wavelet transform (CWT) and artificial neural networks (ANN). The combination of MFSWT and CNN also had a good calculation efficiency. This paper indicated that the combination of MFSWT and CNN is a potential method for automatic identification of noisy segments from wearable ECG recordings.

INDEX TERMS Wearable ECG, signal quality assessment (SQA), convolutional neural network (CNN), modified frequency slice wavelet transform (MFSWT).

I. INTRODUCTION

Cardiovascular disease (CVD) is one of the major diseases that cause human death. According to the latest World Health Organization (WHO) report, about 17.5 million people died from CVDs in 2012, accounting for 30% of all global deaths [1]. The incidence of CVD deaths is predicted to rise to 23 million by 2030 [1]. The costs for CVD-related treatment including medication are also substantial. CVD-related cost in the low- and middle-income countries over the period 2011–2025 is estimated approximately 3.8 trillion U.S. dollars [2]. CVD, which is hailed as ''the world's first killer'', is increasingly threatening people's health and life. Therefore, as a comprehensive reflection of cardiac activity, electrocardiogram (ECG) signal is of great significance in clinic [3]. Advancement of wearable technology has enabled the recording of long-term dynamic ECGs, making behavior monitoring under daily life possible. Monitoring of long-term

The associate editor coordinating the review of this manuscript and approving it for publication was Wanqing Wu.

dynamic ECGs can improve the shortcomings of the traditional short-term ECGs collection in rest state.

Dynamic ECGs are usually weak and are easy to be contaminated by different noises and other human electrical signals, such as electromyographic (EMG) signal. Under different conditions, type and intensity of signal noises are significantly different, which brings the challenge for signal analysis and the following disease diagnosis. Therefore, signal quality assessment for dynamic ECGs is useful [4], [5], and can help to reject the signal episodes with low signals quality [6], [7].

Convolutional neural network (CNN), as a typical deep learning method, allows the 2-D image data to be fed to machine and automatically discover the key signal features without manual intervention and experts' prior knowledge [8]. Thus, it can be used to process the ECG-generated images, by optimizing the filters in the CNN model. In the current study, we tried the CNN method for classifying ECG 2-D images with different signal quality levels. In this way, we can select clinically valuable ECG episodes and reject the noisy ones.

Meanwhile, a time-frequency (T-F) method, named modified frequency slice wavelet transform (MFSWT), was used to convert 1-D ECG signal into 2-D T-F image data, which can be fed to the CNN classifier. There are many common T-F transformation methods, such as short-time Fourier transform (STFT), Wigner-Ville distribution (WVD), and continuous wavelet transform (CWT) [9], as well as the MFSWT method proposed in 2017 [10]. MFSWT can efficiently contain the time-frequency information, such as P-wave, QRS complex and T-wave, in the transformed 2-D images.

In this study, a newly developed 12-lead Lenovo Smart ECG vest was used to record the wearable ECG signals with a sample rate of 500 Hz and a resolution of 16 bits. One thousand 10-s ECG segments were picked up with strict signal quality manually labelling with three quality types: clinically useful segments with good signal quality (type A), clinically useful segments with poor signal quality (type B) and clinically useless segments (pure noises, type C) [11]. Each of the 1,000 10-s ECG segments was transformed into a 2-D time-frequency image using MFSWT, with a pixel size of 200×50. Then, the 2-D image is fed into the CNN to train the classification model. The trained model was evaluated by the indices including accuracy, sensitivity and specificity with a 5-folder cross validation method [12].

II. METHODS

A. MODIFIED FREQUENCY SLICE WAVELET TRANSFORM (MFSWT)

ECG signal, as a continuous time signal closely related to the physiological phenomenon of human diseases, generally has quasi-periodic, non-stationary and dynamic characteristics. Existing study showed that T-F representation method can be better to localize time and frequency information contained in the special components in ECGs, such as P-wave, QRS complex and T-wave [13]. STFT, WVD and CWT are typical T-F representation methods. Yan *et al.* [9] proposed a new T-F method named as frequency slice wavelet transform (FSWT) to accurately locate the signal feature in 2-D T-F plane. FSWT is essentially an extension of the short-time Fourier transform in frequency domain. The advantages include (1) the T-F window of the transform is the center of the observed T-F window; (2) the frequency resolution is controllable; (3) the frequency slice function (FSF) can be freely designed; (4) the signal reconstruction and FSFs are independent of each other [14]. However, FSWT has drawbacks in the application since the T-F positioning accuracy is affected by the transformation scale, and transformation scale is difficult to determine in actual use [15]. Thus, Luo *et al.* [10] proposed an improved MFSWT method and applied it to the heartbeat T-F spectrum analysis. Since MFSWT follows the rules of producing T-F representation from the frequency domain, this method has the characteristics of high T-F resolution and accurate positioning of frequency components of the region of interest.

A bound signal-adaptive FSF was introduced in MFSWT to serve as a dynamic frequency filter. It can realize the adaptive measurement process of signal energy distribution at different observation frequencies. FSF of the narrow frequency window is used to observe the frequency components with small energy, while FSF of wide frequency window is used to observe the frequency components with large energy. The model of MFSWT was expressed as follows:

Assume $\hat{f}(k)$ is the Fourier transform of $f(t)$. MFSWT is expressed in frequency domain as,

$$W_f(t,\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(k)\hat{p}^*\left(\frac{\mu-\omega}{q(f(k))}\right)e^{-ikt}dk \qquad (1)$$

where t and ω are observed time and frequency, respectively, $*$ represents conjugation operator, $\hat{p}$ is the FSF defined as $\hat{p}(x) = e^{-x^2/2}$, q is defined as a scale function of $\hat{f}(k)$. Since q is a function of signal frequency conversion, it enables the transform with signal-adaptive property. In the T-F analysis of the sign signal, q uses the function form of Eq. (2):

$$q = \delta + sign\left(\nabla|\hat{f}(\mu)\right) \qquad (2)$$

where δ is the frequency position of the signal main component. It can be estimated from the frequency position corresponding to maximum $\left|\hat{f}(\mu)\right|$. $\nabla(\cdot)$ is a differential operator, and $sign(\cdot)$ means signum function, which returns 1 if the input is greater than zero, 0 if it is zero, or -1 if it is less than zero. FSF here adopts a Gaussian function form and $\hat{p}(0) = 1$ is always true. Then the original signal can be reconstructed as follows.

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_f(t,\omega)e^{i\omega(t-\tau)}d\tau d\omega \qquad (3)$$

Fig.1 shows the examples of 10-s ECGs, as well as their corresponding MFSWT images. Signals a, d and g in Fig.1 are three typical 10-s ECG segments with different signal quality levels. Signal a is an example of 10-s ECG segments with

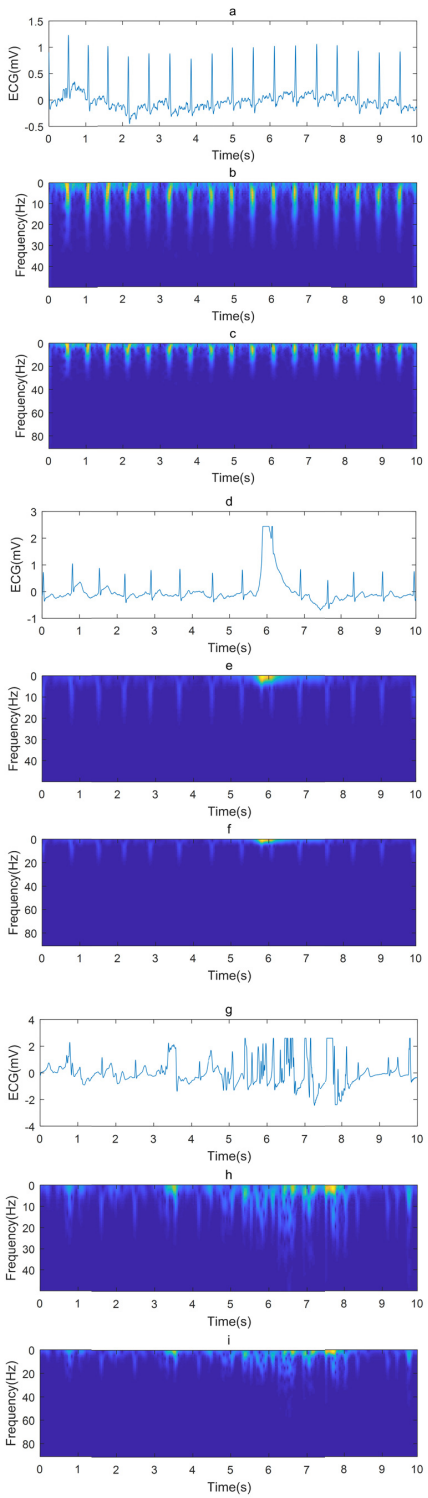


FIGURE 1. Examples of 10-s ECG segments and their corresponding MFSWT images: a. Example of type A ECG signal, b. MFSWT image of a type A ECG with pixel of 200 × 50, c. MFSWT image of a type A ECG with pixel of 560 × 420, d. Example of type B ECG signal, e. MFSWT image of a type B ECG with pixel of 200 × 50, f. MFSWT image of a type B ECG with pixel of 560 × 420, g. Example of type C ECG signal, h. MFSWT image of a type C ECG with pixel of 200 × 50.

very good signal quality (Type A in this study), which clinicians can use to analyze both the rhythm and morphology information. Signal d is an example of 10-s ECG segments with poor signal quality (Type B), but it is still clinically useful. ECG rhythm information can be obtained from this type signal while ECG morphology information cannot be obtained. Signal g is an example of 10-s ECG segments with very serious noises (Type C). Clinically useful information can be hardly obtained from this type signal. For MFSWT images, the time-domain characteristics in ECG waveform, such as P-wave, QRS complex and T-wave have been accurately located in the signal spectrum. In this study, MFSWT was used to generate spectrograms of an ECG signal for CNN-based classifier input. Figure c, f, and i are the initial T-F images with pixel of 560 × 420 obtained by MFSWT. We found that the lower half of the images are roughly the same and the amount of information is very small. In order to improve the training efficiency, we have intercepted and resized the initial T-F graph and got Figure b, e, h with pixel of 200 × 50.

B. CONVOLUTIONAL NEURAL NETWORK (CNN)

The ability of multi-layer CNN model trained with gradient descent to learn complex, high-dimension, non-linear mappings from large collections of examples makes it obvious candidate for image recognition tasks [16]–[22]. CNN has become a popular method for feature extraction and classification without requiring pre-processing and pre-training algorithm [23]. It is a composition of sequences of functions or layers that maps an input vector to an output vector. The input x_k^l is expressed as:

$$x_k^l = \sum_{i=1}^{N_{l-1}} conv2D\left(w_{ik}^{l-1}, s_i^{l-1}\right) + b_k^l \tag{4}$$

Similarly, b_k^l and w_{ik}^{l-1} are the bias and kernel of the k-th neuron at layer l, respectively, s_i^{l-1} is the output of the i-th neuron at layer $l-1$, and $conv2D(.,.)$ means a regular 2-D convolution without zero padding on the boundaries. So the output y_k^l can be described as:

$$y_k^l = f(\sum_{i=1}^{N_{l-1}} conv2D\left(w_{ik}^{l-1}, s_i^{l-1}\right) + b_k^l) \tag{5}$$

Since the single-layer perceptron model cannot handle the linear indivisibility problem, in 1986, Rumelhart *et al.* [24] proposed a multilayer feedforward network trained by the error inverse propagation algorithm—back propagation (BP) network. This method solves the problems that some single-layer perceptrons can't solve. Assuming that the corresponding output vector of the input is $\left[y_1^L, y_2^L, \ldots, y_{N_L}^L\right]$, and its ground truth class vector is $\left[t_1, t_2, \ldots, t_{N_L}\right]$. Then the loss function and incremental error are:

$$E = E\left(y_1^L, y_2^L, \ldots, y_{N_L}^L\right) = \sum_{i=1}^{N_L} \left(y_i^L - t_i\right)^2 \tag{6}$$

$$\Delta_k^l = \frac{\partial E}{\partial x_k^l} \tag{7}$$

Then the network's weight update formula is,

$$\omega_{im}^{l-1} = \omega_{im}^{l-1} - \eta \times \frac{\partial E}{\partial \omega_{im}^{l-1}} \tag{8}$$

where η is the learning rate. The implementation of CNN [8] is summarized in Fig.2.

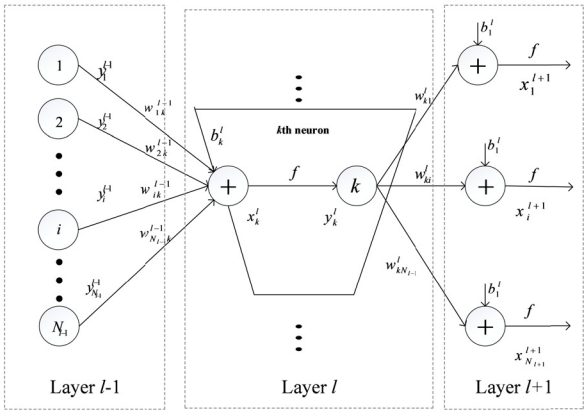


FIGURE 2. The implementation of CNN.

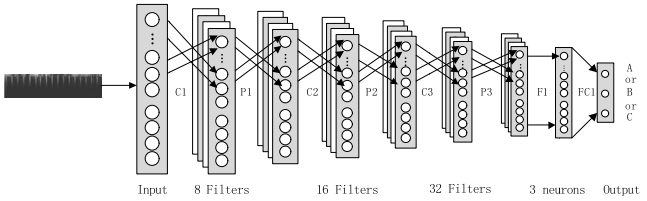


FIGURE 3. The architecture of the network.

A 13-layer CNN structure was constructed. Except the input and output layers, it includes three convolution layers, three ReLU layers, three maximum pooling layers, a flatten layer and a fully connected layer. The implementation process of CNN actually includes the feature extraction process. The model structure in Fig.3 visually shows the process of CNN extracting features. Fig.3 illustrates the architecture of the implemented network and its detailed components for each layer.

The convolution layer extracts various local features of the previous layer by convolution operation [25]. The role of the pooling layer is to semantically combine similar features. The pooling layer makes the feature robust to noise and deformation through pooling operations [26]. The features extracted by each layer represent the original image from different angles in an enhanced manner. As the number of layers increases, its representation becomes more and more abstract [27]. The Flatten layer is used to ''flatten'' the input, that is, to make multidimensional input one-dimensional. Each neuron in the fully connected layer is fully connected to its previous layer. This layer combines the various local features extracted in the previous stage, and finally obtains the posterior probability of each category through the output layer. Table 1 gives the specific parameter settings (after optimization) for the CNN architecture used in this study.

C. COMPARABLE METHODS

To compare the efficiency and accuracy of the proposed method, we chose other T-F analysis and machine learning methods to train the classification model. CWT [28] was used as the comparable method of MFSWT and multilayer perceptron classifier artificial neural networks (ANN) [29] was used as the comparable method of CNN.

TABLE 1. CNN specifications designed for the ECG classification problem.

Parameters	Values
Learning rate	$0.009\times0.96^{(\text{global step}/50)}$
First convolutional layer kernel size	7×7
No. of feature maps in the first convolutional	8
Second convolutional layer kernel size	7×8
No. of feature maps in the second convolutional	16
Third convolutional layer kernel size	7×7
No. of feature maps in the third convolutional	32
Max pooling layer kernel size	2
No. of neurons in the fully connected layer	3
No. of epoch	40
No. of minimal batches	160

TABLE 2. Distribution of each quanlity type of signal.

Signal type	Number	Note
A	595	clinically useful segments with relatively clear and stable signal waveform.
B	189	clinically useful segments with at least several identifiable beats in succession as well as slight noise.
C	216	clinically useless segments which are pure noise.

III. EXPERIMENTAL DESIGN

A. DATA

The data were recorded from a newly developed 12-lead Lenovo Smart ECG device. The database consists of 1,000 10-s ECG segments. ECGs have a sampling rate of 500 Hz and a resolution of 16 bits. These 10-s ECG segments have been manually labeled as three quality types: clinically useful segments with good signal quality (Type A), clinically useful segments with poor signal quality (Type B), and clinically useless segments (pure noise, type C). The distribution of ECG signals of each quality type is shown in Table 2.

B. COMPARATIVE EXPERIMENT

As for CWT, we used wavelet base ''haar'' to transform these 1-D ECG segments into 2-D T-F grayscale images. Then the 2-D images were fed into the 13-layer CNN model and the ANN model respectively for training. As for multilayer perceptron classifier ANN, the T-F image was further transferred into a 1-D vector and were input into the ANN for training and testing. Specific parameter setting for ANN was showed in Table 3.

C. MODEL TRAINING AND TEST

We employ a balanced image data set to train the model. That is, we divided 1,000 records into 5 subsets, and the

TABLE 3. The optimal ANN specifications designed for the ECG classification problem.

Parameters	Values
Learning rate	0.1
No. of hidden layers	1
No. of hidden units in each layer	16
No. of epoch	100

number of records in each subset was 200. The experiment was evaluated using a 5-folder cross-validation. When testing with the first subset, the 200 records in the first subset was used for verification, and the remaining 800 records were used to train the model. Four combinations of different T-F analysis and machine learning methods were tested, i.e., CWT+ANN, CWT+CNN, MFSWT+ANN, and the new proposed MFSWT+ CNN used in this study.

IV. RESULTS

A. PERFORMANCE METRICS

Six widely used metrics, i.e. sensitivity (*Se*), positive predictive rate (*P*+), accuracy (*Acc*), F-measure, area under the ROC curve (AUC) and Kappa coefficient, were used for evaluation of the trained models. Acc and Kappa coefficient refer to the overall system performance, while the remaining indexes is specific to each class, and they measure the generalization ability of the classification algorithm to differentiate events.

$N1_k$ denotes the number of ECG segments in the k-th type correctly identified as being of the k-th type. $N2_k$ denotes the number of ECG segments in the k-th type falsely identified as being of the other two types. $N3_k$ denotes the number of ECG segments in the other two types falsely identified as being of the k-th type. The indices are defined for each type as:

$$Se_k = \frac{N1_k}{N1_k + N2_k} \tag{9}$$

$$P+_k = \frac{N1_k}{N1_k + N3_k} \tag{10}$$

$$Acc = \frac{2 \times \sum_{k=1}^{3} N1_k}{2 \times \sum_{k=1}^{3} N1_k + \sum_{k=1}^{3} N2_k + \sum_{k=1}^{3} N3_k} \tag{11}$$

F-measure is a statistic that is often used to evaluate the quality of a model. It is defined as:

$$F_\beta = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \tag{12}$$

where β is the parameter, P is the precision rate, its value is equal to $P+$, R is the recall rate, and its value is equal to *Se*. Take $\beta = 1$, and $F_1 = \frac{2PR}{P+R}$.

Kappa coefficient can verify consistency and can be used to measure classification accuracy. It is defined as:

$$K = \frac{p_o - p_e}{1 - p_e} \tag{13}$$

where p_o is the overall classification accuracy, and its value is equal to *Acc*. Suppose that the number of real samples in each class is a_1, a_2, a_3 respectively, and the number of samples predicted in each class is b_1, b_2, b_3. The total number of samples is n, and $p_e = \frac{a_1 \times b_1 + a_2 \times b_2 + a_3 \times b_3}{n \times n}$.

ROC curve is based on a series of different ways of classification (boundary value or decision threshold), with true positive rate (*Se*) as the ordinate, the false positive rate (*P*+) as the abscissa, and AUC is defined as the area under the ROC curve, often used to evaluate the classifier with imbalance data. Each folder is tested by a specific classifier with the same parameters and the results are summarized in Table 4. The overall Kappa coefficient is between 0.61 and 0.80, indicating that the model classification results are highly consistent with the actual type of the sample.

TABLE 4. The experimental results.

Fold	F_1			AUC			Kappa
	A	B	C	A	B	C	
1	0.952	0.788	0.976	0.926	0.836	0.977	0.870
2	0.919	0.633	0.884	0.904	0.780	0.926	0.746
3	0.898	0.424	0.659	0.892	0.647	0.786	0.565
4	0.967	0.618	0.674	0.951	0.749	0.793	0.717
5	0.996	0.846	0.867	0.994	0.926	0.909	0.893
Total	0.939	0.661	0.806	0.933	0.786	0.876	0.755

Fig.4 shows the classification results for the testing data, presented for each folder of the 5-folder cross validation. In each panel, the elements in the main diagonal of the matrix denote the number of ECG segments in the k-th type correctly identified as being of the k-th type, i.e., $N1_k$. The sum of the elements in each row of the metric except the main diagonal element denotes the number of ECG segments in the k-th type falsely identified as being of the other two types, i.e., $N2_k$. The sum of the elements in each column of the matrix except the main diagonal element denotes the number of ECG segments in the other two types falsely identified as being of the k-th type, i.e., $N3_k$. The bottom-right corner element in each panel is the *Acc* result. *Acc* is the radio of the number of correct predicted labels and total number of the labels.

For type A, *Se* values were 100%, 90.8%, 85.7%, 100% and 100% respectively for the five folders, and the corresponding *P*+ values were 90.8%, 93.1%, 94.4%, 93.7% and 99.2% respectively. For type B, *Se* values were 68.4%, 65.8%, 47.4%, 55.3% and 89.2% respectively for the five folders, and *P*+ values were 92.9%, 61%, 38.3%, 70% and 80.5%, respectively. For type C, *Se* values were 95.3%, 88.4%, 67.4%, 67.4% and 81.8% and *P*+ values were 100%, 88.4%, 64.4%, 67.4% and 92.3% respectively. Totally, the 5-folder cross validation achieved *Se* values of 95.3%, 65.1% and 80.1% respectively for types A, B and C, and *P*+ values of 92.5%, 67.2% and 81.2%, resulting in a total mean *Acc* value of 86.3%. Standard deviation of *Acc*from 5-folder

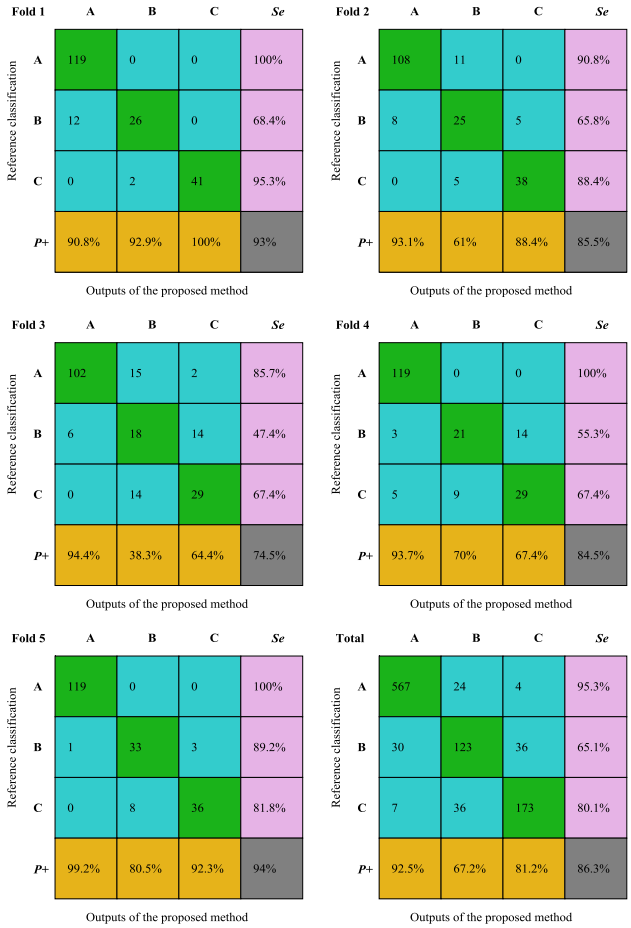


FIGURE 4. Classification results for folders 1∼5 and the total results. Bottom-right corner element in each panel shows accuracy (Acc) result.

TABLE 5. Comparable results from four combination of time-frequency analysis and machine learning methods. Significant differences were indicated as '∗' when compared with the MFSWT + CNN method.

Method	*Acc* (%)	Time consuming (s)
CWT+ANN	69.5±9.00	0.43
CWT+CNN	68.2±6.28	1.17
MFSWT+ANN	81.5±7.30	0.44
MFSWT+CNN	86.3±7.03	1.23

cross validation was 7.03%, indicating the good stability of developed method.

B. COMPARABLE RESULTS

Table 5 shows the results between the comparable T-F analysis/machine learning methods and the method proposed in this study. Herein, we tested four combinations of T-F analysis and machine learning. When using the 2-D images from CWT, classification accuracies for the three signal quality groups were not high, with 69.5% ± 9.00% for CWT&ANN and 68.2% ±6.28% for CWT&CNN. When using the new MFSWT images, the classification accuracies enhanced higher than 80%, and output 81.5% ± 7.30% for MFSWT&ANN and 86.3% ± 7.03% for MFSWT&CNN. Compared with the combination of MFSWT&CNN, other three combination types output significantly lower classification accuracy (*Acc*) from the t-test (as shown in Table 5).

In addition, compared with CNN, ANN classifier had lower calculation cost, less than 450 ms for each record. The experiment was performed using an Intel i5-5200U CPU, a 2.2 GHz dual core clock and a 4 GB DDR3L memory.

V. DISCUSSION AND CONCLUSION

In this work, we presented a unique architecture of CNN to distinguish noisy ECGs from clinically useful ECG segments with the generated T-F images using MFSWT method. High classification results were obtained from the combination of MFSWT method and CNN classifier.

Traditionally, signal quality assessment for ECG signal is a two-classification problem (good or poor quality) [30]–[32]. Studies on three signal quality classification is very rare. A significant feature of this study is the three type classifications for ECG signal quality, which is from the clinical reequipment for long-term ECG monitoring and analysis since in real clinical application, rhythm and morphological features can be calculated from ECGs with good signal quality (type A in this study) while only rhythm features can be calculated from ECGs with poor signal quality (type B). We used the real collected wearable ECG data, rather than the common online public databases, to develop the classification model, to ensure the clinical usefulness.

Another novelty of this study is the MFSWT-based T-F analysis method, which can better capture the tiny changes in the frequency domain and provide better image input for CNN classifier. Luo *et al.* [10] compared the effect of MFSWT with WVD and CWT methods and confirmed that MFSWT was better than comparable methods in the application scenario of heartbeat type classification. Xu *et al.* [33] applied this method in the detection of atrial fibrillation (AF), and enhanced the AF detection accuracy up to 84.85%. We used this T-F analysis method in this study and obtained an accuracy of 86.3% in three signal quality classification for wearable ECGs.

Limitations should be revealed. Frist, the number of signal segments is not large. More data should be further expanded in the follow-up study. Secondly, the labels of signal quality need to be further considered. Although we used a manual manner, there were still some samples with controversy. We identify the studies on automatic labeling algorithm should be our future works. Third, the parameter setting of the used CNN model needs further optimization. The current CNN model parameters were only optimized using a simple grid search method. Further optimization on the parameters is needed. Besides, our accuracy is a bit low because we are concerned about the three type classification problem. In the pre-experiment, we used the same method to do the two-classification experiment, and the accuracy reached more than 90%, similar to the results of other existing studies.

In short, a novel framework based on MFSWT T-F representation and CNN classifier for ECG signal quality

classification was proposed, which could accurately identify the noisy ECG segments for a relatively large amount of data. The final classification accuracy was 86.3% when using a standard 5-folder cross validation, indicating the clinical significance of the method. Considering that CVD is a dynamical process, it is necessary for patients with CVD to be monitored using wearable ECG devices. The proposed method can be used in this situation since it is simple and efficient.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

AVAILABILITY OF DATA

The authors would like to thank the support from the Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent monitoring Lab. The data used in the current study are not publicly available due to the patients' privacy but are available from the corresponding author on reasonable request.

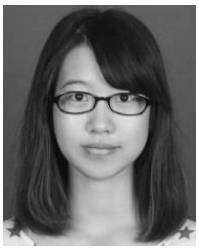
REFERENCES

[1] World Health Organization. (2013). *Cardiovascular Disease*. [Online]. Available: http://www.who.int/cardiovascular_diseases/en/index.html
[2] D. E. Bloom, D. Chisholm, E. Jane-Llopis, K. Prettner, A. Stein, and A. Feigl, "From burden to 'best buys': Reducing the economic impact of non-communicable disease in low-and middle-income countries," Program Global Demogr. Aging, PGDA Working Papers 7511, 2011.
[3] S. S. Barold, "Willem Einthoven and the birth of clinical electrocardiography a hundred years ago," *Cardiac Electrophysiol. Rev.*, vol. 7, no. 1, pp. 99–104, 2003.
[4] Q. Li, R. G. Mark, and G. D. Clifford, "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter," *Physiol. Meas.*, vol. 29, no. 1, pp. 15–32, 2008.
[5] A. E. W. Johnson, J. Behar, F. Andreotti, G. D. Clifford, and J. Oster, "Multimodal heart beat detection using signal quality indices," *Physiol. Meas.*, vol. 36, no. 8, pp. 1665–1677, 2015.
[6] L. P. Wang and J. Dong, "The advance research and analysis of electrocardiogram pattern classification," *Chin. J. Biomed. Eng.*, vol. 29, no. 6 pp. 916–925, 2010.
[7] Y.-T. Zhang, C.-Y. Liu, S.-S. Wei, C.-Z. Wei, and F.-F. Liu, "ECG quality assessment based on a kernel support vector machine and genetic algorithm with a feature matrix," *J. Zhejiang Univ. Sci. C*, vol. 15, no. 7, pp. 564–573, 2014.
[8] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ECG classification by 1-D convolutional neural networks," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.
[9] Z. Yan, A. Miyamoto, Z. Jiang, and X. Liu, "An overall theoretical description of frequency slice wavelet transform," *Mech. Syst. Signal Process.*, vol. 24, no. 2, pp. 491–507, 2010.
[10] K. Luo, J. Li, Z. Wang, and A. Cuschieri, "Patient-specific deep architectural model for ECG classification," *J. Healthcare Eng.*, vol. 2017, 2017, Art. no. 4108720.
[11] G. D. Clifford, J. Behar, Q. Li, and I. Rezek, "Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms," *Physiol. Meas.*, vol. 33, no. 9, pp. 1419–1433, Aug. 2012.
[12] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. Int. Joint Conf. Artif. Intell.*, 1995, pp. 1137–1145.
[13] M. Orini, R. Bailon, L. T. Mainardi, P. Laguna, and P. Flandrin, "Characterization of dynamic interactions between cardiovascular signals by time-frequency coherence," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 3, pp. 663–673, Mar. 2012.
[14] Z. Yan, A. Miyamoto, and Z. Jiang, "Frequency slice wavelet transform for transient vibration response analysis," *Mech. Syst. Signal Process.*, vol. 23, no. 5, pp. 1474–1489, 2009.
[15] Z. Yan, A. Miyamoto, and Z. Jiang, "Frequency slice algorithm for modal signal separation and damping identification," *Comput. Struct.*, vol. 89, nos. 1–2, pp. 14–26, 2011.
[16] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[17] D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification," in *Proc. IJCAI Int. Joint Conf. Artif. Intell.*, vol. 22, no. 1, 2015, p. 1237.
[18] D. Zhang, D. Meng, and J. Han, "Co-saliency detection via a self-paced multiple-instance learning framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 865–878, May 2017.
[19] D. Zhang, J. Han, C. Li, J. Wang, and X. Li, "Detection of co-salient objects by looking deep and wide," *Int. J. Comput. Vis.*, vol. 120, no. 2, pp. 215–232, 2016.
[20] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
[21] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, Jun. 2015.
[22] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, and F. Wu, "Background prior-based salient object detection via deep reconstruction residual," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 8, pp. 1309–1321, Aug. 2015.
[23] S. Chaib, H. Yao, Y. Gu, and M. Amrani, "Deep feature extraction and combination for remote sensing image classification based on pre-trained CNN models," in *Proc. Int. Conf. Digit. Image Process.*, 2017, Art. no. 104203D.
[24] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[26] L. P. Jin and J. Dong, "Deep learning research on clinical electrocardiogram analysis," *Sci. China, Inf. Sci.*, vol. 45, no. 3, pp. 398–416, 2015.
[27] C. Liu *et al.*, "Signal quality assessment and lightweight QRS detection for wearable ECG SmartVest system," *IEEE Internet Things J.*, to be published. doi: 10.1109/JIOT.2018.2844090.
[28] H. C. Jebril and L. Nabli, "ECG waves determining using multi-scaled CWT," in *Proc. Int. Conf. Control, Decis. Inf. Technol. (CoDIT)*, May 2013, pp. 178–184. doi: 10.1109/CoDIT.2013.6689540.
[29] R. Ghongade and A. Ghatol, "A robust and reliable ECG pattern classification using QRS morphological features and ANN," in *Proc. IEEE Region 10 Conf. (TENCON)*, Nov. 2008, pp. 1–6.
[30] U. Satija, B. Ramkumar, and M. S. Manikandan, "Real-time signal quality-aware ECG telemetry system for IoT-based health care monitoring," *IEEE Internet Things*, vol. 4, no. 3, pp. 815–823, Jun. 2017.
[31] C. Liu, P. Li, L. Zhao, F. Liu, and R. Wang, "Real-time signal quality assessment for ECGs collected using mobile phones," in *Proc. Comput. Cardiol.*, vol. 38, Sep. 2011, pp. 357–360.
[32] I. Silva, G. B. Moody, and L. Celi, "Improving the quality of ECGs collected using mobile phones: The PhysioNet/computing in cardiology challenge 2011," in *Proc. Comput. Cardiol.*, vol. 38, Sep. 2011, pp. 273–276.
[33] X. Xu, S. Wei, C. Ma, K. Luo, L. Zhang, and C. Liu, "Atrial fibrillation beat identification using the combination of modified frequency slice wavelet transform and convolutional neural networks," *J. Healthcare Eng.*, vol. 2018, 2018, Art. no. 2102918.

ZHONGYAO ZHAO is currently pursuing the degree in instrument science and engineering with Southeast University, China, where she is currently a Research Assistant with Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Lab. Her research interests include signal processing, machine learning, deep learning, and big data.

CHENGYU LIU received the B.S. and Ph.D. degrees in biomedical engineering from Shandong University, China, in 2005 and 2010, respectively. He has completed a postdoctoral trainings at Shandong University, China, from 2010 to 2013, Newcastle University, U.K., from 2013 to 2014, and Emory University, USA, from 2015 to 2017. He is currently a Professor with the School of Instrument Science and Engineering, Southeast University, China. He is also the Director of Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Lab. He was the PI of 10+ awarded grants. He has published over 150 papers. He is currently a Federation Journal Committee Member of the International Federation for Medical and Biological Engineering and the Chair of China Physiological Signal Challenge. His research interests include mHealth and intelligent monitoring, machine learning and big data processing for cardiovascular signals, device development for CADs, and sleep and emotion monitoring.

YAOWEI LI received the B.S. degree in mathematics and applied mathematics from Shandong Normal University, China, in 2017. She is currently a Research Assistant with Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Lab, Southeast University, China. Her research interests include artificial intelligence, machine learning, and deep learning.



YIXUAN LI is currently pursuing the degree in instrument science and engineering with Southeast University, China, where she is currently a Research Assistant with Southeast-Lenovo Wearable Heart-Sleep-Emotion Intelligent Monitoring Lab. Her research interests include biomedical signal processing and machine learning.

JINGYU WANG received the B.S. degree from the Nanjing University of Posts and Telecommunications, China, and the B.S. degree from the New York Institute of Technology, New York, NY, USA, in 2018. He is currently pursuing the master's degree with Northeastern University, Boston, MA, USA. His research interest include data analytical, big data management, and deep learning.

BOR-SHYH LIN (M'02–SM'15) received the B.S. degree from National Chiao Tung University, Hsinchu, Taiwan, in 1997, and the M.S. and Ph. D. degrees from the Institute of Electrical Engineering, National Taiwan University, Taipei, Taiwan, in 1999 and 2006, respectively. He is currently a Professor with the Institute of Imaging and Biomedical Photonics, National Chiao Tung University. His current research interests include biomedical circuits and systems, biomedical signal processing, and biosensors.

JIANQING LI received the B.S. and M.S. degrees in automatic technology from the School of Instrument Science and Engineering, Southeast University, China, in 1986 and 1992, respectively, and the Ph.D. degree in measurement technology and instruments from Southeast University, where he is a Professor with the School of Instrument Science and Engineering. He is currently a Professor with the School of Basic Medical Sciences, Nanjing Medical University, China, where he acts as the Vice Present. His research interersts include mHealth, wearable ECG systems, multi-parameter physiological signal detection, and wireless networks.