PAPER

A novel method to reduce false alarms in ECG diagnostic systems: capture and quantification of noisy signals

View the article online for updates and enhancements.

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

CrossMark

PAPER

A novel method to reduce false alarms in ECG diagnostic systems: capture and quantification of noisy signals

Wenliang Zhu[1,2], Lishen Qiu[1,2], Wenqiang Cai[3], Jie Yu[3], Deyin Li[3], Wanyue Li[3], Jun Zhong[2], Yan Wang[2] and Lirong Wang[2,3,*]

[1] School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China, People's Republic of China
[2] Suzhou Institute of Biomedical Engineering and Technology, China Academy of Sciences, People's Republic of China
[3] School of Electronics and Information Technology, Soochow University, People's Republic of China
[*] Author to whom any correspondence should be addressed.

E-mail: wanglirong@suda.edu.cn

Abstract

Objective. Muscle artifacts (MA) and electrode motion artifacts (EMA) in electrocardiogram (ECG) signals lead to a large number of false alarms from cardiac diagnostic systems. To reduce false alarms, it is necessary to improve the performance of the diagnostic algorithm in noisy environments by removing excessively noisy signals. However, existing methods focus on signal quality assessment and contain too many artificial features. Here, we present a novel method to flexibly eliminate noisy signals without any artificial features. *Approach.* Our method contains an improved lightweight deep neural network (DNN) to capture the signal portions damaged by EMA and MA, uses the sample entropy to quantize noisy portions, and discards those portions that exceed a defined threshold. Our method was tested in conjunction with Pan-Tompkins (PT), Filter Bank (FB), and 'UNSW' R-peak detection algorithms along with two heartbeat classification algorithms on datasets synthesized from the MIT-BIH Noise Stress Test Database, the China Physiological Signal Challenge 2018 Database and the MIT-BIH Arrhythmia Database. *Main results.* For PT R-peak detection algorithms, the sensitivity (Se) increased noticeably from 89.01% to 99.42% in the synthesized datasets with a signal-to-noise ratio of 6 dB. With the same datasets, the Se of the FB algorithm increased about 9.29%, and a 3.64% increase occurred in the Se of the 'UNSW' algorithm. For heartbeat classification algorithms, the overall F1-score increased about 6% in the synthesized one-heartbeat datasets. It is the first study to utilize a DNN to capture noisy segments of the ECG signal. *Significance.* Too many false alarms can cause alarm fatigue. Our method can be utilized as the preprocessing before signal analysis, thereby reducing false alarms from the ECG diagnostic systems.

1. Introduction

Electrocardiogram (ECG) signal analysis is an effective means for evaluating heart disease, but diagnosis of most cardiac arrhythmias is a challenge because of their brief, sudden, and sometimes asymptomatic nature (Fouassier *et al* 2020). Traditional cardiac data can only be collected when connected to a monitor, and intermittent heart conditions may be difficult to observe. Wearable ECG monitoring devices are a frequent solution (Martin *et al* 2000), allowing patients to be monitored during daily life instead of in a protected hospital environment. However, these devices increase exposure to noise (Moeyersons *et al* 2019). At present, automatic ECG signal diagnosis algorithms claiming to be clinically applicable are appearing, but serious signal interference still causes too many false alarms and 'alarm fatigue,' which desensitizes staff who come to ignore alerts from the system (Tsien and Fackler 1997). Initial preprocessing of ECG signals affects the accuracy of subsequent application algorithms. Preprocessing removes baseline shift, power-line interference, electrode motion artifacts (EMA), muscle artifacts (MA), and other interference (Kaplan Berkaya *et al* 2018, Serhani *et al* 2020). Baseline shift and power-line

interference are easily filtered out, but EMA and MA remain difficult. To reduce false alarms caused by these artifacts in a diagnosis system, there are usually two choices. One option is to filter the signal, but this method often degrades the signal and increases false alarms (Satija *et al* 2019). Another option is to evaluate the quality of the signal and discard the bad parts. Our method discards parts of the signal with too much EMA and MA noise, improving the performance of subsequent application algorithms, and reducing false alarms.

Many researchers have contributed to the evaluation of signal quality levels of ECG signals. Since the 2011 Computing in Cardiology (CinC) Challenge (Silva *et al* 2011), several signal quality indices (SQIs) (Behar *et al* 2012, Clifford *et al* 2012) have been used to evaluate ECG signals. These SQIs are:

a. bSQI: bSQI assesses the agreement level of two QRS detectors and is based on two open-source QRS algorithms: Pan-Tompkins (PT) algorithm (Pan and Tompkins 1985) and wqrs algorithm (Li *et al* 2008). The presence of noise in ECG signals can lower the agreement level between these two QRS algorithms.

b. tSQI: tQSI assesses the consistency of any two ECG beats in a fixed window.

c. iSQI: iSQI assesses the interval abnormal index for the R-R interval time series within a fixed window. A noisy ECG signal reduces the quality of QRS detection algorithms, with abnormal intervals appearing due to missing detections, leading to a low iSQI value.

d. aSQI: aSQI assesses high amplitude artifacts in ECG signals. The more times the amplitude changes, the smaller the value of aSQI.

e. pSQI: pSQI assesses the power spectrum distribution in a frequency range of 0.05–125 Hz and 0.05–45 Hz. Bad quality ECG signals will result in a larger pSQI value.

f. sSQI: sSQI is the third moment (skewness) of the ECG signal distribution.

g. kSQI: kSQI is the fourth moment (kurtosis) of the ECG signal distribution.

h. basSQI: basSQI assesses the power spectrum distribution in a frequency range of 1–40 Hz and 0–45 Hz.

Negin (Yaghmaie *et al* 2018) proposed the dynamic signal quality index (dSQI) that combines four SQIs (bSQI, kSQI, sSQI, pSQI) and employs a support vector machine (SVM) to classify ECG signals as clean normal and abnormal ECG signals and noisy ECG signals, with an accuracy of 96.9% and 96.3%, respectively. Zhao *et al* (Zhao and Zhang 2018) employed Cauchy, rectangular, and trapezoidal distributions to quantify the membership functions of several SQIs (qSQI, kSIQ, pSQI, and basSQI) and to establish a fuzzy vector. This method used a bounded operator for fuzzy synthesis and a weighted membership function for assessment and classification, achieving an accuracy rate of 94.67% in high-quality and low-quality dichotomy tasks. Zhang *et al* (2016) proposed using Lempel–Ziv complexity as an indicator for ECG signal quality. Subsequently, (Liu *et al* 2019) combined typical SQIs with sample entropy (SampEn), fuzzy measure entropy, and Lempel–Ziv complexity and used an SVM classifier to rate the signal quality among five levels, with excellent results. Satija *et al* (2018a) decomposed ECG signals using complete ensemble empirical mode decomposition to extract the characteristics of different noises from the inherent mode decomposition functions, localizing and classifying noises. Moeyersons *et al* (2019) separated the ECG signals into five second signal segments, extracted the autocorrelation function (ACF), extracted the features from the ACF, and finally classified the signal quality among five levels using a random under-sampling Adaboost (RUSBoost) classifier. In addition, Zhang *et al* (2019) converted the time–frequency spectrum of dynamic ECG signals into images of size 257×63 and used multiple cascading convolutional neural networks as classifiers to divide the ECG signals into five levels, with an accuracy of 92.7% against an open database. Finally, Satija *et al* (2018c) gave a detailed summary of relevant studies before 2017, which is an excellent starting point for readers desiring to learn more about ECG signal quality assessment.

Other researchers have identified and eliminated poor quality ECG signals, achieving good results. Mico *et al* (2010) achieved 97% sensitivity and a 16% false detection rate using the MIT-BIH database by filtering MA signals using a sliding window calculation method plus SampEn and its relative insensitivity to noise. Satija *et al* (2018b) used wavelets to decompose ECG signals, and extracted features of different frequency band coefficients after decomposition, and finally used multiple empirical thresholds to locate and classify noises, achieving good results. Bashar *et al* (2019) searched for noise characteristics in the time and frequency domains of ECG signals and more accurately identified poor quality electrocardiographic signals using an experience threshold, reducing the false-positive detection of atrial fibrillation (AF) by 94%.

Examining these works, we find that several problems remain.

First, existing SQIs are artificial features, so there may be the problem of feature redundancy or insufficiency. Moreover, the accuracy of some SQIs depends on the extraction algorithm. For example, the accuracy of pSQI depends on the R-peak detection algorithm, making the accuracy of ECG signal quality classification dependent on the extraction algorithm (Satija *et al* 2018c, Zhang *et al* 2019).

Second, existing studies base the levels of signal quality on the subjective judgment of annotators rather than the subsequent application algorithm, for example the R-peak detection algorithm or the heartbeat classification algorithm. Thus, the signal quality classification may not apply to some application algorithms.

Third, unclear boundaries between signal quality levels lead to many conflicting labels. Although some studies have formulated rules for labeling signal quality level, we found that conflicting labels still exist and are difficult to eliminate.

To address these issues, we propose a new method using an improved lightweight deep neural network to capture signal segments damaged by MA and EMA in ECG signals followed by a SampEn to quantify them. We also propose a simple rule for selecting the appropriate threshold according to the noise robustness of the subsequent application algorithm. In practical applications, when the SampEn value of a noisy part of a signal is greater than the threshold, the noisy part is discarded. Our approach has several advantages:

a. It is the first time a lightweight deep neural network (hereinafter referred to as 'the network') has been used to capture noisy segments in an ECG signal without artificial features.

b. We avoid categorizing the ECG signal quality into levels along with the subjectivity of annotators.

c. We allow the subsequent application algorithms to discard noisy segments, making our approach useful to a variety of applications and with greater flexibility.

We applied our proposed method for signals given to three excellent R-peak detection algorithms and two heartbeat classification algorithms. The results show that our proposed method significantly improves the performance of these algorithms in the presence of noisy signals.

2. Materials and methods

2.1. Datasets
In this section, we introduce the databases used in experiments, introduce the synthesized datasets used for training and testing the network of our method and how we synthesized them, and introduce the synthesized datasets used to validate the effectiveness of our method.

2.1.1. Introduction of databases
Three open-access databases used to synthesize a variety of datasets are presented here.

a. MIT-BIH noise stress test database (MIT-BIH NSTDB): This database includes twelve 30 min ECG records and three 30 min records of noise typical in ambulatory ECG records. Each record has two channels and is sampled at a frequency of 360 Hz. EMA and MA records are considered as the EMA and MA noise datasets used for this work. Each noise dataset is divided into two distinct parts. The first part accounts for 70% of the dataset and is used for mixing with clean ECG signals to train and test the network of our method. The second part accounts for 30% of the dataset and is used for mixing with other clean ECG signals to validate our method's effectiveness.

b. China Physiological Signal Challenge 2018 (CPSC 2018) Database: The ECG records in the CPSC 2018 database come from 11 hospitals and contain nine types of heart signals: one normal and eight abnormal types (9,831 records from 9458 patients with a time length of 7–60 min) (Liu *et al* 2018). These records were sampled at a frequency of 500 Hz. To unify the sampling rate with that of the MIT-BIH NSTDB, we re-sampled these signals at a frequency of 360 Hz.

c. MIT-BIH arrhythmia (MIT-BIH-AR) database: The MIT-BIH-AR database contains 48 two-channel ECG records of 47 patients (Moody and Mark 2002). Each record is 30 min long, and was sampled at the sampling frequency of 360 Hz. The heartbeat types of this database include normal beat (N), AF, ventricular ectopic beat (VEB), left bundle branch block beat (LBBB), right bundle branch block beat (RBBB), supraventricular ectopic beat (SVEB), fusion beats, and unknown beats.
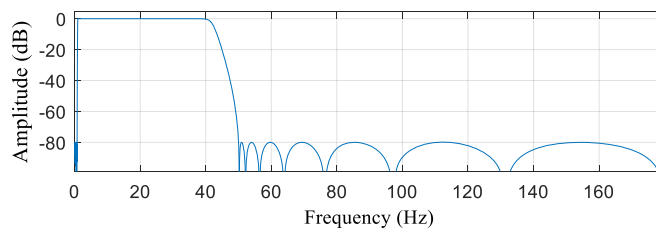
Figure 1. The amplitude–frequency response diagram of the bandpass filter. The filter is designed by MatLab fdatool, Chebyshev II. The bandpass frequency is in the range of 1–40 Hz.

Table 1. Types of synthesized datasets.

No.	Type	Filtered	SNR
1	ECG + MA	Unfiltered	—
2	ECG + MA	Filtered	—
3	ECG + EMA	Unfiltered	—
4	ECG + EMA	Filtered	—
5	ECG + EMA + MA	Unfiltered	—
6	ECG + EMA + MA	Filtered	—
7	ECG + EMA + MA	Filtered	6 dB
8	ECG + EMA + MA	Filtered	0 dB
9	ECG + EMA + MA	Filtered	−6 dB

In this table, ECG, MA, EMA, and SNR are abbreviations for electrocardiogram, muscle artifacts, electrode motion artifacts, and signal-to-noise ratio, respectively.

Table 2. Heartbeat types of clean ECG signals.

Class code	Full name	No. of pieces
N	Normal beat	307
AF	Atrial fibrillation beat	234
LBBB	Left bundle branch block beat	144
RBBB	Right bundle branch block beat	244
SVEB	Supraventricular ectopic beat	255
—	Another abnormal beat	195
Total	—	1379

2.1.2. Datasets for training and testing the network

We synthesized nine datasets to train and test the proposed network of our method. These datasets are mixtures of clean ECG signals and noise signals, as shown in table 1. In this table, datasets Nos. 2, 4, and Nos. 6–9 are filtered by a bandpass filter. The amplitude–frequency response diagram of the filter is shown in figure 1.

The clean ECG signals were clipped manually from the CPSC 2018 Database (Liu *et al* 2018). The heartbeat types in the clean ECG signals are shown in table 2 and include normal beat (N), AF, VEB, LBBB, RBBB, SVEB, and another abnormal beat (ST-segment depression and ST-segment elevated). There are 1379 pieces of clean ECG signals. The duration of each piece is 10 s. In addition, the noise signals come from 70% of EMA and MA noise datasets. To train and test the network, each dataset was divided into the train, validate, and test sets in a ratio of 7:2:1.

The mixing process is shown in figure 2. In this figure, (a) shows a clean ECG signal, (b) and (c) show, respectively, noise signals with the random length, and (d) shows a noise signal added to a clean ECG signal at a random location. In figure 2(e), the noisy part of the signal is labeled as 1, and the others are labeled as 0.

2.1.3. Datasets for validating the effectiveness

The effectiveness of our proposed method was verified using two situations. The first situation is that only an R-peak detector runs in the diagnostic system. The second situation is that an R-peak detector and a heartbeat classifier run in the diagnostic system.
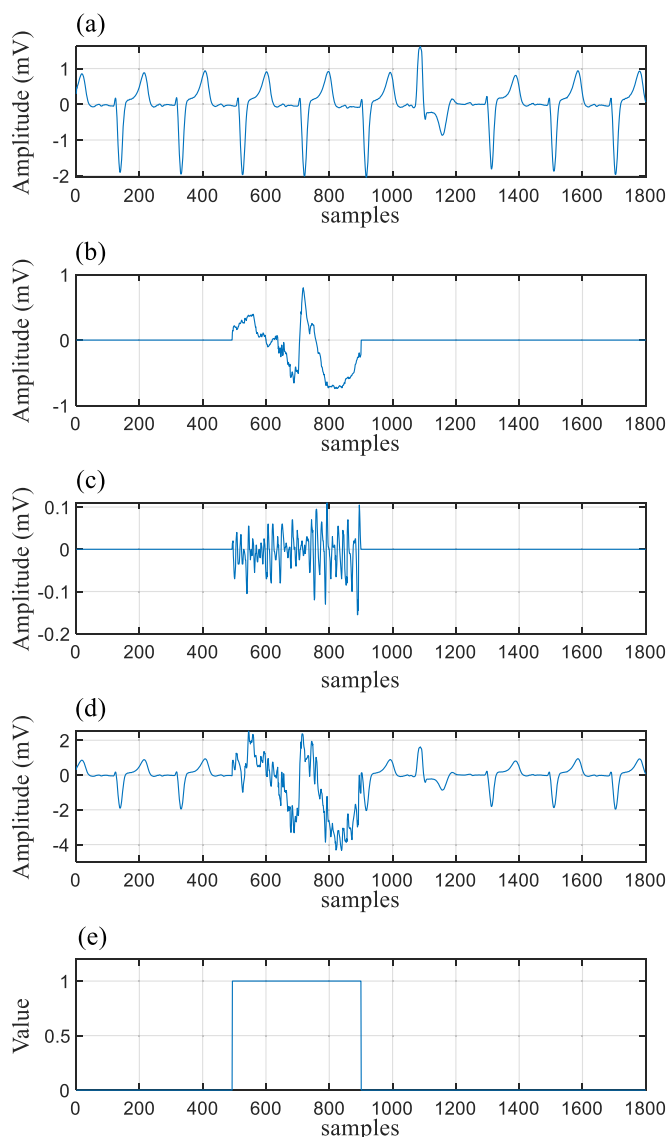
Figure 2. The process of mixing a clean ECG signal with noisy signals: (a) a clean ECG signal, (b) an electrode motion artifacts (EMA) signal, (c) an MA signal, (d) a mixed signal, and (e) label.
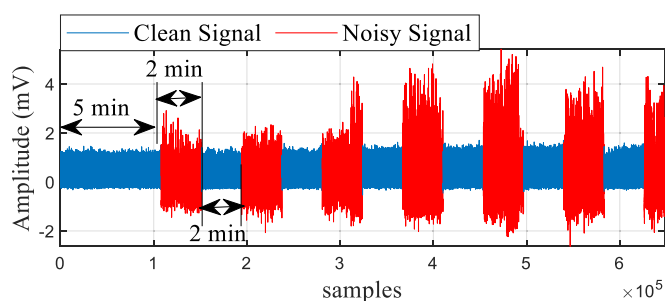
Figure 3. An example of a mixed signal. The blue area contains clean signals. The red area is polluted by noise. The SNR of this signal is −6 dB. The signal duration is 30 min. The noisy duration is 13 min.

2.1.3.1. Datasets for the first situation

In this situation, we separately utilize three R-peak detection algorithms to validate the effectiveness of our method in capturing noisy segments of a signal and set them to value 0, thereby reducing misidentifications in the R-peak detector.

Table 3. R-peak detection algorithms and clean ECG records.

No.	Algorithms	TS dataset		VD dataset	
		SN of clean records	No. of instances	SN of clean records	No. of instances
1	R-peak detection, PT (Pan and Tompkins 1985)	100, 102, 105, 106	4	109, 111, 112, 117	4
2	R-peak detection, FB (Afonso *et al* 1999)	100, 101, 102, 103, 104	5	105, 106, 107, 108, 109	5
3	R-peak detection, UNSW (Khamis *et al* 2016)	100, 101, 102,105,106	5	107, 109, 112, 114, 116	5

In this table, TS dataset is the threshold-search dataset, VD dataset is the validation dataset, and SN is the abbreviation of the serial number.

Table 4. Heartbeat classification algorithms and the number of instances.

No.	Algorithms	TS dataset No. of instances	VD dataset No. of instances
1	Heartbeat classification 1 (Kachuee *et al* 2018)	10803	10798
2	Heartbeat classification 2 (Mousavi and Afghah 2019)	22936	22931

In this table, TS dataset is the threshold-search dataset, VD dataset is the validation dataset.

We synthesize two noise-mixed datasets for each algorithm. The first dataset, called the threshold-search dataset (TS dataset), is used to select an appropriate threshold. The second dataset, called the validation dataset (VD dataset), is used to validate the effectiveness of our method with the threshold.

It is worth noting that to exclude the influence of heartbeat shapes or other factors which disrupt the result of experiments, the selected clean ECG signals must make the R-peak detection algorithm's sensitivity (Se) or positive predictivity (Pp) close to or equal to 100%. It is the basis for choosing clean ECG signals and the reason why different ECG recordings are applied for the different R-peak detection algorithms.

The clean ECG signals of the TS dataset and VD dataset were chosen from the MIT-BIH-AR database (Moody and Mark 2002). Each clean ECG signal is intact and 30 min long, its serial number (SN) and the number of instances are listed in the columns 'SN of clean records' and 'No. of instances', as shown in table 3. In addition, the noise signals come from 30% of the EMA and MA noise datasets.
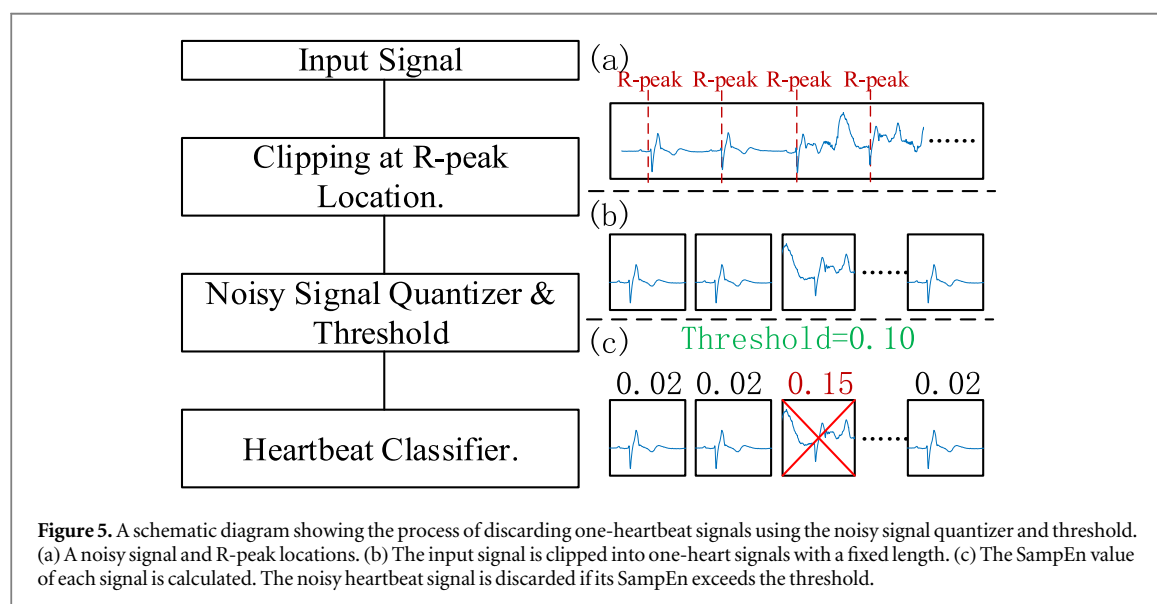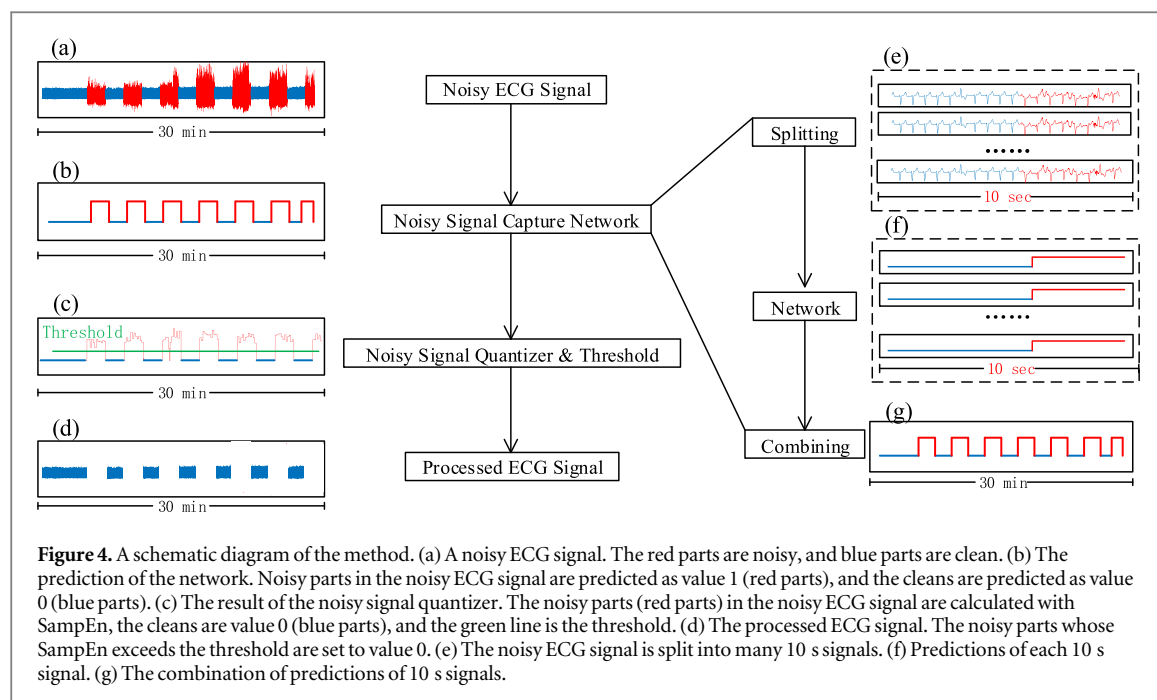
To demonstrate the effectiveness of our method, all the mixed ECG signals have a high SNR level of +6 dB. A high SNR level leads to blurring between clean and noisy ECG signals, increasing the difficulty of validation. The total duration of a mixed signal is 30 min. The duration of the noisy part of a mixed ECG signal is 13 min. Mixed signals were mixed following the first 5 min of each clean signal in 2 min noisy segments alternating with 2 min clean segments. To make it clearer, a −6 dB example of a mixed ECG signal is shown in figure 3. To preserve only MA and EMA, the mixed signals are filtered by a bandpass filter shown in figure 1.

2.1.3.2. Datasets for the second situation

In this situation, the one-heartbeat signals were clipped down from the input signal according to the results of the R-peak detector. We separately employ two heartbeat classification algorithms to validate the SampEn and a threshold has the ability to distinguish between clean and noisy one-heartbeat signals, thereby reducing the misclassifications in the heartbeat classifier.

We also synthesize two noise-mixed datasets for each algorithm, TS dataset, and VD dataset. In these two datasets, the clean ECG signals (one-heartbeat signals) were selected from test sets of the open-access datasets uploaded by the authors of the cited articles, and these signals must make the heartbeat classification algorithm's accuracy or F1-score equal 100%. The open-access datasets consist of a sequence of one-heartbeat signals which were already clipped down by their authors from the MIT-BIH-AR database. The duration of each clean ECG signal in the article of table 4 No.1 is 0.5 s and 0.78 s in the article of No.2. The total number of selected clean ECG signals is listed in the column 'No. of instances', as shown in table 4.

The noise signals come from 30% of EMA and MA noise databases. In the TS and VD datasets, only 50% of the heartbeat signals were mixed with both MA and EMA noise signals with the SNR of +6 dB; all the others are clean.

Figure 4. A schematic diagram of the method. (a) A noisy ECG signal. The red parts are noisy, and blue parts are clean. (b) The prediction of the network. Noisy parts in the noisy ECG signal are predicted as value 1 (red parts), and the cleans are predicted as value 0 (blue parts). (c) The result of the noisy signal quantizer. The noisy parts (red parts) in the noisy ECG signal are calculated with SampEn, the cleans are value 0 (blue parts), and the green line is the threshold. (d) The processed ECG signal. The noisy parts whose SampEn exceeds the threshold are set to value 0. (e) The noisy ECG signal is split into many 10 s signals. (f) Predictions of each 10 s signal. (g) The combination of predictions of 10 s signals.

Figure 5. A schematic diagram showing the process of discarding one-heartbeat signals using the noisy signal quantizer and threshold. (a) A noisy signal and R-peak locations. (b) The input signal is clipped into one-heart signals with a fixed length. (c) The SampEn value of each signal is calculated. The noisy heartbeat signal is discarded if its SampEn exceeds the threshold.

2.2. Method

In this section, we first introduce the principle of our method. Second, we introduce the important components in our method, including the noisy signal capture network, noisy signal quantizer, and the threshold. Third, we introduce the evaluation metrics, steps to validate the effectiveness of our method, and a suggested rule for finding an appropriate threshold. Lastly, we introduce the subsequent application algorithms that are employed to validate our method.

2.2.1. Principle of our method

Our method can be considered as the preprocessing for signals. Once the ECG signals passed through our method, the noisy parts of the signals were set to zero or were directly discarded. Thereby, false detections of subsequent application algorithms were reduced.

We considered two situations for our method:

Only an R-peak detector runs in a diagnostic system;

An R-peak detector and a heartbeat classifier run in a diagnostic system. Usually, the one-heartbeat signals to be classified are clipped down from the input long signal according to the result of the R-peak detector.

In the first situation, we set noisy parts of a noisy ECG signal to value 0, thereby reducing misidentifications in the R-peak detector. A schematic diagram of our method is shown in figure 4.

Figure 6. The structure of the noisy signal capture network. Four stages make up this network: (a) learning to down-sample; (b) global feature extractor; (c) feature fusion; (d) classifier. The structure of the DSconv1D, bottleneck is shown in figure 7.
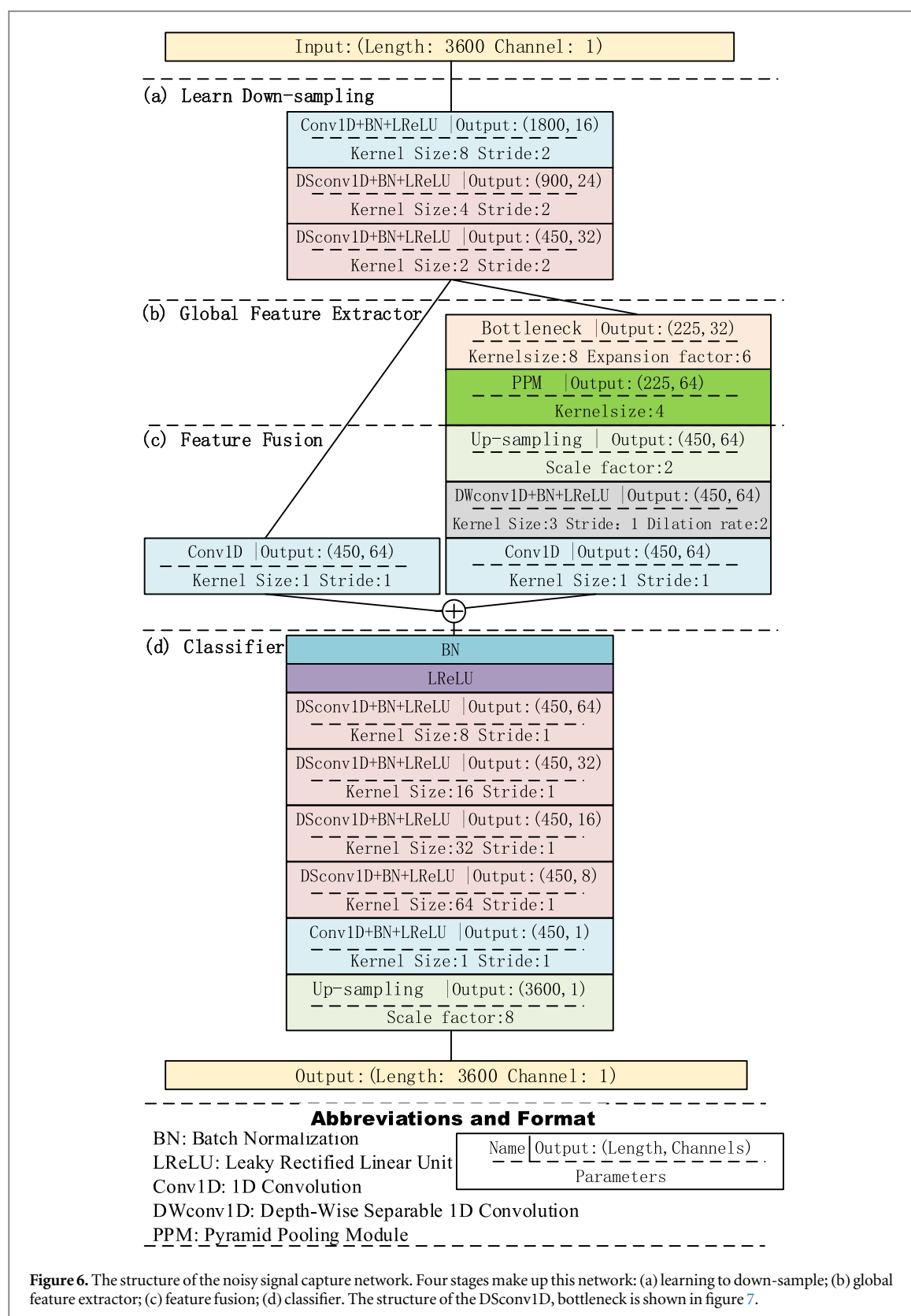
In figure 4(a), a 30 min noisy ECG signal is supposed to be the input for our method. In practice, the length of the input signal can be any length. Then, the noisy signal capture network outputs a prediction, as shown in figure 4(b). We can tell from the prediction which part of the signal is noisy. The noisy signal quantizer outputs the SampEn value of noisy parts and we can set these parts beyond the threshold to zero, as shown in figures 4(c) and (d).

The threshold in our method is the boundary line, which determines what kind of signal should be discarded. In other words, the threshold is the robustness of subsequent application algorithms. It is highly

(a) PPM

Input: (Length:225, Channel:32)

AvgPooling Output: (1, 32) Poolsize:225	AvgPooling Output: (3, 32) Poolsize:75	AvgPooling Output: (5, 32) Poolsize:45	AvgPooling Output: (7, 32) Poolsize:25
Conv1D Output: (1, 8) Kernel Size:1 Stride:1	Conv1D Output: (3, 8) Kernel Size:1 Stride:1	Conv1D Output: (5, 8) Kernel Size:1 Stride:1	Conv1D Output: (7, 8) Kernel Size:1 Stride:1
BN	BN	BN	BN
LReLU	LReLU	LReLU	LReLU
Up-sampling Output: (225,8) Scale factor:225	Up-sampling Output: (225,8) Scale factor:75	Up-sampling Output: (225,8) Scale factor:45	Up-sampling Output: (225,8) Scale factor:25

Concatenate | Output: (Length:225, Channel:32)

(b) Bottleneck

Input : (Length:450, Channel:32)

Conv1D+BN+LReLU | Output: (450, 32*6)
Kernel Size:1 Filter Size:32*6

DWconv1D+BN+LReLU | Output: (450, 32*6)
Kernel Size:3 Stride:1 Dilation rate:1

Conv1D | Output: (225, 32)
Kernel Size:1 Stride:2

Output: (Length:225, Channel:32)

(c) DSconv1D

Input: (Length:L, Channel:C)

DWconv1D Output: (L, C) Kernel Size: K Stride: S Dilation rate:1

BN

LReLU

Conv1D Output: (L, C) Kernel Size:1 Stride:1

BN

LReLU

Output: (Length:L, Channel:C)

Abbreviations and Format

BN: Batch Normalization
LReLU: Leaky Rectified Linear Unit
Conv1D: 1D Convolution
DWconv1D: Depth-Wise Separable 1D Convolution
PPM: Pyramid Pooling Module
L: Length C: Channels K: Kernel Size of Input S:Stride of Input

Name | Output: (Length, Channels)
Parameters

Figure 7. The structure of the PPM, bottleneck, and DSconv1D. The parameters of each layer in the PPM and bottleneck have been given. The parameters (L, C, K, S) of DSconv1D will be set according to the usage in the main structure.

recommended to look for a suitable threshold for the application algorithm before using the full method. If an unbefitting threshold applies to this method, the subsequent application algorithms may not perform well.

Due to the fact that the length of the network input is 10 s, an overlength input signal must be split into the signals in a length of 10 s, as shown in figure 4(e). The network processes a 10 s signal, outputs a prediction, and then recombines the predictions as the final output of the noisy signal capture network.

In the second situation, our method is first applied to the R-peak detector, and we assume that the R-peak detector has a good result. Then, the input signal is clipped into one-heart signals according to the result of the R-peak detector. The process of this situation is shown in fi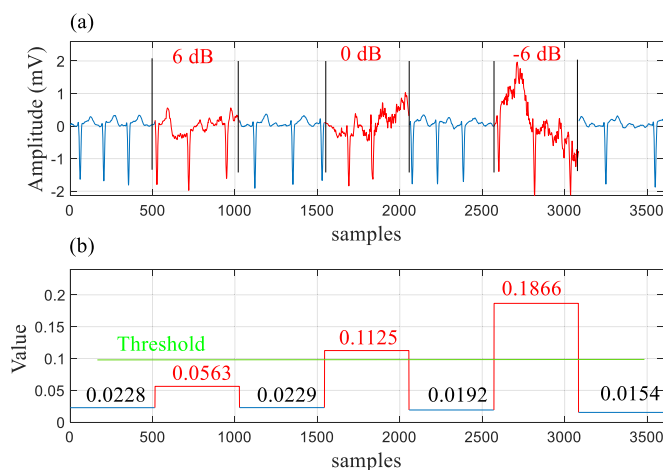gure 5. In this figure, (a) shows a noisy signal and its R-peak locations detected by the R-peak detector, and (b) shows a one-heartbeat signal clipped down from the input signal. Each one-heartbeat signal' SampEn value is calculated. If the one-heartbeat signal's SampEn value surpasses the threshold, it will be discarded, as shown in figure 5(c).

Figure 8. (a) The ECG signal with different SNRs. (b) The SampEn values of noisy parts in the ECG signal. The green line in (b) is the threshold. If the segment's SampEn value exceeds the threshold, the segment will be set to value 0.

2.2.2. Noisy signal capture network

We use the network to capture the noisy part of the signal for three reasons. First, this is the first time the deep learning network has been used to capture noisy parts in an ECG signal. Second, the ECG signal contains a combination of phenomena, including beats, conditions, and sequential patterns (Oh *et al* 2019), although the combinations are altered by EMA and MA. Nonetheless, the signal still has its unique features, which can be extracted by the network. Third, owing to the single scale of SampEn (Michael *et al* 2016), for a long ECG signal, the local value is more effective than the global value.

Figure 6 shows the network structure. The dual-branch network is a new structure proposed to reduce the number of computations required by the U-shaped network (Ronneberger *et al* 2015, Badrinarayanan *et al* 2019). It shows great potential in real-time applications and is better suited to low-memory embedded devices (Mazzini 2018, Poudel *et al* 2018, Yu *et al* 2018). In such a network, one path extracts semantics from deep multiple-scale modules, and the other path restores feature details. We redesign the dual-branch segmentation network (Poudel *et al* 2019) to tailor it to ECG signals, replace the classifier, and adjust many parameters.

Figure 6(a) illustrates the 'Learn Down-sampling' stage, wherein the length of the input is reduced. After many experiments, we determined that a three-layer structure at this stage yielded the best results. The first layer accepts an input ECG signal of 10 s (3600 sampling points), and applies 1D convolution (Conv1D) to augment the number of channels from 1 to 16. The second and third layers perform a 24-channel and a 32-channel DSConv1D process (Howard *et al* 2017) (shown in figure 7(c)), to extract characteristics of the shallow feature map. We perform batch normalization (Ioffe and Szegedy 2015) at the end of each layer to speed up the network optimization. We use a leaky rectified linear unit (Leaky ReLU) instead of an ReLU, as the latter loses the signal characteristics of the negative portion of the signals.

The second stage is global feature extraction, as shown in figure 6(b). We employ one bottleneck module (Sandler *et al* 2018) (shown in figure 7(b)) and one pyramid-pooling module (PPM) (Zhao *et al* 2017) (shown in figure 7(a)) in the network to extract global semantics from the signals. Our tests showed that adding more than one of the same modules did not improve the performance, so we think this combination is sufficient. The bottleneck module in this stage is used for extracting high-dimensional features efficiently, reducing the number of arguments and the computational load. The PPM works best for extracting global semantic information.

The third stage is the feature fusion module, as shown in figure 6(c). This stage fuses features from shallow layers with those from the deep layers and recovers missing detailed features.

In the fourth stage, the classifier, the network combines four DSConv1D modules with different kernel sizes, a 1D convolution (Conv1D), and an up-sampling, as shown in figure 6(d). We found that, compared to the model DSConv1D with the same kernel size, the model DSConv1D with different kernel sizes performed better.

The network uses a sigmoid activation function in the output layer, Adam as the optimization function, and binary cross-entropy as the loss function. The training batch size is 16 with 500 iterations. The code is developed on the Keras platform. The number of total parameters of the network is only 37625.

2.2.3. Noisy signal quantizer and threshold

We use SampEn as a noise quantizer to quantify signal damage for several reasons. First, SNR is the most direct method for quantifying signal damage, but it is difficult to estimate the energy of clean signals. Second, SampEn

is a nonlinear metric for estimating the regularity of a time series and finds similar patterns with the same length in the time series. The scarcer the frequency and possibility of these patterns, the higher the entropy value of the time series. Figure 8 shows the relationship between the SNR and SampEn value. As the SNR decreases, the entropy value increases, indicating that the SampEn value and noise power are closely related (Aboy *et al* 2007). Third, since the entropy value is almost unaffected by the clean signal energy (Mico *et al* 2010), the SampEn value of a signal without EMA and MA noise differs significantly from that of noisy signals.

For a discrete-time series of length *N*:

$$\{x(n)\} = x(1),\, x(2),...,x(N). \tag{1}$$

We calculate the SampEn value as follows. In this work, m is defined as the length of subsequences taken from the discrete-time series, and r is defined as the maximal distance between the maximal value in any two subsequences. Here, m and r are 2 and 0.25, respectively (Mico *et al* 2010).

1. Constitute a vector sequence with length m and N-$m + 1$ dimensions:

$$X_m(i) = \{x(i),\, x(i+1),...,x(i+m-1)\},$$
$$1 \leqslant i \leqslant N - m + 1. \tag{2}$$

2. We define the distance between vectors $X_m(i)$, $X_m(j)$ as the absolute value of the maximum difference between the corresponding elements:

$$d[X_m(i),\, X_m(j)] = \max_{k=0,...,m-1}(|x(i+k) - x(j+k)|)$$
$$1 \leqslant j \leqslant N - m,\, j \neq i. \tag{3}$$

3. For a given value $X_m(i)$, calculate the number of $d[X_m(i),\, X_m(j)] \leqslant r$, and define as B_i. For $1 \leqslant i \leqslant N - m$, we define $B_i^m(r)$ as:

$$B_i^m(r) = \frac{1}{N-m-1} B_i. \tag{4}$$

4. We define $B^m(r)$ as the probability of matching m points under the similarity tolerance r of two sequences:

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r). \tag{5}$$

5. We increase the length to $m+1$, calculate the value of $d[X_{m+1}(i),\, X_{m+1}(j)] \leqslant r$, calling it $A_i(r)$. We defined $A_i^m(r)$ as:

$$A_i^m(r) = \frac{1}{N-m-1} A_i(r). \tag{6}$$

6. Define $A^m(r)$ as the two sequences matching the probability of the $m + 1$ point:

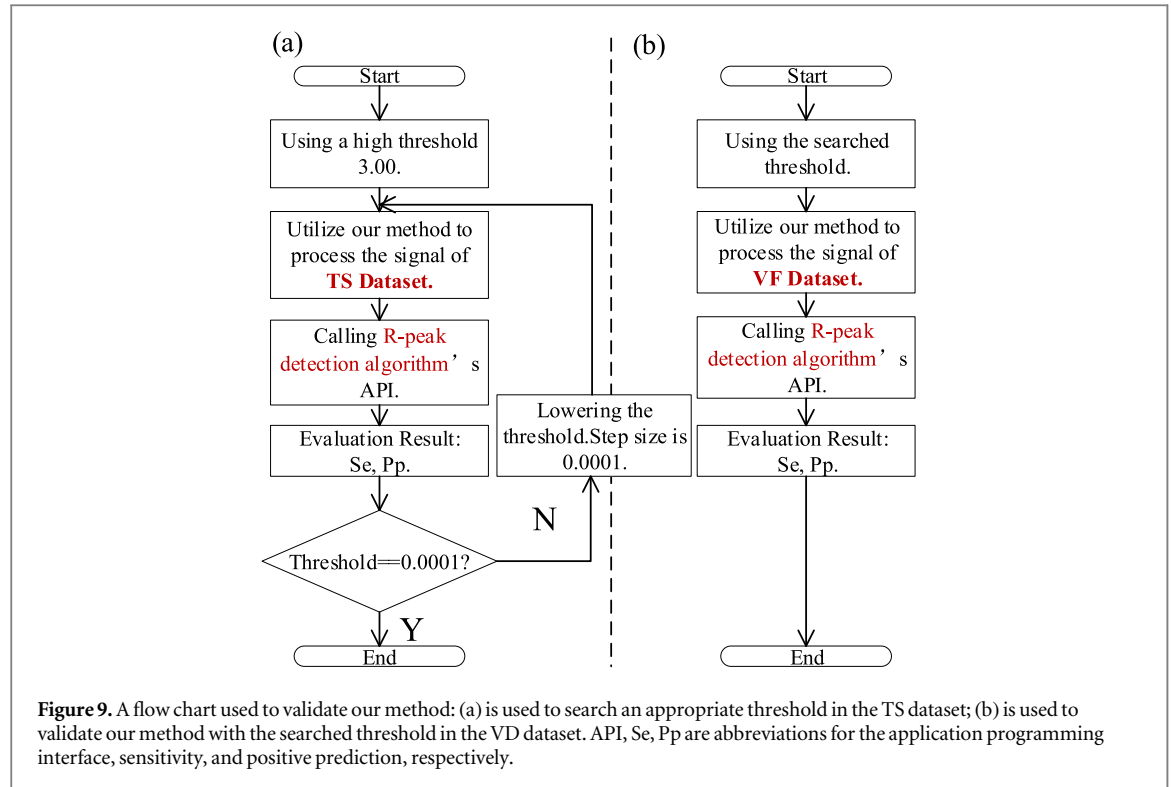$$A^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r). \tag{7}$$

7. We define the SampEn as:

$$SampEn(m,\, r) = \lim_{N \to \infty} \left[-\ln \frac{A^m(r)}{B^m(r)} \right]. \tag{8}$$

If N is a finite value, the definition becomes:

$$SampEn(m,\, r) = -\ln \frac{A^m(r)}{B^m(r)}. \tag{9}$$

The threshold is an important part of our method, and easy to use, as shown in the green line in figure 8(b). The flexibility of our method is reflected in this threshold, since it can be defined with a different value for

Figure 9. A flow chart used to validate our method: (a) is used to search an appropriate threshold in the TS dataset; (b) is used to validate our method with the searched threshold in the VD dataset. API, Se, Pp are abbreviations for the application programming interface, sensitivity, and positive prediction, respectively.

different subsequent application algorithms, for example, different r-peak detection algorithms and heartbeat detection algorithms.

2.2.4. Steps, evaluation metrics and a rule

2.2.4.1. Validation of the network via two accuracy metrics

The noisy signal capture network is an important part of our method, and we firstly validate it using the datasets in table 1. The evaluation metrics are defined as follows:

Point-wise accuracy (PAcc):

$$\mathrm{P}Acc = \frac{1}{m}\sum_{i=1}^{m} \frac{\sum_{j=1}^{N}(x_{i,j})}{N}, \; x_{i,j} = \begin{cases} 1, \; P_{i,j} = T_{i,j} \\ 0, \; P_{i,j} \neq T_{i,j} \end{cases} \tag{10}$$

where m is the batch size of the test set, N is the length of a prediction of the network, and $P_{i,j}$ denotes the j value in the i prediction of the network. Here, $T_{i,j}$ denotes the j value in the i label.

This metric is used to evaluate the mean proportion of the correctly classified number of samples in all signals of the test set.

Mean of intersection over union (mIoU):

$$mIoU = \frac{1}{m}\sum_{i=1}^{m} \frac{\sum_{j=1}^{N}(x_{i,j})}{\sum_{j=1}^{N}(x_{i,j}) + \sum_{j=1}^{N}(y_{i,j}) + \sum_{j=1}^{N}(z_{i,j})},$$
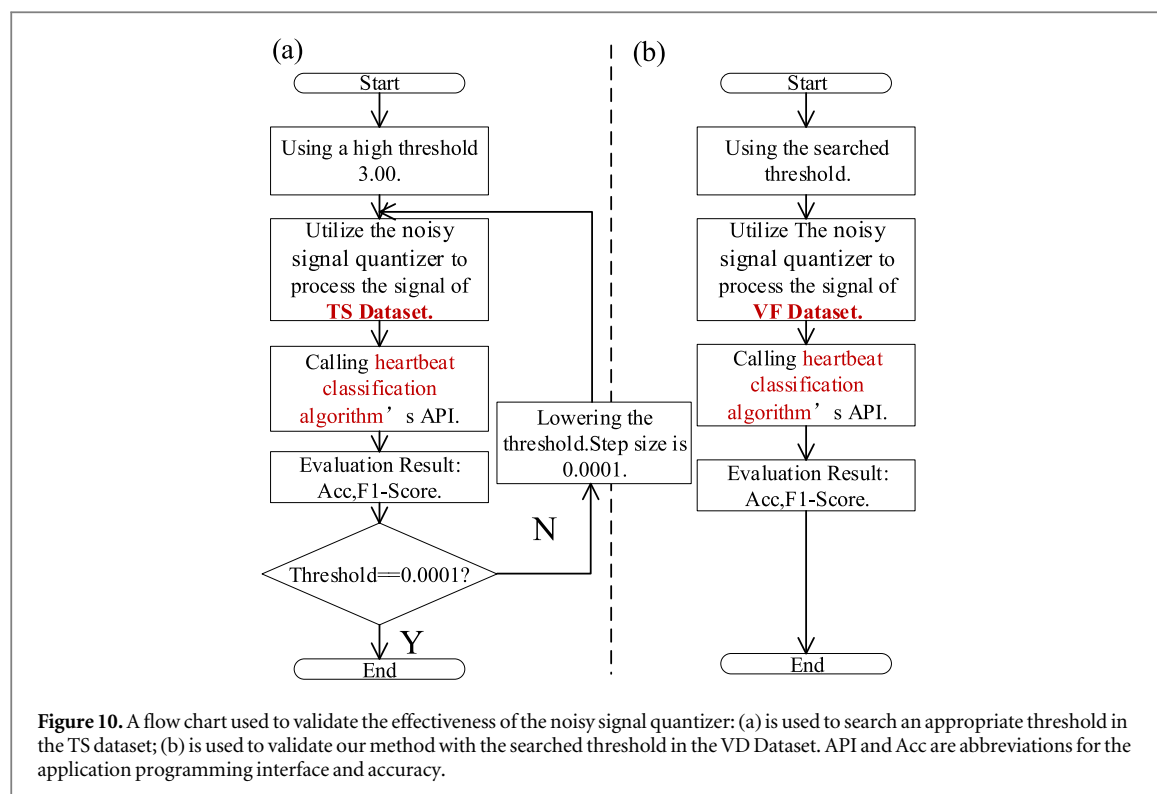
$$x_{i,j} = \begin{cases} 1, \; P_{i,j} = 1, \; T_{i,j} = 1 \\ 0, \; others \end{cases},$$

$$y_{i,j} = \begin{cases} 1, \; P_{i,j} = 1, \; T_{i,j} = 0 \\ 0, \; others \end{cases},$$

$$z_{i,j} = \begin{cases} 1, \; P_{i,j} = 0, \; T_{i,j} = 0 \\ 0, \; others \end{cases} \tag{11}$$

where m is the batch size of the test set, N is the length of a prediction of the network, and $P_{i,j}$ denotes the j value in the i prediction of the network. Here, $T_{i,j}$ denotes the j value in the i label.

This metric is used to evaluate the matching confidence between true areas and predicted areas in a signal.

Figure 10. A flow chart used to validate the effectiveness of the noisy signal quantizer: (a) is used to search an appropriate threshold in the TS dataset; (b) is used to validate our method with the searched threshold in the VD Dataset. API and Acc are abbreviations for the application programming interface and accuracy.

2.2.4.2. Validation of our method via the sensitivity and positive prediction of R-peak detection algorithms
This is the first situation in which we considered that only an R-peak detector works in the diagnosis system. We must obtain appropriate thresholds for each algorithm before the validation. They are different because they are related to the noise robustness of the algorithm. Figure 9(a) shows the flow chart used to search a threshold. At the beginning of this flow chart, we give an initial threshold of 3.0 for our method; it is a fairly high value. Then, we apply this threshold to our method to process signals in the TS dataset and call the application programming interface (API) of the R-peak detection algorithm with the processed signals. We record the evaluation result, Se, Pp, and the average discarding length (the length of a signal which is set to zero). Then, we again lower the threshold with the step size of 0.0001. We repeat these steps until the threshold equals 0.0001. In the end, a reference curve is obtained. The appropriate threshold is selected from the curve. Figure 9(b) is the flow chart used to validate our method with the appropriate threshold in the VD dataset.

The evaluation metrics, Se, and Pp are defined as follows:

$$Se = \frac{TP}{TP + FN} \tag{12}$$

$$Pp = \frac{TP}{TP + FP} \tag{13}$$

where *TP*, *FP*, *FN*, and *TN* are true-positive numbers, false-positive numbers, false-negative numbers, and true-negative numbers, respectively.

2.2.4.3. Validation of our method via the accuracy and F1-score of heartbeat classification algorithms
We validate our method in the second situation, where an R-peak detector and a heartbeat classifier work in the diagnostic system. Our whole method can be used for the R-peak detector just like the first situation. Due to the fixed length of the one-heartbeat signals, it is sufficient to apply a noisy signal quantizer and the threshold to discard the noisy one-heartbeat signal.

We focus on validating the ability of the noisy signal quantizer and the threshold. As in the first situation, we must select an appropriate threshold from the TS dataset, using the flow chart shown in figure 10(a).
Figure 10(b) shows the flow chart for validating the noisy signal quantizer using the searched threshold in the VD dataset.

The evaluation metrics of accuracy (Acc) and the F1-score are defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

$$F1 - score = \frac{2 \times TP}{2 \times TP + FP + FN}. \tag{15}$$

For a certain class in a multi-classification problem, *TP*, *FP*, *FN*, and *TN* are true-positive numbers, false-positive numbers, false-negative numbers, and true-negative numbers, respectively. The averages of two metrics among classes were calculated to give a final evaluation.

2.2.4.4. A rule for finding the appropriate threshold
The appropriate threshold is not absolute, but it is appropriate if the discarding length and the improved performance are acceptable.

In this work, there is a suggested rule for finding an appropriate threshold. The rule is:

a. To find a threshold range in the reference curve in which the average discarding length (or discarding number) of noisy signals is less than the real noisy length of each signal;

b. In this range, find a maximal value of the metric that we focus on, and the corresponding threshold is the appropriate one.

2.2.5. Subsequent application algorithms
Three R-peak detection algorithms and two heartbeat classification algorithms were employed as the subsequent application algorithms in this work to synergistically validate the effectiveness of our method. Due to the different code versions and development environments, only five algorithms can work well on our computers. These algorithms are all well documented, open-sourced, and uploaded by their authors.

a. The PT R-peak detection algorithm: this algorithm is one of the most used. It has an excellent performance with an accuracy of 99.3% in the MIT-BIH-AR database (Pan and Tompkins 1985).

b. The filter bank (FB) R-peak detection algorithm: this algorithm enables independent time and frequency analysis to be performed on the ECG signal, and the feature computed from a set of the sub-bands and a heuristic detection strategy are used to fuse decisions from multiple one-channel beat detection algorithms. The overall beat detection algorithm has an Se of 99.59% and a Pp of 99.56% in the MIT-BIH-AR database (Afonso *et al* 1999).

c. The UNSW R-peak detection algorithm: this algorithm generates a feature signal containing information about ECG amplitude and derivative, which is filtered according to its frequency content and an adaptive threshold is applied. It has a Se of 99.76% and a Pp of 99.80% in the MIT-BIH-AR database (Khamis *et al* 2016).

d. The first heartbeat classification algorithm: this algorithm uses a deep convolution neural network to classify heartbeats into one of five different arrhythmias (N, SVEB, VEB, F, Q) (Kachuee *et al* 2018). This algorithm makes predictions with an average accuracy of 93.4% in the MIT-BIH arrhythmia database.

e. The second heartbeat classification algorithm: this algorithm develops a sequence to sequence a deep learning approach to classify heartbeats into one of three different arrhythmias (N, SVEB, VEB) (Mousavi and Afghah 2019). This algorithm makes predictions with average accuracies of 99.53% in the MIT-BIH arrhythmia database.

3. Results

The method for eliminating noisy segments in ECG signals has been proposed. In this work, two situations for our method were considered. The first situation is where an R-detector runs in a diagnostic system. The second situation is where an R-peak detector and a one-heartbeat signal classifier collaboratively work in a diagnostic system.

In this section, we first introduce the result of the network and show processed signals. Second, we introduce the results of our method applied to the first situation. Third, we introduce the results of our method applied to the second situation.

3.1. Results of the noise capture network
The overall behavior of the noise capture network identifies the noisy segments of ECG signals. We used the point-wise accuracy (PAcc) and the mean of the class-wise intersection over union (mIoU) to evaluate the performance, with the results shown in table 5.
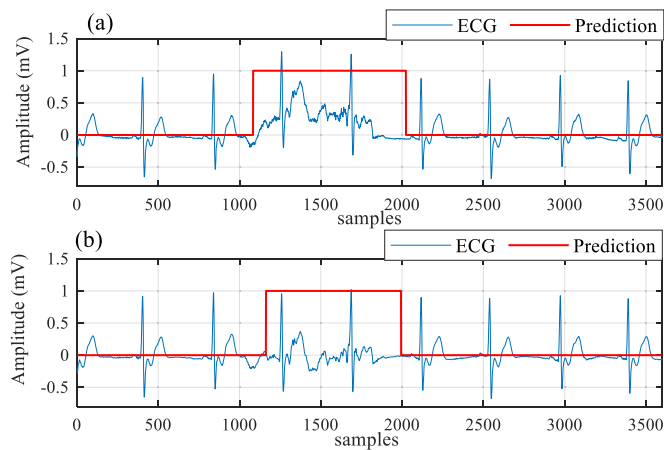
Figure 11. An example of noisy ECG signals and predictions. (a) The ECG signal is destroyed by the EMA and is unfiltered. (b) The signal is destroyed by the EMA and is filtered by a passband filter. The red lines in these two figures are the prediction of the network.
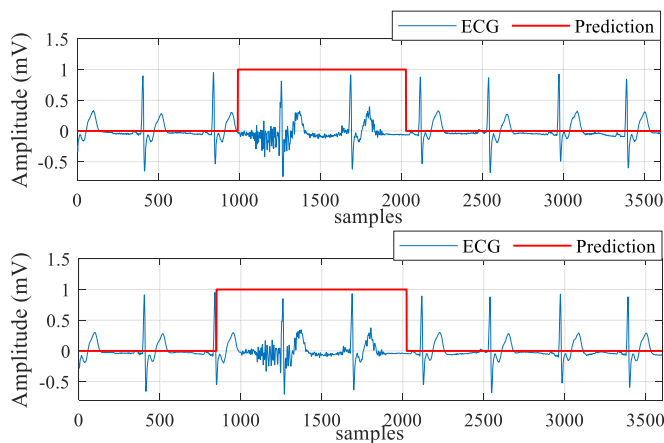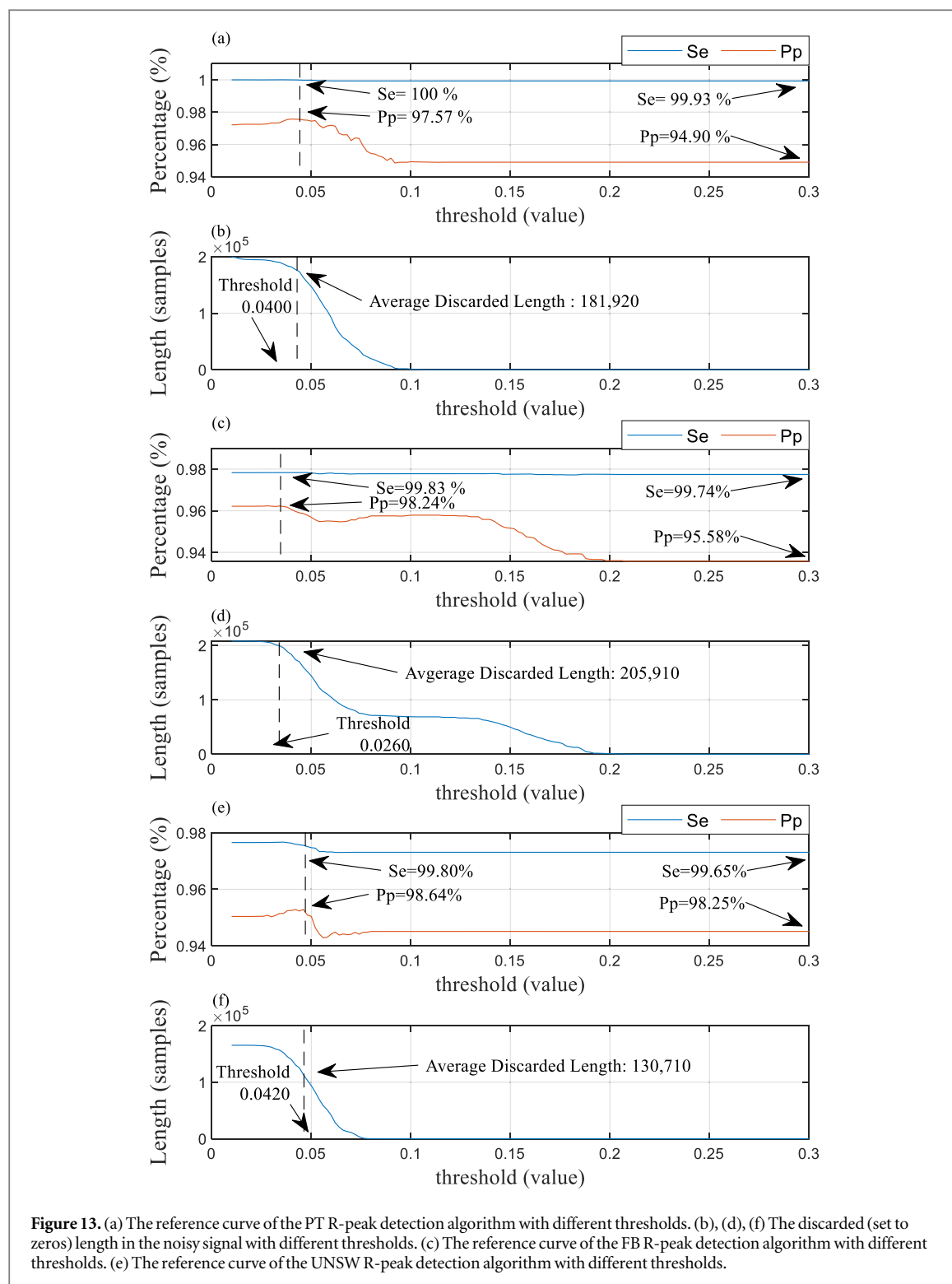

Figure 12. An example of noisy ECG signals and predictions. (a) The signal is destroyed by the MA and is unfiltered. (b) The signal is noised by the MA and is filtered by a passband filter.

Table 5. Performance of the network for various datasets.

Index	Type	Filtered	PAcc (%)	mIoU (%)
1	EMA + ECG	Unfiltered	96.99	90.41
2		Filtered	96.13	85.34
3	MA + ECG	Unfiltered	97.04	89.21
4		Filtered	97.34	90.70
5	EMA + MA + ECG	Unfiltered	98.42	95.01
6		Filtered	98.45	94.98
7	EMA + MA + ECG, +6 dB	Unfiltered	96.55	91.04
8	EMA + MA + ECG, 0 dB	Unfiltered	98.47	95.46
9	EMA + MA + ECG, −6 dB	Unfiltered	98.65	96.14
Average:	—	—	97.56	92.03

In this table, ECG, EMA, and MA are abbreviations for electrocardiogram, electrode motion artifacts, and muscle artifacts. The evaluation metrics PAcc and mIoU are abbreviations for point-wise accuracy and mean of intersection over union.

Among these signal types (ECG + EMA, ECG + MA, and ECG + EMA + MA), the network has the best performance in the signal type ECG + MA + EMA, with a small difference in the performance between filtered and unfiltered signals, as shown in table 5: No. 5 and No. 6.

Figure 13. (a) The reference curve of the PT R-peak detection algorithm with different thresholds. (b), (d), (f) The discarded (set to zeros) length in the noisy signal with different thresholds. (c) The reference curve of the FB R-peak detection algorithm with different thresholds. (e) The reference curve of the UNSW R-peak detection algorithm with different thresholds.

In the ECG + EMA datasets, the point accuracy performance of the filtered signal was slightly lower than the unfiltered one. We hold the opinion that EMA noise is wide-band noise containing many low-frequency components (Tam and Webster 1977), and the bandpass filter weakens low-frequency components and features of the noisy signal. This can be clearly seen in figure 11.

Owing to the spectral characteristics of MA, the waveform did not change significantly before and after filtering, as shown in figure 12. This example in figure 12 was randomly selected from the results of the network. The earlier part of the prediction in figure 12(b) is an incorrect network output, and the prediction in figure 12(a) is completely correct. This indicates that the network has room for improvement. However, there are only a few such examples. As a whole, there is only a minor difference between the results, as shown in table 5: Nos. 3 and 4.

The network has excellent results in these quantitative datasets, as shown in table 5: Nos. 7, 8, and 9.
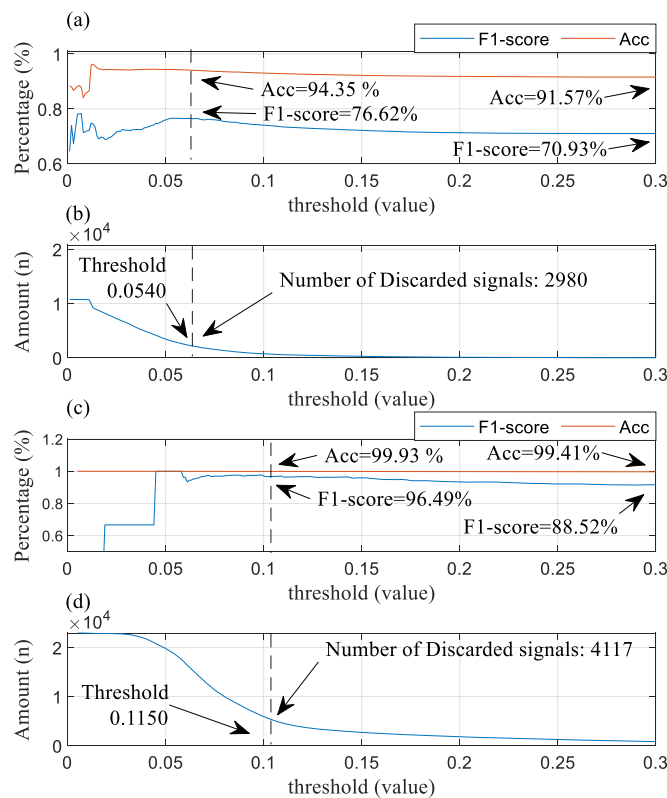
Figure 14. (a) The reference curve of the first heartbeat classification algorithm with different thresholds. (b), (d) The curve of the quantity of discarded signals with different thresholds. (c) The reference curve of the second heartbeat classification algorithm with different thresholds. Acc is the accuracy.

Table 6. Results of the VD datasets.

Algorithm	Without our method		With our method			
	Se (%)	Pp (%)	Se (%)	Pp (%)	Threshold	Discarded length (minutes)
Pan-Tompkins	92.77	89.01	99.90	99.42	0.0400	12.47
Filter bank	99.21	90.24	99.89	99.53	0.0260	12.13
UNSW	99.05	95.98	1	99.62	0.0420	12.70

The discarded length is the average length of noisy signals that are set to zeros in the VD dataset.

3.2. Results of R-peak detection algorithms

In this situation, only an R-peak detector runs in the diagnostic system; the system is used for alarming abnormal heart rate or heart rate variability analysis. EMA and MA cause R-peak identification errors, which in turn cause false alarms in the diagnostic system. Our method was validated in conjunction with every R-peak detection algorithm.

The experiment was completed following the flow chart in figure 9(a), and the reference curve used to select a threshold is shown in figure 13. The network was trained with the dataset ECG + EMA + MA (table 1, No. 6). For the R-peak detection algorithms, we consider Pp in particular because the Se is high enough.

The reference curve of the PT R-peak detection algorithm is shown in figures 13(a) and (b). With the descent of the threshold, the Se and Pp are increasing. The Pp decreases after the threshold close to 0.05, which may be due to the strategy of the R-peak detection algorithm and the misidentification of the network. In the range of 0.01 to 3.0, the average discarded length is less than the real noisy signal length of 13 min. When the threshold descends to 0.0400, the average discarding length is only 181920 points, about 8.42 min, less than the real noisy length of 13 min, and the Pp is maximal; we consider the threshold is appropriate, and it is 0.0400. The reference curves of the FB and UNSW R-peak detection algorithms are shown in figures 13(c)–(f). According to the reference curves, the FB algorithm has its threshold of 0.0260 and the UNWS algorithm's threshold is 0.0420.

We apply the threshold to our method and validate using the VD datasets in table 3, and the results are shown in table 6. In this table, the performance of three R-peak detection algorithms has an obvious promotion.

Table 7. Results of the VD datasets.

Algorithm	Without our method		With our method			
	F1-score (%)	Acc (%)	F1-score (%)	Acc (%)	Threshold	Discarded amount (n)
The first algorithm	70.90	91.74	76.09	94.29	0.0540	2885
The second algorithm	92.43	99.63	99.07	99.97	0.1150	3132

In this table, Acc is the accuracy. The discarded amount is the number of discarded one-heart signals in the VD dataset.

For the PT algorithm, Se increases to 99.90%, with a maximum increase of 7%, and Pp increases to 99.42%, with an increase over 10%. For the FB algorithm, the Pp increases from 90.24% to 99.53%. For the UNSW algorithm, the Pp increases to 99.62% with the maximum average discarding length of 12.70 min.

3.3. Results of heartbeat classification algorithms

In this situation, an R-peak detector and a heartbeat classifier synergistically work in the diagnostic system. When classifying heartbeats, EMA and MA cause misclassification, and false alarms were happening in the system. Our method was validated in conjunction with two classification algorithms, and the reference curves are shown in figure 14.

For the first heartbeat classification algorithm, the reference curve is shown in figures 14(a) and (b), At the location of the threshold 0.0540, the number of discarded one-heartbeat signals is only 2980, less than 50% of the TS dataset (about 5401), And we only focus on the F1-score due to the high-enough accuracy. With the threshold of 0.0540, the F1-score increases to its first maximal value of 76.62%. We consider the threshold is desirable.

For the second heartbeat classification algorithm, the reference curve is shown in figures 14(c) and (d). In the threshold area where the number of discarded one-heartbeat signals is less than 50% of the TS dataset (about 11468), we obtain the maximal F1-score of 96.49%, the discarding number is 4117, and the threshold is 0.1150.

We apply the threshold to our method and the results are shown in table 7. The first algorithm's F1-score increases from 70.90% to 76.09%. The second algorithm's F1-score has an increase of about 6% and reaches 99.07%.

4. Discussion

To reduce the false alarms from diagnosis systems caused by EMA and MA, we proposed a new method requiring no artificial features and flexible enough to use with several R-peak detection algorithms and heartbeat classification algorithms. Experimental results show that the proposed method demonstrates a good performance, with several details meriting some discussion.

It is feasible to use deep neural networks to capture parts of the signals damaged by MA and EMA. Since the time–frequency domain characteristics of these two kinds of noises are quite similar to the characteristics of ECG signals, it is difficult for ordinary filtering methods to filter out these two kinds of noise. Many researchers have proposed methods to filter out these two kinds of noise, with others thinking that deleting the noisiest parts of the signals is more straightforward. We developed our hypothesis that ECG signals containing noise retain characteristics that can be extracted by a deep neural network alongside clean signals. To validate the performance of the network, we synthesized a variety of signals, including signals of occasional perturbations in static state, signals with different SNRs in a state of motion, and signals with hardware filters. The experimental results show that our method effectively captures the noisy parts of each synthetic dataset. We found that the worse the signal quality, the better the performance of the network.

We also applied our method to three R-peak detection algorithms. The results show the method effectively captures the signal components damaged by EMA and MA and discards the noisy parts with SampEn values higher than a threshold. This method improved the performance of all three R-peak detection algorithms in noisy environments. However, if the network performs poorly, it will wrongly identify parts of clean signals while ignoring noisy parts, resulting in the incorrect deletion of clean signals. In this case, the performance of R-peak detection algorithms will not improve, and too many signals will be deleted as a result. In other words, the performance of an R-peak detection algorithm can be improved only if the network has the ability to capture noisy signal portions accurately. Our experiments with three R-peak detection algorithms show the feasibility of our approach for capturing the noisy parts of ECG signals.

Since the 2011 CinC ECG signal quality assessment challenge, ECG signal quality assessment has become a significant topic of interest. There are many excellent studies, and many of them classify ECG signals according to quality levels; however, we think the greatest issue with these efforts is the subjective classification of signal

quality by labelers rather than from an objective algorithm. Further, this subjectivity often leads to excessive filtering. In contrast, our automated solution produced experimental results showing it to be effective for three R-peak detection and two single-beat classification algorithms. This indicates that our method is more flexible in its implementation and application.

There are some limitations to our approach. First, our method only applies to signals containing only EMA and MA noise. If there is a large baseline drift or power frequency interference in the signal, the performance of our proposed method will be affected. Because the time required to calculate the SampEn value is related to the signal complexity, a long computation delay will occur if the signal is long and complex. To solve this problem, we suggest expanding the types of training data and using a better noisy signal quantizer. We plan to explore this in future research. Additionally, the small amount of training data produces some false predictions with the test dataset.

5. Conclusion

There are many clinical health reasons to monitor ECG signals, for example, heart rate variability analysis and heartbeat diagnosis. An R-peak detector and a heartbeat classifier are indispensable parts of a diagnostic system. But ECG signals are often disturbed by EMA and MA, causing false alarms from the diagnostic system. To reduce the false alarms in the diagnostic systems, it is important to eliminate noisy segments of ECG signals. The proposed novel approach is based on a lightweight deep neural network to capture the parts of ECG signals damaged by EMA and MA, a SampEn used to quantify the damaged part, and a decision threshold to discard damaged parts. In this work, the effectiveness of this proposed method has been validated in synthetic datasets. To the best of our knowledge, our work is the first to propose a deep learning network to identify the noisy parts of an ECG signal. Our future objective is to expand this work by incorporating more quantizing methods to increase the robustness and identify more types of noise.

Acknowledgments

ORCID iDs

Lirong Wang ● https://orcid.org/0000-0001-9172-3746

References

Aboy M, Cuesta-Frau D, Austin D and Mico-Tormos P 2007 Characterization of sample entropy in the context of biomedical signal analysis *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2007** 5943–6
Afonso V X, Tompkins W J, Nguyen T Q and Luo S 1999 ECG beat detection using filter banks *IEEE Trans. Biomed. Eng.* **46** 192–202
Badrinarayanan V, Kendall A and Cipolla R 2019 SegNet: a deep convolutional encoder-decoder architecture for image segmentation *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 2481–95
Bashar S K, Ding E, Walkey A J, McManus D D and Chon K H 2019 Noise detection in electrocardiogram signals for intensive care unit patients *IEEE Access* **7** 88357–68
Behar J, Oster J, Li Q and Clifford G D 2012 A single channel ECG quality metric *Computing in Cardiology* pp 381–4
Clifford G D, Behar J, Li Q and Rezek I 2012 Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms *Physiol. Meas.* **33** 1419–33
Fouassier D, Roy X, Blanchard A and Hulot J S 2020 Assessment of signal quality measured with a smart 12-lead ECG acquisition T-shirt *Ann. Noninvasive Electrocardiol.* **25** e12682
Howard A G, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M and Adam H 2017 MobileNets: efficient convolutional neural networks for mobile vision applications arXiv:1704.04861
Ioffe S and Szegedy C 2015 Batch normalization: accelerating deep network training by reducing internal covariate shift arXiv:1502.03167
Kachuee M, Fazeli S and Sarrafzadeh M 2018 ECG heartbeat classification: a deep transferable representation *2018 IEEE Int. Conf. on Healthcare Informatics (ICHI)* pp 443–4
Kaplan Berkaya S, Uysal A K, Sora Gunal E, Ergin S, Gunal S and Gulmezoglu M B 2018 A survey on ECG analysis *Biomed. Signal Process. Control* **43** 216–35
Khamis H, Weiss R, Xie Y, Chang C-W, Lovell N H and Redmond S J 2016 QRS detection algorithm for telehealth electrocardiogram recordings *IEEE Trans. Biomed. Eng.* **63** 1377–88
Li Q, Mark R G and Clifford G D 2008 Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter *Physiol. Meas.* **29** 15
Liu C Y, Zhang X Y, Zhao L N, Liu F F, Chen X W, Yao Y J and Li J Q 2019 Signal quality assessment and lightweight QRS detection for wearable ECG SmartVest system *IEEE Internet of Things J.* **6** 1363–74
Liu F F *et al* 2018 An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection *J. Medical Imaging Health Inform.* **8** 1368–73

Martin T, Jovanov E and Raskovic D 2000 Issues in wearable computing for medical monitoring applications: a case study of a wearable ECG monitoring device (https://doi.org/10.1109/ISWC.2000.888463)

Mazzini D 2018 Guided upsampling network for real-time semantic segmentation *British Machine Vision Conf.* arXiv:1812.07421

Michael B A, Richard, Van E A and Emmerik 2016 Multiscale entropy: a tool for understanding the complexity of postural control *J. Sport Health Ence.* **5** 44–51

Mico P, Mora M, Cuesta-Frau D and Aboy M 2010 Automatic segmentation of long-term ECG signals corrupted with broadband noise based on sample entropy *Comput. Methods Programs Biomed.* **98** 118–29

Moeyersons J *et al* 2019 Artefact detection and quality assessment of ambulatory ECG signals *Comput. Methods Programs Biomed.* **182** 105050

Moody G B and Mark R G 2002 The impact of the MIT-BIH arrhythmia database *IEEE Eng. Med. Biol. Mag.* **20** 45–50

Mousavi S and Afghah F 2019 Inter- and intra-patient ECG heartbeat classification for arrhythmia detection: a sequence to sequence deep learning approach *ICASSP 2019–2019 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp 1308–12 arXiv:1812.07421

Oh S L, Ng E Y K, Tan R S and Acharya U R 2019 Automated beat-wise arrhythmia diagnosis using modified U-net on extended electrocardiographic recordings with heterogeneous arrhythmia types *Comput. Biol. Med.* **105** 92–101

Pan J and Tompkins W J 1985 A real-time QRS detection algorithm *IEEE Trans. Biomed. Eng.* **32** 230–6

Poudel R P K, Bonde U, Liwicki S and Zach C 2018 ContextNet: exploring context and detail for semantic segmentation in real-time *British Machine Vision Conf.* arXiv:1805.04554

Poudel R P K, Liwicki S and Cipolla R 2019 Fast-SCNN: fast semantic segmentation network arXiv:1902.04502

Ronneberger O, Fischer P and Brox T 2015 U-Net: convolutional networks for biomedical image segmentation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* arXiv:1505.04597

Sandler M, Howard A, Zhu M, Zhmoginov A and Chen L-C 2018 MobileNetV2: inverted residuals and linear bottlenecks *2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 4510–20 arXiv:1801.04381

Satija U, Ramkumar B and Manikandan M S 2018a Automated ECG noise detection and classification system for unsupervised healthcare monitoring *IEEE J. Biomed. Health Inform.* **22** 722–32

Satija U, Ramkumar B and Manikandan M S 2018b An automated ECG signal quality assessment method for unsupervised diagnostic systems *Biocybernetics Biomed. Eng.* **38** 54–70

Satija U, Ramkumar B and Manikandan M S 2018c A review of signal processing techniques for electrocardiogram signal quality assessment *IEEE Rev. Biomed. Eng.* **11** 36–52

Satija U, Ramkumar B and Manikandan M S 2019 A new automated signal quality-aware ECG beat classification method for unsupervised ECG diagnosis environments *IEEE Sensors J.* **19** 277–86

Serhani M A, Kassabi H T E, Ismail H and Nujum Navaz A 2020 ECG monitoring systems: review, architecture, processes, and key challenges *Sensors (Basel)* **20** 1796

Silva I, Moody G B and Celi L 2011 Improving the quality of ECGs collected using mobile phones: the PhysioNet/Computing in cardiology challenge 2011 *2011 Comput. Cardiol.* **38** 273–6

Tam H W and Webster J G 1977 Minimizing electrode motion artifact by skin abrasion *IEEE Trans. Biomed. Eng.* **24** 134–9

Tsien C L and Fackler J C 1997 Poor prognosis for existing monitors in the intensive care unit *Crit. Care Med.* **25** 614–9

Yaghmaie N, Maddah-Ali M A, Jelinek H F and Mazrbanrad F 2018 Dynamic signal quality index for electrocardiograms *Physiol. Meas.* **39** 105008

Yu C, Wang J, Peng C, Gao C, Yu G and Sang N 2018 BiSeNet: bilateral segmentation network for Real-time semantic segmentation *European Conf. on Computer Vision* pp 325–41 arXiv:1808.00897

Zhang Q, Fu L and Gu L 2019 A cascaded convolutional neural network for assessing signal quality of dynamic ECG *Comput. Math. Methods Med.* **2019** 7095137

Zhang Y, Wei S, Di Maria C and Liu C 2016 Using Lempel-Ziv complexity to assess ECG signal quality *J. Med. Biol. Eng.* **36** 625–34

Zhao H, Shi J, Qi X, Wang X and Jia J 2017 Pyramid scene parsing network *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 6230–9 arXiv:1612.01105

Zhao Z and Zhang Y 2018 SQI quality evaluation mechanism of single-lead ECG signal based on simple heuristic fusion and fuzzy comprehensive evaluation *Front Physiol* **9** 727