Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

ECG quality assessment based on hand-crafted statistics and deep-learned S-transform spectrogram features

GUOYANG LIU[a], XIAO HAN[a,b], LAN TIAN[a,b,*], WEIDONG ZHOU[a], HUI LIU[b]

[a] *School of Microelectronics, Shandong University, Jinan 250100, PR China*
[b] *Shandong Artificial Intelligence Institute, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, PR China*

ARTICLE INFO

Article history:
Received 9 January 2021
Accepted 1 July 2021

Keywords:
Stockwell transform
Feature fusion
Convolutional neural network
ECG quality assessment

ABSTRACT

Background and Objective Electrocardiogram (ECG) quality assessment is significant for automatic diagnosis of cardiovascular disease and reducing the massive workload of reviewing continuous ECGs. Hence, how to design an appropriate algorithm for objectively evaluating the multi-lead ECG recordings is particularly important. Despite the deep learning methods performing well in many fields, as a data-driven method, it may not be entirely suitable for ECG analysis due to the difficulty in obtaining sufficient data and the low signal-to-noise ratio of ECG recordings. In this study, with the aim of providing an accurate and automatic ECG quality assessment scheme, we propose an innovative ECG quality assessment algorithm based on hand-crafted statistical features and deep-learned spectral features.

Methods In this paper, a novel approach, combining the deep-learned Stockwell transform (S-Transform) spectrogram features and hand-crafted statistical features, is proposed for ECG quality assessment. Firstly, a double-input convolutional neural network (CNN) is established. Then, the S-Transform with a novel online augmentation scheme is performed on the multi-lead raw ECG signal received from one input layer to obtain proper time-frequency representation. After that, the CNN with three convolutional layers is employed to extract robust deep-learned features automatically. Simultaneously, the hand-crafted statistical features, including lead-fall, baseline drift, and R peak features, are calculated and fed into another input layer for feature fusion training. Finally, the deep-learned and hand-crafted features are concatenated and further fused by a fully connected layer for quality classification. Furthermore, a log-odds analysis scheme combining with a gradient-based method can localize the abnormal zone in time, frequency, and spatial domains.

Results and Conclusion Our proposed method is evaluated on a publicly available database with 10-fold cross-validation. The experimental results demonstrate that the proposed assessment algorithm reached a mean accuracy of 93.09%, a mean F1-score of 0.8472, and a sensitivity of 0.9767. Moreover, comprehensive experiments indicate that the fusion of CNN features and statistical features has complementary advantages and ideal interpretability, achieving end-to-end multi-lead ECG assessment with satisfying performance.

1. Introduction

As a non-invasive and inexpensive diagnostic tool, Electrocardiogram (ECG) has been widely applied in diagnosing cardiovascular disease [1–3]. Traditional ECG analysis requires doctors to diagnose and treat patients according to their ECG waveform information. However, the ECG signals recorded by wearable devices can be significantly contaminated by numerous noises such as electromyographic (EMG) signals and low-voltage signals. It has been shown that poor-quality ECGs can cause multiple false alarms that may seriously threaten patient safety [4,5]. Worse yet, some noises often have similar morphology and overlapped frequency bands with the normal ECG recordings [6]. This makes the traditional ECG signal quality analysis based on manual observation time-consuming, laborious, and subjective [7,8]. Therefore, an automatic ECG quality assessment method with high performance is highly in demand for addressing these issues. Simultaneously, with the development of 5G technology and health Internet of Things technology, the remote ECG monitoring and diagnosis system have

[*] Corresponding author at: School of Microelectronics, Shandong University, Jinan 250100, PR China.
E-mail address: tianlan65@sdu.edu.cn (L. TIAN).

https://doi.org/10.1016/j.cmpb.2021.106269
0169-2607/© 2021 Elsevier B.V. All rights reserved.

also put forward higher performance requirements for the automatic quality assessment of ECG signals [9].

The existing mainstream approaches of ECG quality assessment can be approximately concluded into two categories. The first one is the traditional machine learning-based methods. For example, Li et al. [10] extracted more than ten signal quality metrics such as pSQI, baseSQI, and pcaSQI, and SVM is utilized to conduct the quality assessment; Johannesen [11] demonstrated a rule-based ECG quality assessment algorithm by leveraging hand-crafted features such as global high-frequency noise, global low-frequency noise, and average RR interval. The heuristically determined threshold is set for assessing the ECG quality. Clifford et al. [12] presented a data fusion scheme for determining the acceptability of ECGs collected in noisy ambulatory environments. Six SQIs, including iSQI, bSQI, fSQI, sSQI, kSQI, and pSQI and five classifiers (i.e., NB, SVM, MLP, and ANN) were utilized for ECG quality assessment. Johannesen and Galeotti [13] proposed a two-step threshold algorithm to realize quality classification, which first rejects ECGs with macroscopic errors and subsequently quantifies the noise on a continuous scale. Recently, Zhang et al. [14] employed the waveform features, including lead-fall feature, baseline wander feature and baseline drift feature, power spectrum feature, and non-linear features for feeding into random forest and support vector machine (SVM) to classify the ECG quality. For hand-crafted features, each feature has a specific physical meaning and corresponds to a specific description for ECG signals. However, it is difficult to represent ECG signal features from all aspects, even if all hand-crafted features are integrated.

Another mainstream approach is deep learning methods. Some studies have shown that time-frequency transformation can reveal time-frequency domain characteristics more comprehensively in the particular components of ECGs, such as P-wave, QRS complex, and T-wave [15–17]. The time-frequency representation of the raw ECG signals can increase the dimensionality. Hence, it is usually classified by the deep learning-based methods, which is proved to be more suitable for handling high-dimensionality data such as multi-channel images [18,19]. Zhao et al. [20] adopted the modified frequency slice wavelet transform (MFSWT) to extract ECG features into the 2-D time-frequency representation. The transformed image-based data was sent into a Convolutional Neural Network (CNN) for subsequent tri-class classification. Huerta et al. [21] transformed the raw ECG signals into 2-D image-based representation by continuous wavelet transform (CWT), and then the pre-trained Alexnet is utilized for finetuning. Zhang et al. [22] applied the Short-Time Fourier Transform (STFT) to acquire the time-frequency spectrum of the ECG recordings, and then they were fed into a CNN branch for feature extraction. This feature is integrated with another CNN branch feature for the final decision. Besides, Zhang et al. [23] collected multiple ECG features in terms of spectral distribution, signal complexity, horizontal and vertical variation of waves and sent them into a 7-layer Long Short-Term Memory (LSTM) network to better capture the time-related features. Furthermore, Zhou et al. [24] constructed a 1D-CNN to classify single-lead ECG signals on two publicly available databases and obtained a satisfactory performance. However, this method can only deal with the single-lead ECGs. Compared with the hand-crafted features, the deep-learned features describe the ECG recordings from another point of view. Though deep learning approaches and some feature fusion methods were investigated in many fields [25–27], the interpretability and the relationship between these features have seldomly been presented.

S-Transform is another time-frequency analysis method proposed by Stockwell et al. [28], which inherits and develops STFT and CWT. It has been widely applied in ECG signal analysis. For example, Ari et al. [29] leveraged the S-Transform for ECG signal enhancement, removing noise components from the time–frequency domain represented noisy ECG signal. Zidelmal et al. [30] adopted the S-Transform in QRS detection and tested their algorithm with the MIT-BIH arrhythmia database (MITDB). In this study, we introduce the S-Transform combining with CNN for ECG quality assessment. This is the first attempt to implement ECG quality assessment by adopting S-Transform analysis to the best of our knowledge.

In this paper, the S-Transform spectrogram is calculated for time-frequency image-based representation, and its feature is automatically extracted by CNN modules. On the other hand, statistical features (i.e., hand-crafted features), including Lead-fall, Baseline-drift, and R peak features, are collected and combined with CNN features for feature fusion decision. To overcome the limitation of the few training samples, a novel online augmentation method is proposed to improve the generalization ability of the model significantly. Moreover, we introduce a log-odds analysis method to measure the contribution of each type of feature and employ a gradient-based method to localize the abnormal ECG recordings in time-frequency domain.

The rest of the paper is organized as follows: Section 2 gives a detailed description of the employed database and the proposed method. Section 3 demonstrates the experimental results and discusses the significance of our work. Finally, Section 4 concludes the advantages and limitations of the proposed method.

2. Materials and methods

2.1. Materials

In this paper, the database is from the Physionet/CinC Challenge 2011, recorded at 500 Hz, 16 bit per sample, and 5μV resolution. Each 12-lead ECG recording (leads I–III, aVR, aVL, aVF, V1–V6) was 10s long and had been bandpass filtered within 0.05–100 Hz. All the ECG recordings were manually annotated by 23 volunteers, who identified themselves as 2 cardiologists, 1 (non-cardiologist) physician, 5 ECG analysts, 5 others with some experience reading ECGs, and 10 volunteers who had never read ECGs previously. Each ECG recording was randomly presented to volunteers that gave a grade of *A*, *B*, *C*, *D*, and *F*, which corresponded to numerical values of 0.95, 0.85, 0.75, 0.6, and 0. Most of the volunteers graded a few of the ECGs more than once as a result of the random selection process. Then, the average grade value was calculated for each ECG recordings. The ECG recording with which at least two grade values were available, average value greater than 0.7, and no more than one grade annotated as *F* would be labeled as ‘acceptable’. On the other hand, if the average value could not reach 0.7 and at least two grade values were available, it would be labeled as ‘unacceptable’. Otherwise, the ECG recording would be labeled as ‘indeterminate’ [31]. The labeled ECG recordings are divided into dataset *A* (set A, 1000 ECG recordings) and dataset *B* (set *B*, 500 ECG recordings), where dataset *B* serving as a test set is not publicly available. This paper adopts the set *A* (including 773 acceptable ECG signals and 225 unacceptable ECG signals) to evaluate our proposed ECG quality assessment algorithm.

2.2. Methods

2.1.1. S-Transform spectrogram

S-Transform maintains a direct relationship with the Fourier transform, which ensures efficient computing speed. At the same time, it has different resolutions at different frequencies. The Gaussian window of S-Transform can provide higher time resolution of high frequency and higher frequency resolution of low frequency. Besides, the frequency of normal ECG signals is relatively low, and therefore the S-Transform spectrum is an effective method to characterize ECG signals. The hyper-parameter *p* is introduced to control the resolution of the S-Transform. With the increase of p value, the Gaussian window width of S transformation increases, resulting in the decrease of frequency domain resolution of high frequency ECG signal and the increase of time domain resolution of low frequency ECG signal. In this study, the p value is set to 0.3. The S-Transform fine-tuned with parameter p is given as follows:

$$S(\tau, f) = \frac{|f|}{p\sqrt{2\pi}} \int_{-\infty}^{+\infty} x(t)\, e^{-\frac{(t-\tau)^2 f^2}{2p^2}} e^{-2i\pi ft} dt \tag{1}$$

where $x(t)$ is the ECG signal to be analyzed; τ and f are observed time and frequency respectively; .. represents the time-frequency matrix obtained by S-Transform. S-Transform is also known as "phase orthogonal" CWT. Thus, from the perspective of CWT, S-Transform can be written as follows:

$$S(\tau, f) = e^{-2\pi i \tau f} \sqrt{|f|} W(\tau, a) \tag{2}$$

where a represents the scale inversely proportional to frequency, and $W(\tau, a)$ is the CWT of the signal with a special complex Morlet wavelet satisfying the following equation:

$$\Phi(t) = \frac{1}{p\sqrt{2\pi}} e^{-\frac{t^2}{2p^2}} e^{-2i\pi t} \tag{3}$$

Finally, the S-Transform spectrogram of ECG signals is used to represent the energy distribution of ECG signals in the time-frequency domain, as shown below:

$$|S(\tau, f)|^2 = S(\tau, f) S^*(\tau, f) \tag{4}$$

In this work, the frequency range of the S-Transform spectrogram is selected to be between 1 and 25 Hz, which is based on the fact that the main frequency band of the ECG signal (including QRS complex, P wave, and T wave) is concentrated in this range. Compared with the time domain waveform of ECG signal, the S-Transform spectrogram of ECG signal can reflect the time-frequency domain characteristics of ECG signal more precisely so as to describe the dynamic change process of ECG signal more intuitively. Therefore, S-Transform is an effective method to evaluate the quality of ECG signals.

Fig. 1 shows the S-Transform spectrogram of acceptable and unacceptable ECG records. It can be seen that the acceptable raw ECG waveform is regular, and its corresponding S-Transform spectrogram has many regular ridge-like localized patterns. Though some noises exist, they do not affect the identification of the QRS complex. However, the unacceptable ECG segment is obviously irregular. As illustrated in Fig. 1(b) and (d), the Gaussian noises that occurred in the record may be caused by EMG signals and other artificial interferences, making this ECG recording labeled ‘Unacceptable’.

2.2.2. Statistical feature extraction

Feature 1: Lead-fall. In the dynamic ECG collection procedure, poor electrode contact or lead movement could cause a signal waveform that seems like a straight line. In this case, the ECG tends to be of unacceptable quality. Hence, the number of continuous constant voltage in each lead is calculated to describe this feature. Fig. 2 demonstrates an example of the lead fall. Let Num_n denotes the maximum number of the identical continuous value in the n^{th} lead. Then, the lead-fall feature vector $\mathbf{F}_{lf} \in \mathbb{R}^{1\times 12}$ can be expressed as:

$$\mathbf{F}_{lf} = [Num_1, Num_2, Num_3, \ldots, Num_{12}] \tag{5}$$

Feature 2: Baseline drift. Baseline drift is one of the main noises in online ECG collection. ECG recordings with too severe baseline drift cannot be used as a reference for clinical diagnosis. In this study, we filter the original signal using an 8-order low-pass Butterworth filter with a cut-off frequency of 0.01 Hz and then extract the maximum value to estimate the extent of baseline drift. An example corresponding to this feature is illustrated in Fig. 3. Let Bas_n denotes the maximum value of n^{th} lead over the filtered signal. The baseline drift feature $\mathbf{F}_{bd} \in \mathbb{R}^{1\times 12}$ is defined as follows:

$$\mathbf{F}_{bd} = [Bas_1, Bas_2, Bas_3, \ldots, Bas_{12}] \tag{6}$$

Feature 3: R peak features. Since R peak is the symbolic band of ECG signal, as presented in Fig. 4, we adopt the number of absolute values that equal to the maximum as the quality index of R peaks. Let Mas_n denote the number of absolute values that equal to the maximum in n^{th} lead, then the R peak feature$\mathbf{F}_r \in \mathbb{R}^{1\times 12}$can be described as:

$$\mathbf{F}_r = [Mas_1, Mas_2, Mas_3, \ldots, Mas_{12}] \tag{7}$$

Let $\mathbf{F}_{SF} \in \mathbb{R}^{3\times 12}$ be the total statistical feature matrix:

$$\mathbf{F}_{SF} = \begin{bmatrix} F_{lf1} & \cdots & F_{lfN} \\ F_{bd1} & \cdots & F_{bdN} \\ F_{r1} & \cdots & F_{rN} \end{bmatrix} \tag{8}$$

where N is the number of the lead.

2.2.3. Double-input deep convolutional neural network

As a kind of popular artificial neural network, convolutional neural network has been widely used in image recognition. In this paper, a double-input model of convolutional neural network is designed and its architecture is shown in Fig. 5. The training process of the whole network is described by a pseudo-code flow chart (see Fig. 6). In the network initialization phase, three types of statistical features are extracted to serve as the input data. In the training stage, the dataset is divided into mini-batch and then fed into the model for subsequent training. For each iteration, we randomly select 3000 points from each sample with length of 5000 points, and then down-sample ten times to reduce computational burden. After that, the S-Transform is applied to each trail. This procedure can be seen as a certain online augmentation method for improving the generalization performance of the model. After transformed ECG data passing through three CNN modules (each CNN module including a convolution layer, a batch normalization layer, and a maxpooling layer), the obtained feature map is flattened and connected with corresponding statistical features prepared in the initialization stage, and then fused through a fully-connected layer, which is connected with softmax layer. The final softmax mapping score is compared with the corresponding input label to calculate the cross-entropy loss value. Finally, the network is updated through back propagation and the trained network is obtained. It is noteworthy that actually, the input length of the ST branch is 6 s, which is randomly selected in the training stage to perform online-augmentation. Hence, in the inference stage, we segment the signals with the rectangle windows in 0–6 s,2–8 s,4–10 s, and then execute the forward propagation three times. Finally, three softmax mapping scores are averagely fused to obtain the final label.

The output softmax score of the proposed double-input CNN can represent the probability of a sample belongs to a certain class. Let $\mathbf{X} \in \mathbb{R}^{2340}$ be the concatenated feature vector, $\mathbf{w}_1 \in \mathbb{R}^{2340}$ and $\mathbf{w}_2 \in \mathbb{R}^{2340}$ the weights of fully-connected layer for acceptable class and unacceptable class, b_1 and b_2 the corresponding bias term, and $z_1 = \mathbf{w}_1^\top \mathbf{X} + b_1$ the output value of fully-connected layer that corresponds to the acceptable class. Then the probability P that represents a sample belonging to the acceptable class can be expressed as follows:

$$\begin{aligned} P &= \mathrm{softmax}(z_1) \\ &= \frac{\exp\left(\mathbf{w}_1^\top \mathbf{X} + b_1\right)}{\exp\left(\mathbf{w}_1^\top \mathbf{X} + b_1\right) + \exp\left(\mathbf{w}_2^\top \mathbf{X} + b_2\right)} \\ &= \frac{1}{1 + \exp\left(\Delta \mathbf{w}^\top \mathbf{X} + \Delta b\right)} \end{aligned} \tag{9}$$

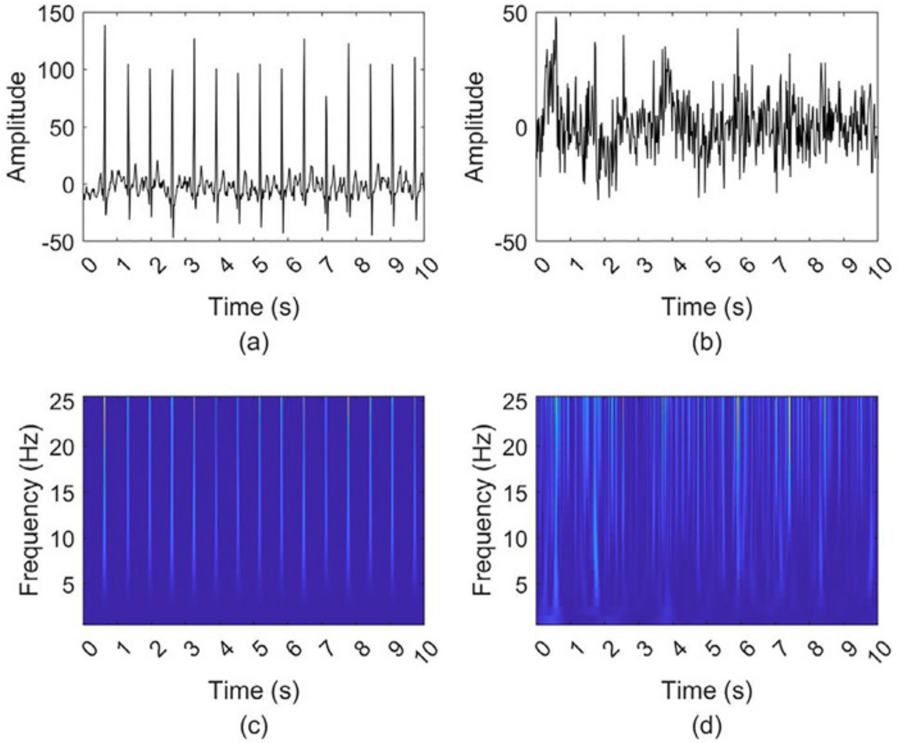


Fig. 1. Example of two ECG recordings and their corresponding S-Transform spectrogram. (a) and (b) depict the acceptable and unacceptable single-lead ECG signals randomly selected in the database, and (c) and (d) are their corresponding S-Transform spectrograms.

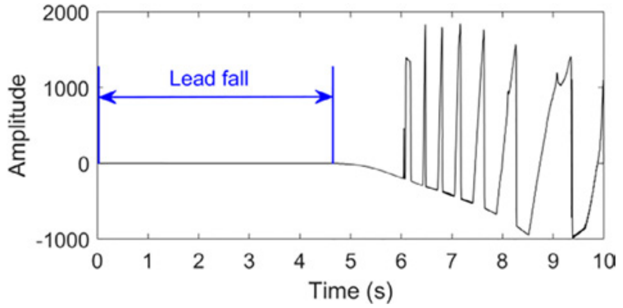


Fig. 2. An example of ECG recording consisting of lead fall. The recording values between two blue lines are equal to zero due to the lead fall.

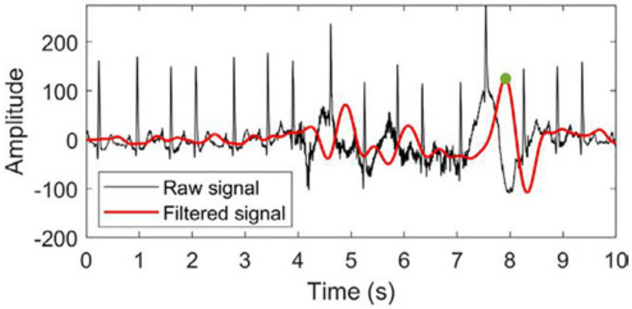


Fig. 3. An example of baseline drift feature extraction. The bold red line is the signal filtered by a low-pass filter, and the solid green dot indicates the maximum value of the filtered signal.

where $\Delta\mathbf{w} = \mathbf{w}_2 - \mathbf{w}_1$ and $\Delta b = b_2 - b_1$. In statistical analysis, odds are an expression of relative probabilities which is defined as the ratio of the probability of two opposed events. Let $C = 12$ be the number of leads, then the odds for acceptable class are given

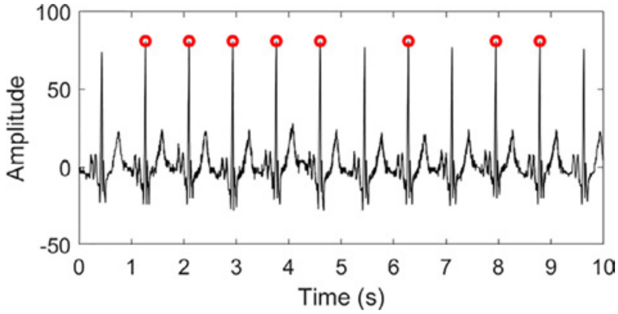


Fig. 4. An example of extracting R peak features. The red circles marked on raw ECG signal denote the R peaks.

as:

$$
\begin{aligned}
O_p &= \tfrac{P}{1-P} \\
&= \exp\left(-\Delta\mathbf{w}^{\top}\mathbf{X} - \Delta b\right) \\
&= \exp\left(-\left(\Delta\mathbf{w}_h^{\top}\mathbf{X}_h + \Delta\mathbf{w}_d^{\top}\mathbf{X}_d\right) - \Delta b\right) \\
&= \exp\left(-\left(\sum_{c=1}^{C}\Delta\mathbf{w}_{hc}^{\top}\mathbf{X}_{hc} + \Delta\mathbf{w}_d^{\top}\mathbf{X}_d\right) - \Delta b\right) \\
&= \exp\left(-\Delta b\right)\exp\left(-\Delta\mathbf{w}_d^{\top}\mathbf{X}_d\right)\prod_{c=1}^{C}\exp\left(-\Delta\mathbf{w}_{hc}^{\top}\mathbf{X}_{hc}\right)
\end{aligned}
\tag{10}
$$

where $\mathbf{X}_h \in \mathbb{R}^{36}$ is the flattened vector of the $\mathbf{F}_{SF}$, $\mathbf{X}_{hc} \in \mathbb{R}^3$ the statistical feature vector extracted in c^{th} lead, $\mathbf{X}_d \in \mathbb{R}^{2304}$ the flattened feature map of the last max-pooling layer. $\mathbf{w}_h$ and $\mathbf{w}_d$ are the learned weights corresponding to $\mathbf{X}_h$ and $\mathbf{X}_d$. According to the definition of the odds, we define $O_{hc} = \exp(-\Delta\mathbf{w}_{hc}^{\top}\mathbf{X}_{hc})$ as the corresponding odds of statistical features in each lead, $O_d = \exp(-\Delta\mathbf{w}_d^{\top}\mathbf{X}_d)$ as the corresponding odds of deep-learned features, and $O_b = \exp(-\Delta b)$ as the corresponding odds of bias term, which is a constant value. Then the odds O_p can be expressed

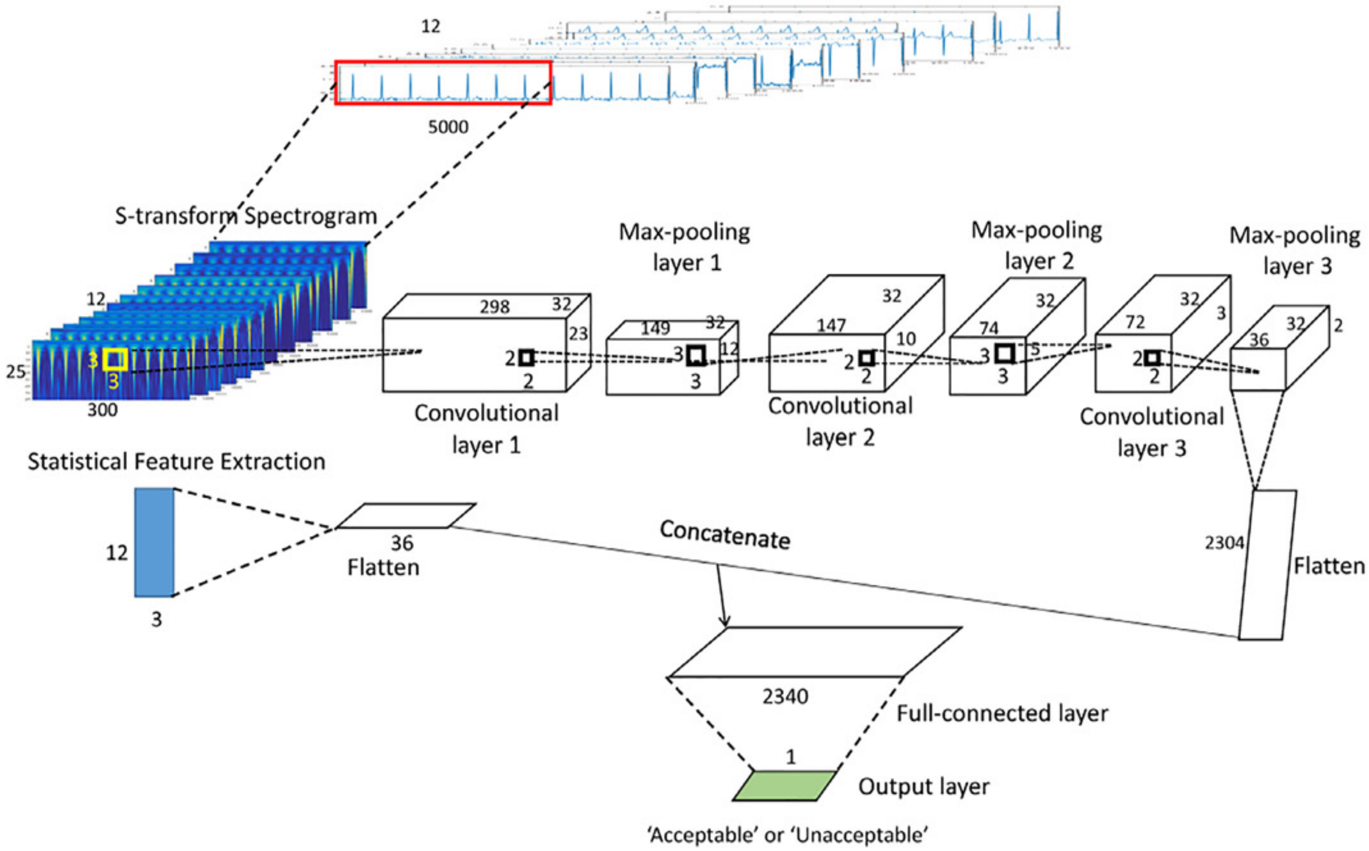
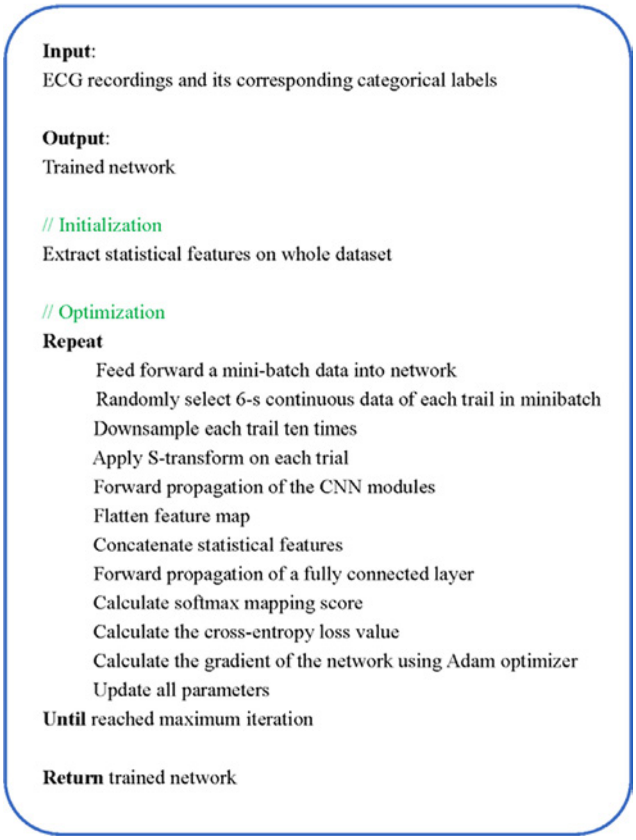


Fig. 5. The proposed double-input CNN architecture.



```
Input:
ECG recordings and its corresponding categorical labels

Output:
Trained network

// Initialization
Extract statistical features on whole dataset

// Optimization
Repeat
    Feed forward a mini-batch data into network
    Randomly select 6-s continuous data of each trail in minibatch
    Downsample each trail ten times
    Apply S-transform on each trial
    Forward propagation of the CNN modules
    Flatten feature map
    Concatenate statistical features
    Forward propagation of a fully connected layer
    Calculate softmax mapping score
    Calculate the cross-entropy loss value
    Calculate the gradient of the network using Adam optimizer
    Update all parameters
Until reached maximum iteration

Return trained network
```

Fig. 6. The pseudocode of model training process.

as the product of these sub-odds:

$$O_p = O_b O_d \underbrace{\left(O_{h1} O_{h2} \cdots O_{hC}\right)}_{O_h} \tag{11}$$

and the corresponding log-odds are as follows:

$$\ln\left(O_p\right) = \ln\left(O_b\right) + \ln\left(O_d\right) + \underbrace{\sum_{c=1}^{C} \ln\left(O_{hc}\right)}_{\ln\left(O_h\right)} \tag{12}$$

We can observe that the contribution of each type of feature can be simply evaluated by log-odds, and the larger the absolute log-odds, the more significant the impact on output score. It is worth noting that the contribution of statistical features corresponding to each lead can also be computed.

The learning rate of the proposed network is 0.0005 and the coefficient of L_2 norm is set to 0.0005. The whole CNN model is updated with Adam optimizer [32] for 500 epochs. In this study, all experiments are carried out in MATLAB 2020a, running in a workstation with a i9-9820 × 3.30 GHz CPU, a NVIDIA GTX 2080 SUPER GPU and 64 GB memory.

3. Results and discussion

3.1. Performance metrics

In this study, we calculate Sensitivity, Specificity, Precision, F1-score, and Accuracy to comprehensively evaluate the proposed algorithm. Assume positive be the acceptable ECG signal, and negative be the unacceptable ECG signal, and let TP, TN, FN, and FP be the abbreviation of true positive, true negative, false negative and

Table 1
The ablation study results on five comparative experiments. The highest accuracy of each evaluation index is marked in boldface.

No.	Method	Sensitivity	Specificity	Precision	F1-Score	Accuracy
A	SF	94.70%	75.56%	93.01%	83.38%	90.38%
B	ST-CNN	96.64%	66.67%	90.88%	76.91%	89.88%
C	AugST-CNN	**98.19%**	62.67%	90.04%	73.90%	90.18%
D	ST-CNN+SF	95.86%	76.00%	93.21%	83.73%	91.38%
E	AugST-CNN+SF	97.67%	**77.33%**	**93.67%**	**84.72%**	**93.09%**

false positive. Then these evaluation indicators can be expressed as follows:

$$\text{Sensitivity} = \text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{14}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

$$F1 - \text{score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{16}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FN + FP} \tag{17}$$

3.2. Ablation study

Table 1 shows a series of ablation experiments related to the proposed method and Fig. 7 illustrates five confusion matrices for corresponding experiments. Experiment *A* establishes a fully connected neural network and takes the statistical features (SF) as the input. Experiment *B* and experiment *C* construct the S-Transform CNN (ST-CNN) and online augmented ST-CNN (AugST-CNN), respectively. Experiment *D* and experiment *E* adopt the double-input CNN architecture, which can take advantage of the statistical features and deep-learned features. Similarly, compared with experiment *D*, the online augmentation method is added in experiment *E*. From experiments *A* and *B*/*C*, we can observe that the Sensitivity of the ST-CNN method outperforms the hand-crafted feature method, and experiment *C* performs the best Sensitivity at 98.19%. At the same time, the Specificity is not as good as the hand-crafted feature extraction method. In other words, the ST-CNN method tends to rec-

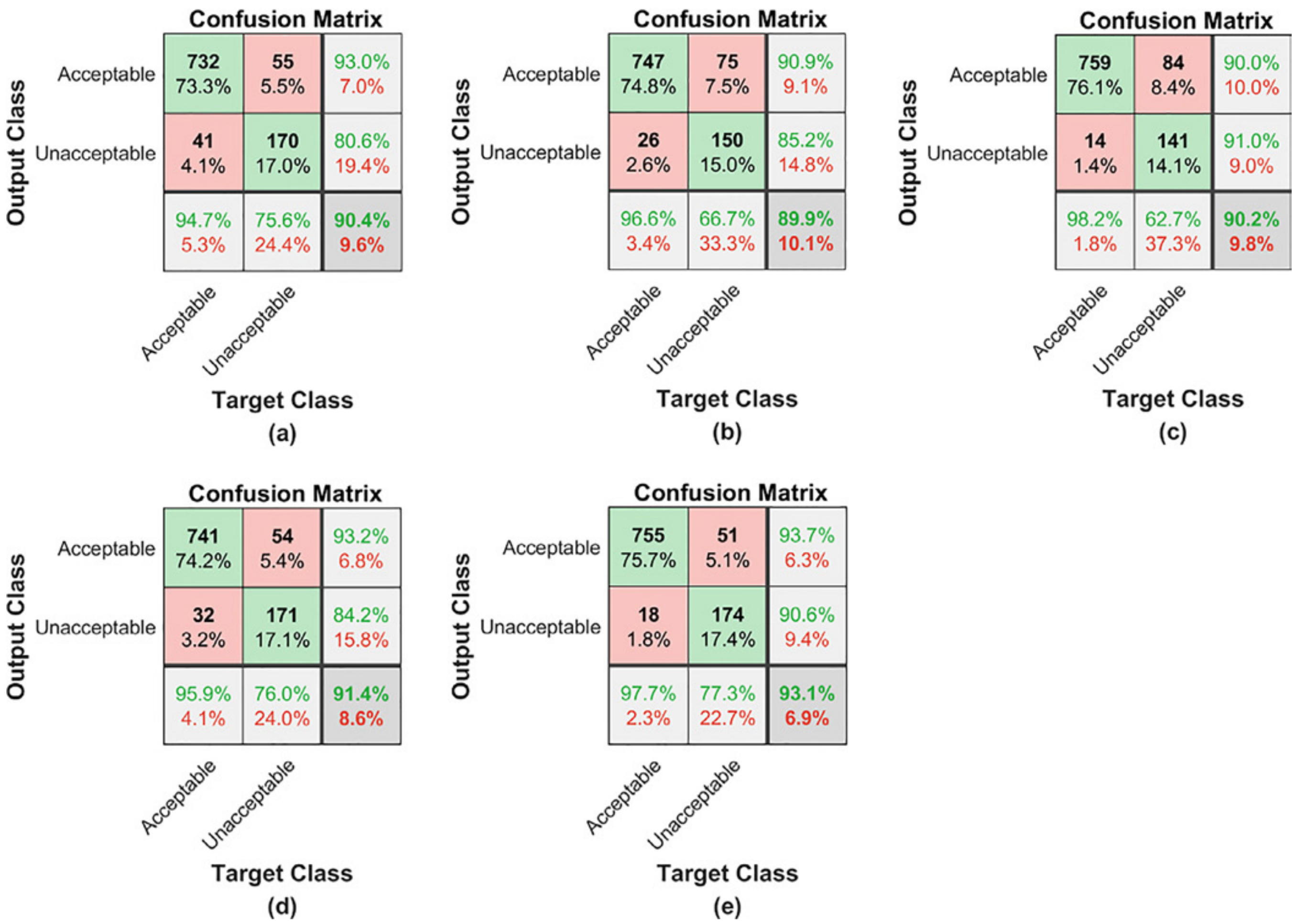


Fig. 7. (a)–(e) are the confusion matrices corresponding to experiment A–E.

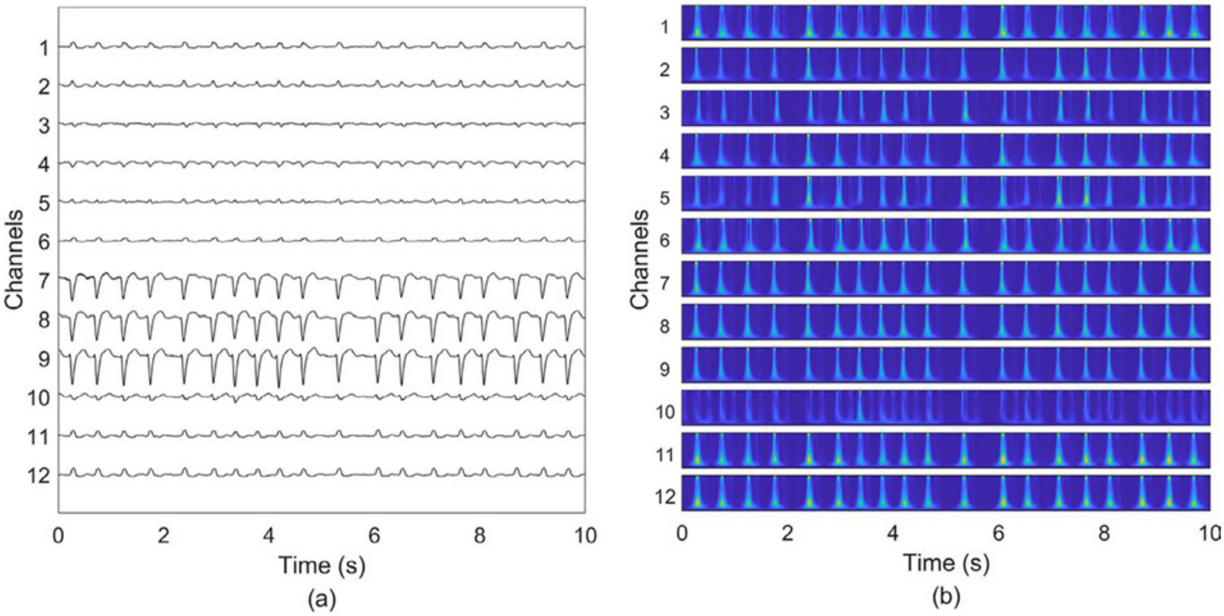


Fig. 8. (a) illustrates an example of ECG recording labeled 'Acceptable' and (b) is its corresponding S-Transform spectrogram.

ognize more ECG signals as acceptable signals, which has the advantage of not missing valuable signals in the subsequent processing stage. From this perspective, the features extracted by CNN automatically and statistical features are complementary. Comparing B/C and D/E experiments, it can be seen that the double-input CNN method, which combines the features extracted by ST-CNN automatically and statistical features, combines the advantages of the two features and has higher Sensitivity and Specificity. Simultaneously, the comprehensive evaluation indexes, including Precision, F1-score, and Accuracy, have been improved. In addition, comparing experiments D and E, the improvement of the generalization ability of the model is depicted. It can be seen that the method proposed in this paper based on augmented S-Transform convolutional neural network and statistical feature performed best overall, with Sensitivity, Specificity, Precision, F1 score, and Accuracy of 97.67%, 77.33%, 93.67%, 84.72%, and 93.09%, respectively. Compared with the S-Transform time-frequency representation method and the hand-crafted feature extraction method, this method improves the accuracy by 2.91% and 2.71%, respectively.

3.3. Case study

Fig. 8 illustrates an example tagged as an acceptable ECG signal in the Physionet/CinC Challenge 2011 database. For this multi-lead ECG segment, the hand-crafted feature extraction method will give a wrong classification label. In contrast, the augmented ST-CNN method gives the correct label (i.e., experiment A is classified incorrectly, but experiment B, C, D, and E are classified correctly). The statistical features concentrate on the maximum value and continuous identical value, so it is not surprising that the record is classified as unacceptable. In contrast, the record can be judged as an acceptable signal by augmented ST-CNN method because it has the ability to capture the inherent relative amplitude features of the signal.

The ECG signal in Fig. 9 is labeled as unacceptable. When only the ST-CNN is performed, the classification result will be wrong. In comparison, the classification result will be correct if statistical features are added for feature fusion (i.e., experiments B and C are misclassified, while experiments A, D, and E are correctly classified). It can be inferred that ST-CNN is not very sensitive to the lead-fall, and take more attention on the performance of whole 12 leads.

3.4. Interpretability study

Double-input CNN architecture has the capability of automatically fusing multiple features. Eq. (12) proves that given an output score, we can obtain the contribution of each type of feature by computing the log-odds. Further, the feasibility of separating the statistical feature contribution of each lead indicates we can spatially localize the abnormal ECG lead. On the other hand, the rich information contained in the S-Transform spectrogram promotes us to explore the possibility of precisely localizing the abnormal ECG recording in the time-frequency domain. To achieve this aim, we employed the Grad-CAM technique [33], which exploits the gradient of the 'Acceptable' softmax score with respect to the last max-pooling layer in the ST branch to find which parts of the multi-lead ECG dominate the output score. Fig. 10 illustrates an interpretability analysis of a sample with an 'Acceptable' label. We can observe from the raw ECG signal shown in Fig. 10(a) that there is an obvious baseline drift noise in the 8th lead at 3rd s. Meanwhile, the S-Transform spectrogram illustrated in Fig. 10 (c) demonstrates an unexpected noise with low frequency at the corresponding region. Fig. 10 (d) and (f) plot the importance map in the time domain and time-frequency domain. It can be seen that a lower Grad-CAM score is obtained in the low-frequency region at 3rd s. This proves that the Grad-CAM technique is able to precisely localize the beginning and end ranges of acceptable and non-acceptable zones. In addition, the contribution of each type of feature is depicted in Fig. 10(e), which shows that the deep-learned feature has higher confidence in labeling this sample as 'Acceptable' and dominates the final output score. Fig. 10(b) provides the spatial localization based on log-odds of each lead computed by statistical features, with results correctly localizing the lead suffered from noise. Fig. 11 shows another interpretability analysis example. It can be seen from Fig. 11(a) and (c) that a continuous noise that occurred in the 7th lead makes the sample have an unacceptable quality. The importance plots illustrated in Fig. 11(d) and (f) indicate two noises with high-frequency occurred at approximately 2nd and 3rd s have a significant impact on the final output result.

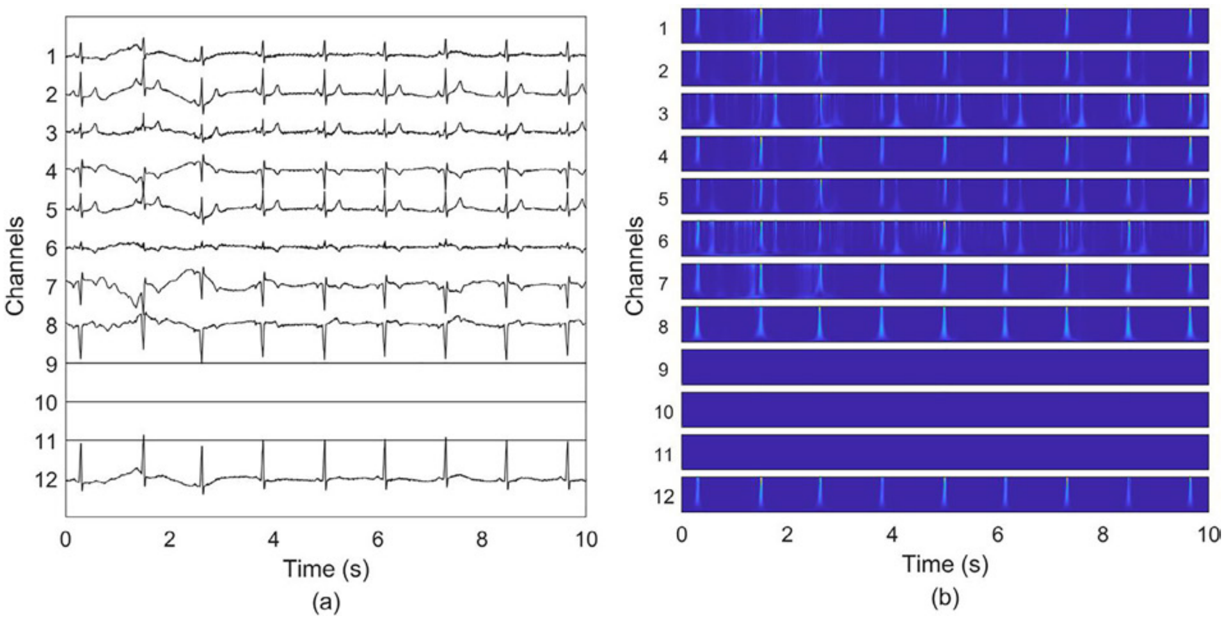


Fig. 9. (a) illustrates an example of ECG recording labeled ‘Unacceptable’ and (b) is its corresponding S-Transform spectrogram.

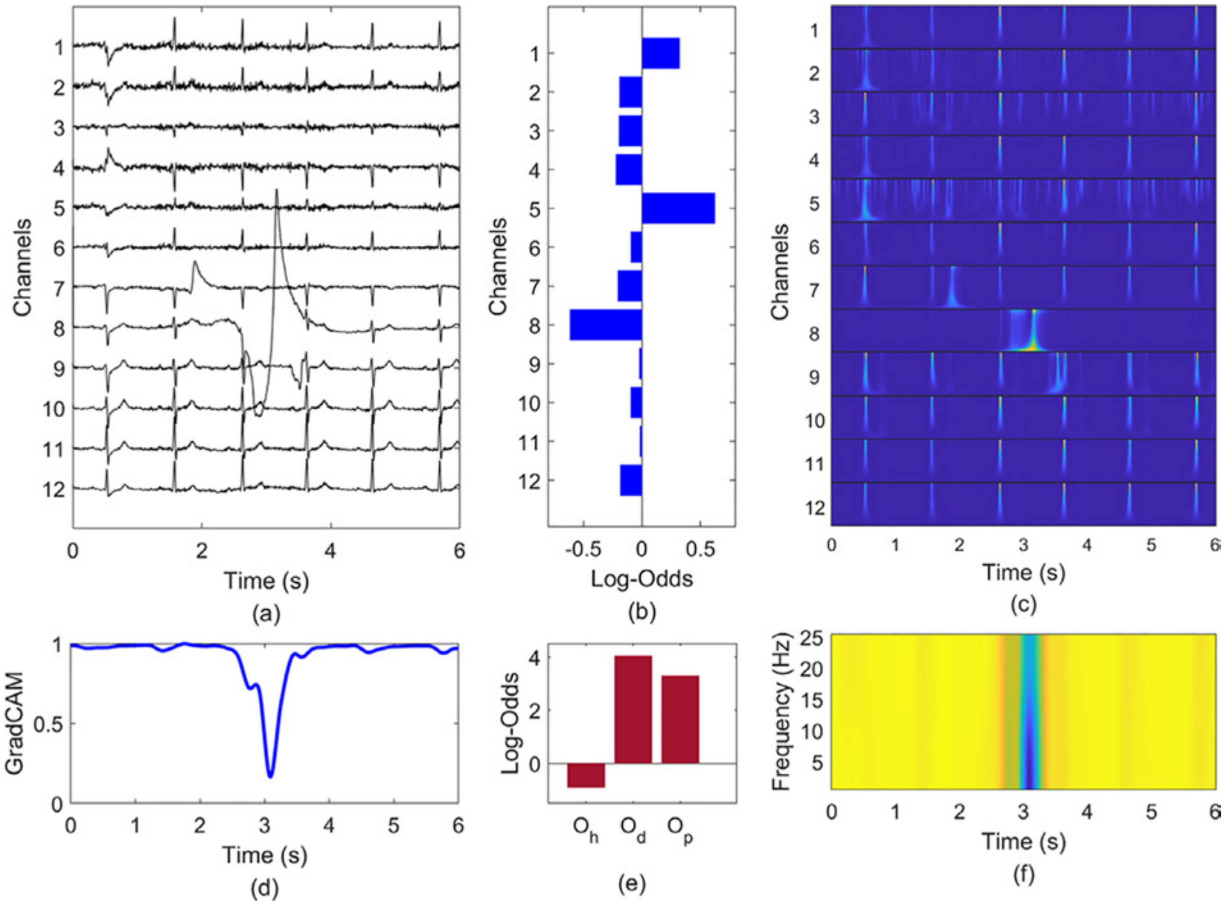


Fig. 10. An example of interpretability analysis. (a) and (c) are the raw multi-lead ECG sample labeled ‘Acceptable’ and its corresponding S-Transform spectrogram, respectively. (b) is the log-odds of each lead computed by corresponding statistical features. (f) is the importance map derived from the Grad-CAM technique, and (d) is its mean value calculated by squeezing the frequency dimension. (e) illustrates the log-odds that correspond to statistical feature, deep-learned feature, and output score.

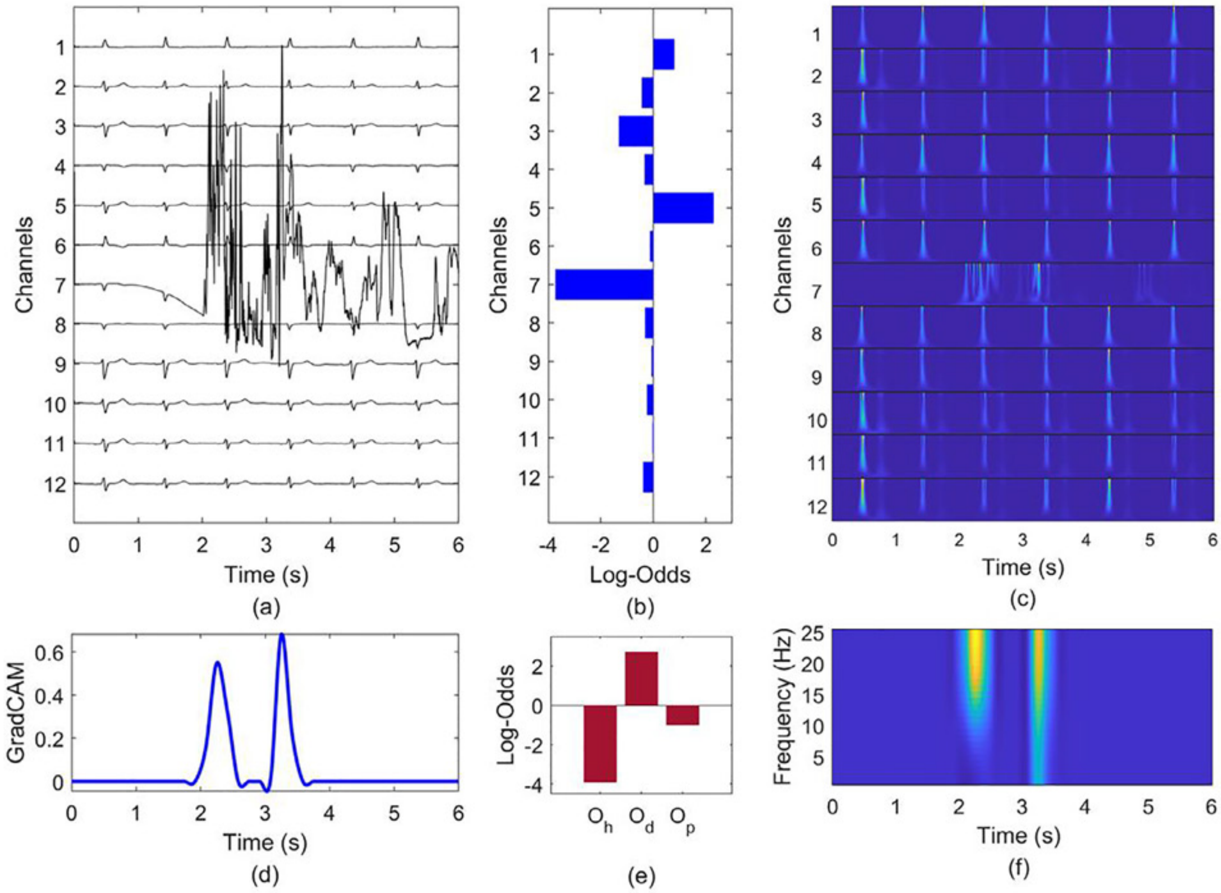


Fig. 11. An example of interpretability analysis. (a) and (c) are the raw multi-lead ECG sample labeled 'Unacceptable' and its corresponding S-Transform spectrogram, respectively. (b) is the log-odds of each lead computed by corresponding statistical features. (f) is the importance map derived from the Grad-CAM technique, and (d) is its mean value calculated by squeezing the frequency dimension. (e) illustrates the log-odds that correspond to statistical feature, deep-learned feature, and output score.

Table 2
Performance comparison with other methods.

Authors	Method	Evaluation	Sensitivity	Specificity	Accuracy
Liu et al. [34]	Integrative signal quality index (ISQI)	Rule-based	90.7%	89.8%	90.0%
Maan et al. [35]	Reconstruction Matrix	Rule-based	97.0%	75.1%	92.2%
Johannesen and Galeotti [13]	Two-step algorithm	Rule-based	95.0%	83.1%	92.3%
Hayn et al. [36]	Four measures	Rule-based	96.1%	84.0%	93.4%
Shahriari et al. [37]	Structural Image Similarity Metric	70% for training, 30% for testing	83.9%	77.7%	82.5%
Zhang et al. [23]	LSTM-ECG	Not mentioned	97.2%	81.2%	93.5%
Our work	Statistical features and S-transform CNN	10-fold cross- validation	97.7%	77.3%	93.1%

We can observe from Fig. 11(b) and (e) that the statistical feature dominates the final output score, and the corresponding most unacceptable lead is localized.

3.5. Performance comparison

The Physionet/CinC Challenge 2011 database employed in this paper is also used in many other works. Table 2 lists other methods assessed on Physionet/CinC Challenge 2011 database. Liu et al. [34] proposed an ISQI indicator that considers the straight line, huge impulse, Gaussian noise, and detector error, with results achieving an accuracy of 90% with a high specificity of 89.8%. Maan et al. [35] transformed the ECG to Vectorcardiogram and reconstruct it by inverse matrix, obtaining an accuracy of 92.2%. Johannesen and Galeotti [13] fulfilled an accuracy of 92.3% on set *A* with a two-step algorithm, which first rejects ECGs with macroscopic errors and subsequently quantifies the noise. Hayn et al. [36] introduced four quality measures, including empty lead and spike detection, number of lead crossing points, and QRS detection. Their method accomplished a higher accuracy of 93.4% but lower sensitivity of 96.1%. Shahriari et al. [37] proposed an image-based ECG quality assessment approach based on Structural Similarity Measure (SSIM), yielding a cross-validation accuracy of 82.5% on set *A*. Zhang et al. [23] designed an LSTM network for ECG quality assessment and reported a relatively high accuracy. However, their evaluation method was not mentioned. Therefore, the performance measures, even when higher, are not fully comparable. In this paper, a novel ECG quality assessment method is proposed, which innovatively combines the hand-crafted statistics and deep-learned S-Transform spectrogram features and obtains a mean accuracy of 93.09% with higher sensitivity of 97.7% in 10-fold cross-validation. Further, the proposed method has ideal interpretability and can realize abnormal ECG localization in time, frequency, and spatial domains.

4. Conclusion

The novelties of this paper can be concluded in the following aspects. We explore the performance of deep-learned S-Transform spectrogram features and introduce a novel online augmentation scheme for the first time, which can achieve higher sensitivity and generalization ability in evaluating ECG quality. Simultaneously, by fusing statistical features with deep-learned features using a double-input CNN architecture, we obtain a noticeable improvement in classification results. Through the case study, we also confirm that the two-type features are complementarity. Furthermore, the comprehensive interpretability analysis proves that our method can effectively localize the abnormal ECG recording in the time, frequency, and spatial domains.

Though the effectiveness of our proposed method has been proved, some limitations should be revealed. Firstly, the scale of the currently used database is relatively small. In the following study, more labeled data collected in challenging conditions should be further expanded to verify the generalization ability of the proposed method. Secondly, the structure of the adopted CNN and utilized statistical features is not fully optimized, and more novel network structures and statistical features should be evaluated to further improve the classification results in future work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The support of the Research Funds of Science and Technology Innovation Committee of Shenzhen Municipality (No. JCYJ20180305164357463), National Natural Science Foundation of China (No. 11474185), International Science and Technology Cooperation Project of Shandong Academy of Sciences (No. 2019GHZD04), Postgraduate Education Reform Fund of Shandong University (No. XYJG2020108), and Key R&D Project of Shandong Province (No. 2016GGX101028) are gratefully acknowledged.

References

[1] B.J. Maron, R.A. Friedman, P. Kligfield, B.D. Levine, S. Viskin, B.R. Chaitman, P.M. Okin, J.P. Saul, L. Salberg, G.F. Van Hare, Assessment of the 12-lead ECG as a screening test for detection of cardiovascular disease in healthy general populations of young people (12–25 years of age) a scientific statement from the American Heart Association and the American College of Cardiology, Circulation 130 (2014) 1303–1334.

[2] J. Jin, Screening for cardiovascular disease risk with ECG, JAMA 319 (2018) 2346.

[3] S.K. Berkaya, A.K. Uysal, E.S. Gunal, S. Ergin, S. Gunal, M.B. Gulmezoglu, A survey on ECG analysis, Biomed. Signal Process. Control 43 (2018) 216–235.

[4] C. Orphanidou, T. Bonnici, P. Charlton, D. Clifton, D. Vallance, L. Tarassenko, Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring, IEEE J. Biomed. Health Inform. 19 (2014) 832–838.

[5] B.J. Drew, P. Harris, J.K. Zègre-Hemsey, T. Mammone, D. Schindler, R. Salas-Boni, Y. Bai, A. Tinoco, Q. Ding, X. Hu, Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients, PLoS One 9 (2014) e110274.

[6] G.D. Clifford, F. Azuaje, P. Mcsharry, ECG statistics, noise, artifacts, and missing data, Adv. Methods Tools ECG Data Anal. 6 (2006) 18.

[7] J. Behar, J. Oster, Q. Li, G.D. Clifford, ECG signal quality during arrhythmia and its application to false alarm reduction, IEEE Trans. Biomed. Eng. 60 (2013) 1660–1666.

[8] Q. Li, G.D. Clifford, Signal quality and data fusion for false alarm reduction in the intensive care unit, J. Electrocardiol. 45 (2012) 596–603.

[9] C. Liu, X. Zhang, L. Zhao, F. Liu, X. Chen, Y. Yao, J. Li, Signal quality assessment and lightweight QRS detection for wearable ECG SmartVest system, IEEE Internet Things J. 6 (2018) 1363–1374.

[10] Q. Li, C. Rajagopalan, G.D. Clifford, A machine learning approach to multi--level ECG signal quality classification, Comput. Methods Programs Biomed. 117 (2014) 435–447.

[11] L. Johannesen, Assessment of ECG quality on an Android platform, in: Proceedings of the 2011 Computing in Cardiology, IEEE, 2011, pp. 433–436.

[12] G. Clifford, J. Behar, Q. Li, I. Rezek, Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms, Physiol. Meas. 33 (2012) 1419.

[13] L. Johannesen, L. Galeotti, Automatic ECG quality scoring methodology: mimicking human annotators, Physiol. Meas. 33 (2012) 1479.

[14] Y. Zhang, S. Wei, L. Zhang, C. Liu, Comparing the performance of random forest, SVM and their variants for ECG quality assessment combined with nonlinear features, J. Med. Biol. Eng. 39 (2019) 381–392.

[15] M. Yochum, C. Renaud, S. Jacquir, Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT, Biomed. Signal Process. Control 25 (2016) 46–52.

[16] M. Orini, R. Bailón, L.T. Mainardi, P. Laguna, P. Flandrin, Characterization of dynamic interactions between cardiovascular signals by time-frequency coherence, IEEE Trans. Biomed. Eng. 59 (2011) 663–673.

[17] F. Mochimaru, Y. Fujimoto, Y. Ishikawa, Detecting the fetal electrocardiogram by wavelet theory-based methods, Prog. Biomed. Res. 7 (2002) 185–193.

[18] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.

[19] D. Shen, G. Wu, H.I. Suk, Deep learning in medical image analysis, Annu. Rev. Biomed. Eng. 19 (2017) 221–248.

[20] Z. Zhao, C. Liu, Y. Li, Y. Li, J. Wang, B.S. Lin, J. Li, Noise rejection for wearable ECGs using modified frequency slice wavelet transform and convolutional neural networks, IEEE Access 7 (2019) 34060–34067.

[21] A. Huerta, A. Martínez-Rodrigo, V.B. González, A. Quesada, J. Rieta, R. Alcaraz, Quality assessment of very long-term ECG recordings using a convolutional neural network, in: Proceedings of the 2019 E-Health and Bioengineering Conference (EHB), IEEE, 2019, pp. 1–4.

[22] Q. Zhang, L. Fu, L. Gu, A cascaded convolutional neural network for assessing signal quality of dynamic ECG, Comput. Math. Methods Med. 2019 (2019) 7095137, doi:10.1155/2019/7095137.

[23] J. Zhang, L. Wang, W. Zhang, J. Yao, A signal quality assessment method for electrocardiography acquired by mobile device, in: Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2018, pp. 1–3.

[24] X. Zhou, X. Zhu, K. Nakamura, N. Mahito, ECG quality assessment using 1D–convolutional neural network, in: Proceedings of the 2018 14th IEEE International Conference on Signal Processing (ICSP), IEEE, 2018, pp. 780–784.

[25] L. Nanni, Y.M. Costa, D.R. Lucio, C.N. Silla, S. Brahnam, Combining visual and acoustic features for audio classification tasks, Pattern Recognit. Lett. 88 (2017) 49–56.

[26] Z. Farhoudi, S. Setayeshi, Fusion of deep learning features with mixture of brain emotional learning for audio-visual emotion recognition, Speech Commun. 127 (2021) 92–103.

[27] G. Kłosowski, T. Rymarczyk, D. Wójcik, S. Skowron, T. Cieplak, P. Adamkiewicz, The use of time-frequency moments as inputs of LSTM network for ECG signal classification, Electronics 9 (2020) 1452.

[28] R.G. Stockwell, L. Mansinha, R. Lowe, Localization of the complex spectrum: the S transform, IEEE Trans. Signal Process. 44 (1996) 998–1001.

[29] S. Ari, M.K. Das, A. Chacko, ECG signal enhancement using S-Transform, Comput. Biol. Med. 43 (2013) 649–660.

[30] Z. Zidelmal, A. Amirou, D. Ould-Abdeslam, A. Moukadem, A. Dieterlen, QRS detection using S-Transform and Shannon energy, Comput. Methods Programs Biomed. 116 (2014) 1–9.

[31] I. Silva, G.B. Moody, L. Celi, Improving the quality of ECGs collected using mobile phones: the physionet computing in cardiology challenge 2011, in: Proceedings of the 2011 Computing in Cardiology, IEEE, 2011, pp. 273–276.

[32] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint https://arxiv.org/pdf/1412.6980.pdf, 2014.

[33] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad–cam: visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.

[34] C. Liu, P. Li, L. Zhao, F. Liu, R. Wang, Real-time signal quality assessment for ECGs collected using mobile phones„ in: Proceedings of the Computing in Cardiology At: Hangzhou, IEEE, 2011, pp. 357–360.

[35] A.C. Maan, E.W. Van Zwet, S. Man, S.M. Oliveira-Martens, M.J. Schalij, C.A. Swenne, Assessment of signal quality and electrode placement in ECGs using a reconstruction matrix„ in: Proceedings of the Computing in Cardiology, IEEE, 2011, pp. 289–292.

[36] D. Hayn, B. Jammerbund, G. Schreier, QRS detection based ECG quality assessment, Physiol. Meas. 33 (2012) 1449.

[37] Y. Shahriari, R. Fidler, M.M. Pelter, Y. Bai, A. Villaroman, X. Hu, Electrocardiogram signal quality assessment based on structural image similarity metric, IEEE Trans. Biomed. Eng. 65 (2017) 745–753.