



Comparison of Machine Learning Algorithms for the Quality Assessment of Wearable ECG Signals Via Lenovo H3 Devices

Fan Fu¹ · Wentao Xiang¹ · Yukun An¹ · Bin Liu¹ · Xianqing Chen³ · Songsheng Zhu^{1,2} · Jianqing Li^{1,2} 

Received: 25 September 2020 / Accepted: 8 December 2020 / Published online: 16 April 2021
© The Author(s) 2021

Abstract

Purpose Electrocardiogram (ECG) signals collected from wearable devices are easily corrupted with surrounding noise and artefacts, where the signal-to-noise ratio (SNR) of wearable ECG signals is significantly lower than that from hospital ECG machines. To meet the requirements for monitoring heart disease via wearable devices, eliminating useless or poor-quality ECG signals (e.g., lead-falls and low SNRs) can be solved by signal quality assessment algorithms.

Methods To compensate for the deficiency of the existing ECG quality assessment system, a wearable ECG signal dataset from heart disease patients collected by Lenovo H3 devices was constructed. Then, this paper compares the performance of three machine learning algorithms, i.e., the traditional support vector machine (SVM), least-squares SVM (LS-SVM) and long short-term memory (LSTM) algorithms. Different non-morphological signal quality indices (i.e., the approximate entropy (ApEn), sample entropy (SaEn), fuzzy measure entropy (FMEn), Hurst exponent (HE), kurtosis (K) and power spectral density (PSD) features) extracted from the original ECG signals are fed into the three algorithms as input.

Results The true positive rate, true negative rate, sensitivity and accuracy are used to evaluate the performance of each method, and the LSTM algorithm achieves the best results on these metrics (97.14%, 86.8%, 97.46% and 95.47%, respectively).

Conclusions Among the three algorithms, the LSTM-based quality assessment method is the most suitable for the signals collected by the Lenovo H3 devices. The results also show that the combination of statistical features can effectively evaluate the quality of ECG signals.

Keywords Electrocardiogram (ECG) signal quality assessment · Machine learning · Long short-term memory (LSTM) · Support vector machine (SVM) · Least-squares support vector machine (LS-SVM)

1 Introduction

According to [1], the total number of people with cardiovascular disease (CVD) in China is approximately 290 million people, and it is the main cause of death. However, most cardiac arrhythmias cannot be confirmed because they are transient, paroxysmal, and sometimes asymptomatic. Hospitals currently use ambulatory devices (e.g., Holter monitors) to provide real-time dynamic monitoring of patients with suspected heart disease, typically over 24 h. Although they are accurate, these devices are expensive, uncomfortable to wear and affect patients' daily physiological activities. Studies have shown that continuous electrocardiogram (ECG) monitoring for 4 weeks can improve the accuracy of detecting intermittent atrial fibrillation (AF) by more than 5 percentage points [2] compared to short-term detection. Over the past few years, a number of wearable devices

Fan Fu and Wentao Xiang contributed equally to this work.

✉ Songsheng Zhu
zhu@njmu.edu.cn

✉ Jianqing Li
jqli@njmu.edu.cn

¹ The Key Laboratory of Clinical and Medical Engineering, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 211166, China

² The State Key Laboratory of Bioelectronics, School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China

³ College of Engineering, Zhejiang Normal University, Jinhua 321004, China

using textile electrodes have been invented for continuous non-invasive ECG monitoring. However, the ECG signals collected by wearable devices can be contaminated by environmental and electrode displacement noise, leading to a low signal-to-noise ratio (SNR). Research also shows that when a patient is walking, only 46.7–50.7% of wearable ECG signals are of high quality, and it is difficult to distinguish the P waves, which are an important feature of AF detection, from the rest of the signals [2]. These low-quality wearable ECG signals may lead to medical misdiagnosis [3]. To address this problem, automatic estimation of ECG signal quality is required.

Applying machine learning algorithms to ECG quality assessment has been widely researched [3–11]. The improvement in ECG signal quality mainly relied on two aspects: classification algorithms and quality assessment indices, as indicated in [8–11]. Among these classification algorithms, the support vector machine (SVM) algorithm is widely applied in the context of ECG quality assessment [6–12] and has achieved optimal performance [11, 12]. As also mentioned in [8], SVM and neural network algorithms showed extremely similar results in ECG signal quality assessment. Later, Zhang and his team compared the performance of four ECG quality assessment methods, including the random forest (RF), kernel SVM, least-squares SVM (LS-SVM), and multi-surface proximal SVM-based oblique RF methods, on seven characteristics and their combinations [11]; the best performance was achieved by the LS-SVM classifier with an accuracy of 92.2%. The above-mentioned comparison focused on the ECG recordings from the 2011 PhysioNet Computing in Cardiology Challenge (PCICC), which are routinely collected from wet electrodes. However, wearable devices use dry electrodes to collect ECG signals, which result in motion artifacts, especially when the subject is moving. Such noise will not occur in data recorded by wet electrodes. We compared the signal quality assessment performance by using traditional SVM and LS-SVM models for wearable ECG signals. Long short-term memory (LSTM) networks have achieved excellent results in ECG signal classification due to the nature of the temporal characteristics [13]. However, to our knowledge, it has not been applied to the assessment of ECG signal quality. Therefore, an LSTM network was also tested in this paper, given its excellent performance on continuous temporal signals.

Earlier studies failed to find effective indices for ECG signal quality assessment. For example, the overfitting problem occurred in [7], where the classification accuracy for the test set was only 83.6%. Regarding the ECG signal, quality assessment indices can be divided into morphology and non-morphological indices. However, the QRS and P waves collected in different environments or patients have varying morphologies [13], and the classification algorithm has a lack of generalizability. Therefore, this study focused on

non-morphological indices to assess the quality of wearable ECG signals (acceptable signals were labelled ‘1’, and unacceptable signals were labelled ‘0’) by considering the signal complexity, correlation, Gaussian and power spectral density (PSD) characteristics, and entropy features (approximate entropy (ApEn), sample entropy (SaEn), and fuzzy measure entropy (FMEn)), which have been used in quality assessment for a long time to describe the complexity of signals [14, 15]. The Hurst exponent (HE) indicates the duration of extreme values in the sample, describing the strength of the abnormality and the correlation of ECG signals [16]. The kurtosis (K) is a measure of a signal’s Gaussianity, defining the randomness of the signal [17]. In [8–12], PSD features were shown to be effective in evaluating the quality of ECG signals. Therefore, PSD features were also included the best feature group for the final comparison of the three classifiers (i.e., the SVM, LS-SVM and LSTM algorithms).

Recently, studies have focused on wearable ECG signals for quality assessment [12, 18]. However, most wearable ECG signals are recorded in patients without heart disease [18] or in healthy adults [12]. Note that although the authors of [12] tested a dataset with five categories (e.g., clear QRS complexes (level A), clarity of the majority of QRS complexes (level B), challenges with identifying the QRS complexes in 2–3-s time windows (level C), at least one 4–5-s signal episode for identifying QRS complexes (level D) and a lower signal quality than level D (level E); for more details, see [12]) and then their further classify the five categories into acceptable (levels A–D) and unacceptable (level E) categories. However, if the dataset lacks abnormal ECGs from heart disease patients, the system may classify a patient’s heart condition as unacceptable during the test. Therefore, unlike the above-mentioned studies, this paper focused on wearable ECG signals recorded by Lenovo H3 devices from patients with heart disease.

In this study, to compensate for the lack of ECG quality assessment research, a special dataset, WECG-HD, was established. Three machine learning algorithms were compared to build a better ECG quality assessment system. Section 2 describes the data, feature extraction and classifiers we used in the signal quality assessment task. Section 3 discusses the experimental results. Finally, the conclusions are summarized in Sect. 4.

2 Materials and Methods

The quality assessment procedure is displayed in Fig. 1. The procedure is divided into three steps: lead-fall detection, feature and training parameter preparation, and comparison.

To reduce the complexity of the training parameters, the wearable ECG signals are normalized, and lead-fall detection is performed. According to Hayn’s research [19],

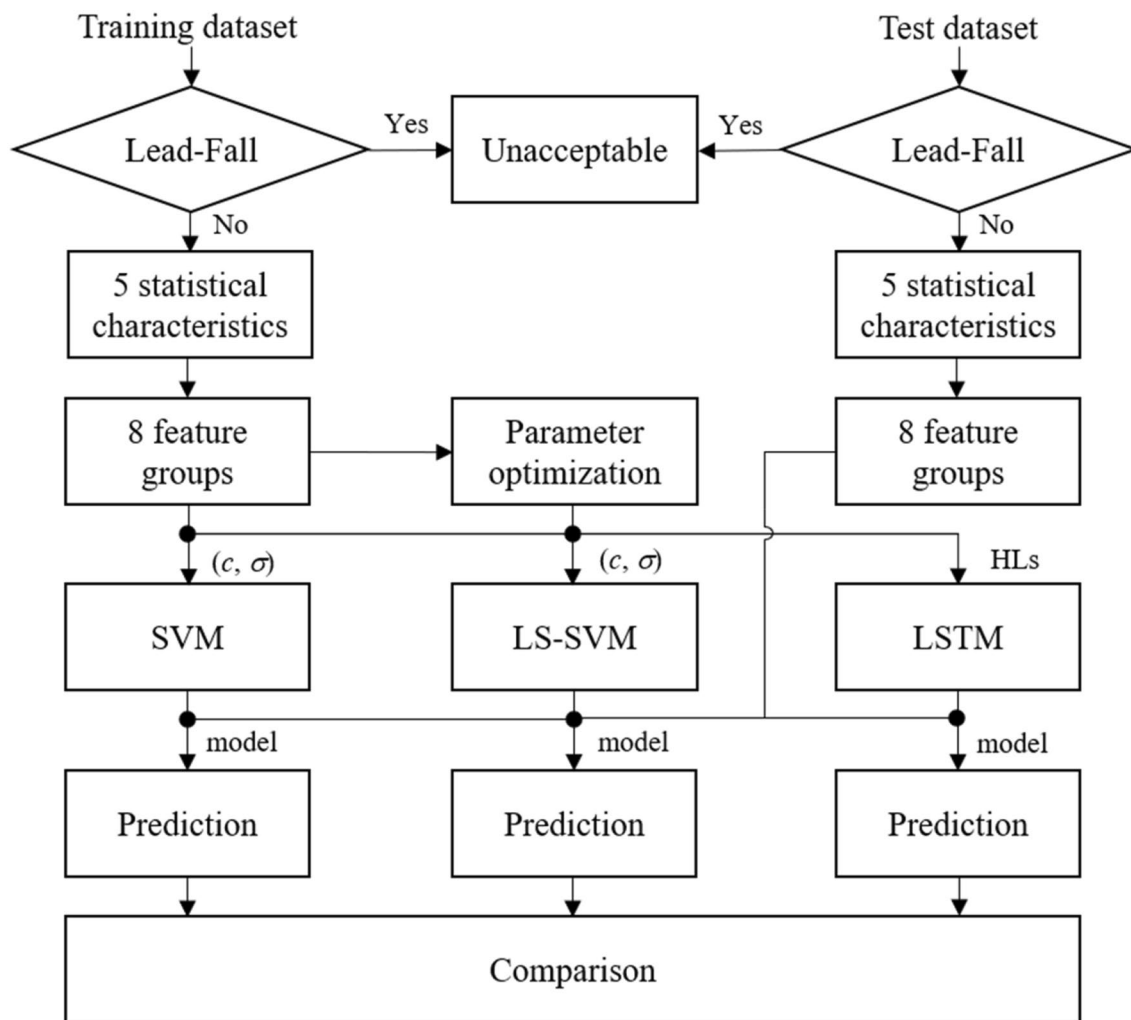


Fig. 1 The quality assessment procedure

if the ECG amplitude remains constant for a pre-set time (80%), then lead-fall (i.e., some electrodes are not in contact with the skin or the contact is poor [20]) is assumed to have occurred.

To select the optimal statistical features, five statistical features (i.e., the ApEn, SaEn, FME_n, HE, and K) calculated from the normalized signals are divided into eight groups considering the Gaussian characteristics, complexity, and correlation to compare the classification performance of the three classifiers (i.e., the SVM, LS-SVM, and LSTM classifiers). Then, tenfold cross-validation and grid search (GS)-based parameter optimization are used to find the training parameters of the SVM and LS-SVM classifiers for each of the eight groups. Note that the number of hidden layers (HLs) of the LSTM network is also explored.

The test dataset is used to evaluate the performance of the eight groups.

2.1 Data Recording and Labelling

The dataset used in this study included 24-h dynamic ECG signals collected from 200 heart disease patients (including patients with premature beats, AF, arrhythmia, etc.) using a Lenovo H3 wearable ECG device (designed by our laboratory and Lenovo) with two dry electrodes and two wet electrodes at a sampling rate of 400 Hz and a resolution of 12 bits. The Lenovo H3 device can simultaneously collect three-lead ECG signals, although this article used only one of the leads for the quality assessment study. The Lenovo H3 hardware integrates a simple filter, including the suppression of the power frequency and myoelectric and breathing interference, etc. The filtering range is from 0.05 to 40 Hz. The device records ECG signals for 90 h and stores them in the data storage module. The ECG data will be uploaded to the mobile app once the ECG recorder

is paired with a smart phone via low-power Bluetooth. The waveform, heart rate and other physiological information can be displayed in real time by an algorithm embedded in a mobile phone.

In total, 9000 single-channel wearable ECG segments (10-s each) were randomly selected and labelled by two individuals (clinical cardiologists) as acceptable or unacceptable (Table 1). A third clinical heart specialist examined the results, identified disagreements and made the final classification annotation. Typical unacceptable (a–c) and acceptable (d, e) signals are shown in Fig. 2.

2.2 Feature Description

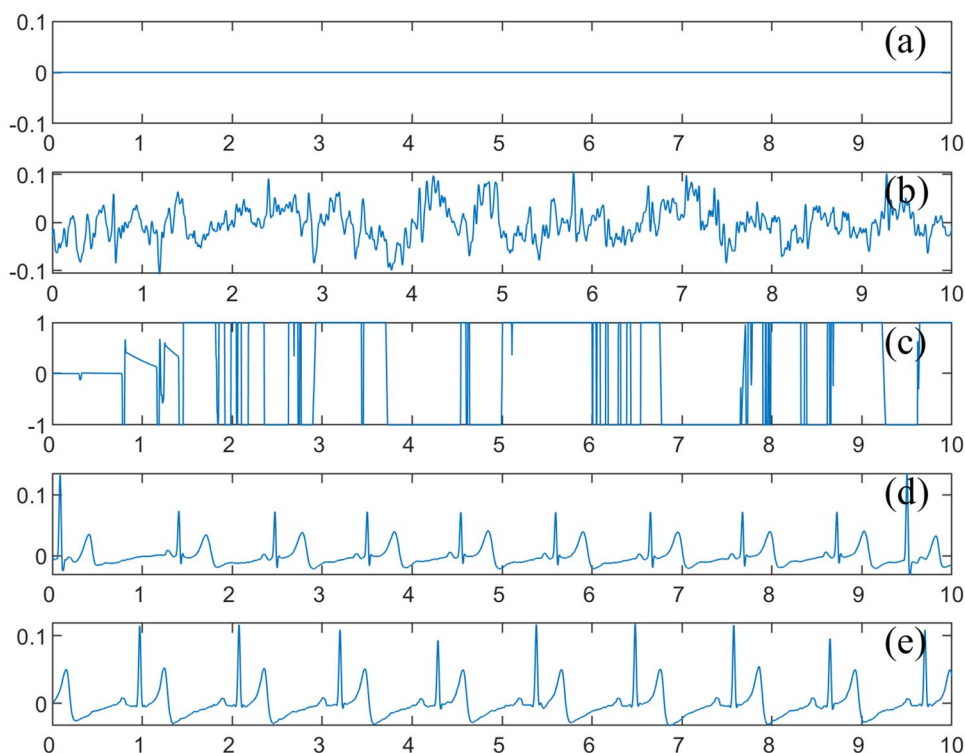
Effective features for ECG signal quality assessment are essential for classification. Hence, in this study, five statistical features (selected after the experiment) and the PSD features were extracted.

To investigate the performance of the five statistical characteristics in quantifying the quality of wearable ECG

Table 1 Details of the wearable ECG dataset

Dataset	Recordings			Time length (s)	Sampling rate
	Acceptable	Unacceptable	Total		
Training	3995	1005	5000	10	400 Hz
Test	3356	644	4000	10	400 Hz

Fig. 2 Typical 10-s signal-channel wearable ECG signals collected by the Lenovo H3 wearable ECG device: **a** ‘lead-fall’, **b** ‘unacceptable (high surrounding noise)’, **c** ‘unacceptable (electrode displacement)’, **d** and **e** ‘acceptable’



signals by the machine learning algorithms, eight groups are created (Table 3). Note that groups 1 and 2 represent the Gaussianity and correlation of the wearable ECG signals, respectively. Group 3 is the combination of the Gaussianity and the correlation. Inspired by [11], the ApEn may reduce the accuracy of the final classification, and the other two entropies (e.g., the SaEn and FMEn) with and without the ApEn are represented by groups 4 and 5, respectively. We include the three entropies together with all the possible two-feature (K and HE) combinations for groups 6–8.

After the comparison, the PSD features are added to the feature group, achieving the best performance for each classifier.

2.2.1 Approximate Entropy (ApEn)

The ApEn is a non-linear dynamic parameter for quantifying the regularity and unpredictability of time-series fluctuations [21]. A positive ApEn value represents the complexity (i.e., a regular signal is indicated by a small ApEn value, and vice versa) and the possibility of new information in the time series.

2.2.2 Sample Entropy (SaEn)

The SaEn is widely used to express the degree of signal regularity/irregularity [22]. The SaEn is simpler than the ApEn and more independent of the recording length. Moreover,

it exhibits relative consistency under a variety of circumstances and eliminates autocorrelation [23].

2.2.3 Fuzzy measure entropy (FMEn)

The SaEn is based on the Heaviside step function with a rigid boundary, which may have poor stability [24]. To address this problem, the FMEn is utilized to combine both local and global similarities in the entropy calculation, giving a better discrimination ability for time-series data [16].

2.2.4 Hurst Exponent (HE)

The HE is often used to characterize the long-term memory of time-series data. A large HE indicates a highly correlated sequence. Inspired by [25], in this paper, the HE is calculated by fitting function $(\log(R(n)/S(n)), \log(n))$, where $S(n)$ is the standard deviation of the original time series within n points and $R(n)$ is defined as follows:

$$R(n) = \max(0, W_1, W_2, \dots, W_n, \dots, W_N) - \min(0, W_1, W_2, \dots, W_n, \dots, W_N) \tag{1}$$

where $W_n = (x(1) + x(2) + \dots + x(n)) - n \times \mu$ ($n = 1, 2, \dots, N$), $x(n)$ is the ECG signal of the n -th sample and μ is the mean of the original time series across the n points.

2.2.5 Kurtosis (K)

The K value is used to quantify the Gaussian characteristic of the signals [16] and is given as follows:

$$K = \frac{1}{N} \sum_{n=1}^N \left(\frac{x(n) - \mu}{S(n)} \right)^4 \tag{2}$$

In general, the K value of a signal is equal to 3 for a Gaussian distribution; note that the K value of electromyography interference is approximately 5, and therefore, the K value for a high-quality normal sinus ECG signal should be slightly larger than 5 [26]. The K values for baseline drift and 50-Hz power line interference are both < 5 [27].

2.2.6 Power Spectral Density (PSD)

PSD features are widely used for assessing ECG quality [8–12]. According to [11], the frequency band of a normal ECG signal ranges from 0.05 to 100 Hz. The frequency bands of low- and high-frequency noise in the ECG signal are 0–1 Hz and 10–1000 Hz, respectively, which means that the high-frequency noise can cover the band of a normal ECG signal. Here, five PSD features (i.e., PSD_l : the power of low-frequency signals between 0 and 1 Hz; PSD_n : the power of the sinus signals between 0.05 and 100 Hz; PSD_h : the power of

high-frequency signals between 10 and 1000 Hz; $PSD_{h/n}$: the ratio of PSD_h to PSD_n ; and $PSD_{l/n}$: the ratio of PSD_l to PSD_n) are employed to reflect the energy distribution and the ratio of different frequency bands in the original wearable ECG signals.

2.3 Assessment Algorithm

In this section, the three classifiers used in this paper are introduced.

2.3.1 Support Vector Machine (SVM) Classifier

Linear SVMs are widely used in simple classification problems. To address the non-linear classification problem of the quality of wearable ECG signals, a kernel function is utilized by mapping the features to a high-dimensional space, where the non-linear problem can be transformed into a constrained quadratic optimization problem without a significant increase in the computational complexity. Such an optimization problem is defined as follows [28]:

$$\begin{aligned} \min_{w,b} & \frac{1}{2} \|w\|^2 \\ \text{s.t.} & y_n (w \cdot x_n + b) \geq 1, \quad n = 1, 2, \dots, N \end{aligned} \tag{3}$$

where x_n is the n -th support vector, w and b are the parameters to be calculated, y_n is the known label for the training vector, and c and k are the penalty parameter and kernel function, respectively. Then, the decision function $f(x)$ is given as:

$$f(x) = \text{sign} \left(\sum_{x_n \in SV} \alpha_n y_n k(x_n, x) + b \right) \tag{4}$$

where x is the sample in the test dataset and SV denotes the support vectors.

Here, three traditional kernel functions (i.e., the linear kernel, polynomial kernel and radial basis function kernel) are tested in our study. The best performance is achieved by the radial basis function kernel, $k(x_n, x)$, which is defined as follows:

$$k(x_n, x) = \exp \left(- \frac{\|x - x_n\|^2}{2\sigma^2} \right) \tag{5}$$

where σ is the scale parameter [10].

2.3.2 Least-Squares Support Vector Machine (LS-SVM) Classifier

According to [11], the LS-SVM classifier had a better ability to assess the quality of ECG signals in the PCICC dataset

than the SVM. The optimization problem of the LS-SVM is given as:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + \frac{c}{2} \sum_{n=1}^N e_n^2 \tag{6}$$

$$s.t. \ y_n(w \cdot x_n + b) \geq 1 - e_n, \quad n = 1, 2, \dots, N$$

where e_n is the training error of the n -th sample and the other parameters are the same as those defined for the SVM. The decision function of the LS-SVM is consistent with that of the SVM (Eq. (4)).

2.3.3 Long Short-Term Memory (LSTM) Classifier

LSTM is an algorithm developed on the basis of a recurrent neural network, which is mainly composed of three gates: a forget (F) gate, an input (I) gate, and an output (O) gate [13]. The LSTM formula is given as follows:

$$F_i^t = sig \left(\sum_{j=1}^N w_F[i,j]x_t[j] + \sum_{j=1}^{N_h} u_F[i,j]h_{t-1}[j] + b_F[i] \right) \tag{7}$$

$$I_i^t = sig \left(\sum_{j=1}^N w_I[i,j]x_t[j] + \sum_{j=1}^{N_h} u_I[i,j]h_{t-1}[j] + b_I[i] \right) \tag{8}$$

$$O_i^t = sig \left(\sum_{j=1}^N w_O[i,j]x_t[j] + \sum_{j=1}^{N_h} u_O[i,j]h_{t-1}[j] + b_O[i] \right) \tag{9}$$

where F_t , I_t and O_t are intermediate vectors at time t , N is the length of x , N_h ($1 \leq i \leq N_h$) is the number of HLs and sig denotes the sigmoid function. The weight (w , u) and bias parameters (b) are updated upon the completion of the program. The output vector (h_t) of the LSTM classifier at time t is:

$$h_t[i] = O_t[i] \times tanh(c_t[i]), \tag{10}$$

$$c_t[i] = F_t[i] \times c_{t-1}[i] + I_t[i] \times m_t[i]$$

where c_t is the intermediate vector at time t , m_t is the activation function formed by the tanh activation function, and m_t is defined as follows:

$$m_t^i = tanh \left(\sum_{j=1}^N w_O[i,j]x_t[j] + \sum_{j=1}^{N_h} u_O[i,j]h_{t-1}[j] + b_O[i] \right). \tag{11}$$

2.4 Classifier Parameter Determination

The performance of the SVM and LS-SVM algorithms is sensitive to the scale parameter σ and the penalty parameter c , as

mentioned in [10]. To address this problem, the GS method is used to search the (c, σ) combinations in a specified 2D parameter space to find the highest classification accuracy. The search range for each parameter is from 2^{-8} to 2^8 . Once the obtained optimal parameter is equal to the boundary, the range is expanded from 2^{-10} to 2^{16} , and the algorithm is optimized again. To avoid overfitting in the training set due to the obtained (c, σ) combination, the search step is set to 1. More details can be found in Sect. 3.2.

Once the parameter space is established, k -fold cross-validation is performed. As indicated in [10], a large value for k leads to a high computational complexity, and a small value for k can make the results insufficiently robust. Therefore, according to [29], a usual choice for k is 10.

The performance of the LSTM classifier is affected by the number of HLs, as mentioned in [13]. We also performed a GS for the number of HLs in the LSTM network in the range of 10–200.

2.5 Evaluation Metrics

Four metrics, namely, the true positive rate (PR), true negative rate (NR), sensitivity (Se) and accuracy (Acc), are used to evaluate the performance of each situation and algorithm. In addition, $Acc-CV$ (the average Acc across the 10 folds from the tenfold cross-validation step) is used in the parameter optimization process to evaluate the (c, σ) parameter combination. The definitions of PR , NR , Se , and Acc are given as follows:

$$PR = \frac{T_1}{T_1 + F_1} \times 100\% \tag{12}$$

$$NR = \frac{T_0}{T_0 + F_0} \times 100\% \tag{13}$$

$$Se = \frac{T_1}{T_1 + F_0} \times 100\% \tag{14}$$

$$Acc = \frac{T_1 + T_0}{T_1 + T_0 + F_1 + F_0} \times 100\% \tag{15}$$

where T_i ($\{i: i = 1, \text{‘acceptable’}; i = 0, \text{‘unacceptable’}\}$) indicates the number of correctly predicted signals and F_i represents the number of incorrectly predicted signals for label i .

3 Results and Discussions

The results of this paper are presented in four parts: lead-fall detection, parameter optimization, statistical characteristic comparison and classifier comparison.

3.1 Lead-Fall Detection

We report the lead-fall detection results in Table 2. *PR* and *NR* are calculated for lead-fall detection in the training set and the test set. From this table, we can see that a total of 464 and 179 10-s lead-fall signals are detected in the training set and the test set, respectively. Note that *PR* is equal to 100% in the training and test sets, while *NR* is 46.17% and 27.8% in the training and test sets, respectively. The results can be explained by the fact that poor signal quality contains lead-falls, electrode motion noise, motion artefacts and muscle artefacts [30].

3.2 Optimized Parameters

In this part, tenfold cross-validation and GS are utilized to search for optimal (c , σ) pairs for the SVM and LS-SVM classifiers for each of the eight feature groups, where the optimal (c , σ) pairs are given in Table 3. It can be seen from the parameter optimization results in Table 3 that the training parameters (c , σ) corresponding to each feature group are different since the large differences in the feature space are mapped by different inputs for different groups. For the same feature group, there is a large difference between the model trained with and without the best parameters because the training parameters (c , σ) can affect the solution of the optimal classification hyperplane. For example, the 3D view of parameter optimization for group 8 (shown in Fig. 3) shows the performance change, where *Acc-CV* is the criterion for evaluating the parameter combination. Note that the smallest *Acc-CV*

value is approximately 88%, as shown in Fig. 3, because of the imbalanced dataset between the composition of the ‘acceptable’ parts (3356 10-s segments) and the ‘unacceptable’ parts (644 10-s segments) in the test dataset (see Table 1).

For the LSTM classifier, we found that a relatively good result can be obtained when the number of HLs is equal to 80.

3.3 Comparison of the Statistical Characteristics

To investigate the effectiveness of the classifiers on quality assessment in the WECG-HD dataset, the three algorithms (SVM, LS-SVM and LSTM) are trained on the 8th statistical group with optimal parameters. The details of the quality assessment on the test set are summarized in Table 4. According to this table, all the feature groups except group 2 succeeded in identifying the ‘unacceptable’ signals, which means that a single HE may not be appropriate in this situation. In groups 1, 2 and 3, we note that these feature groups tend to classify signals as ‘acceptable’ since the *PR* value is higher than that of the other groups. According to the results of the three classifiers in groups 4 and 5, *Acc* is significantly increased when entropy is used to evaluate the ECG signal quality. This finding indicates that entropy is an important index for ECG quality assessment. As also shown in Table 4, the difference between groups 3 and 4 indicates that adding the ApEn feature can slightly improve the classification accuracy of the SVM classifier (from 95 to 95.03%) in the quality assessment method used in this paper.

For the SVM classifier, the classification performance generally improves as the number of features increases. For the LSTM classifier, the addition of entropy leads to a sharp increase in the classification accuracy, as seen in the comparison of the first three groups and the last five groups. This finding further illustrates the importance of entropy. Among the three classification algorithms, the classification accuracies of 90.33%, 90.22%, 92.67%, 91.5%, 91.77%, 91.25% and 92.65% for feature groups 1 and 3–8 obtained

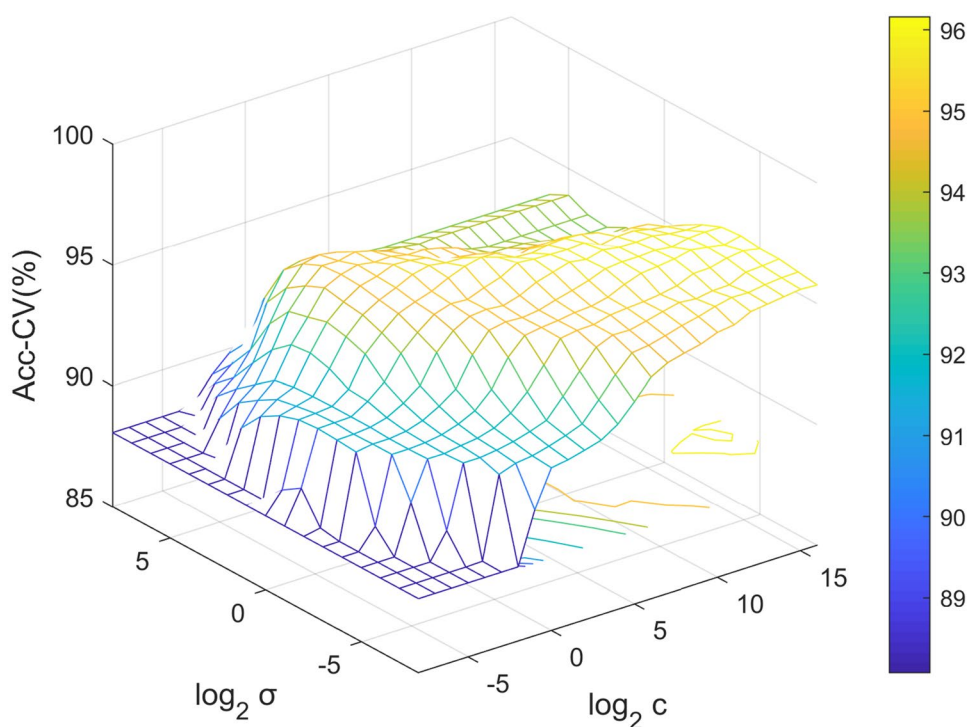
Table 2 Lead-fall detection results

Dataset	Classification results		
	Detected lead-fall numbers	<i>PR</i>	<i>NR</i>
Training	464	100%	46.17%
Test	179	100%	27.8%

Table 3 Parameters for SVM and LS-SVM

Group	Feature(s)	SVM		LS-SVM	
		c	σ	c	σ
1	K	512	0.125	32	2
2	HE	16,384	8	2048	0.00390625
3	K + HE	2	16	2048	4
4	SaEn + FMEEn	32,768	0.25	2048	0.015625
5	SaEn + FMEEn + ApEn	65,536	0.125	16,384	0.00390625
6	SaEn + FMEEn + ApEn + K	65,536	0.03125	32,768	0.125
7	SaEn + FMEEn + ApEn + HE	16,384	0.25	4096	0.0078125
8	SaEn + FMEEn + ApEn + K + HE	32,768	0.03125	32,768	1

Fig. 3 3D view of the tenfold cross-validation results using the GS method in Group 8



by the LS-SVM classifier are the lowest. The reason is that the LS-SVM classifier cannot obtain the optimal classification hyperplane to achieve a faster solution speed than the SVM classifier. The SVM and LSTM classifiers show similar accuracies (differences < 1%) for all the groups, while the LSTM classifier achieves the best classification accuracy of 95.8% (group 8).

3.4 Comparison of the Classifiers

In this section, the PSD features are also added to the feature groups to further investigate the performance of the classifiers. As shown in Table 5, the LSTM classifier achieves the best evaluation results, while the LS-SVM classifier achieves the worst evaluation results, which is consistent with the findings from Table 4. We note that the LSTM classifier has the highest retention rate ($PR = 98.66\%$) for ‘acceptable’ signals based on the highest classification accuracy. It is obvious that the application of PSD features can improve the performance of the SVM and LSTM classifiers (the *Acc* values for the LSTM and SVM classifiers increased from 95.8 to 96.73% and from 95.63 to 96.1%, respectively) but not that of the LS-SVM classifier (the *Acc* value for the LS-SVM classifier decreased from 92.67 to 91.83%).

4 Conclusion

In this paper, a wearable ECG signal dataset is established for the study of ECG quality assessment for practical applications. To find the most suitable classifier for evaluating the quality of wearable ECG signals, three machine learning algorithms (including the newly tested LSTM algorithm) are compared with the statistical features and PSD characteristics derived from the original dataset. As shown in the results, even though the best classification accuracy of the SVM algorithm is 96.1%, it is slightly lower than that of the LSTM algorithm (96.73%). The LS-SVM uses a faster optimization method, but its performance on the wearable ECG database is not satisfactory. The LSTM classifier, which many researchers have ignored, is the best evaluator (i.e., it achieves the highest classification accuracy of 96.73%) among the three classifiers compared in this paper. In addition, a combination of statistical features is utilized to evaluate the quality of ECG signals. It is noteworthy that such statistical features perform well in ECG signal quality assessment, and the best classification result for 10-s ECG segments can reach 95.8% via the LSTM algorithm. To further improve the quality assessment accuracy, PSD features are added with an improvement of 0.93 percentage points (i.e., from 95.8 to 96.73%) by the LSTM algorithm.

Table 4 Classification results on statistical features

Group	Method	Results on the test set			
		PR	NR	Se	Acc
1	SVM	99.85%	53.57%	91.81%	92.4%
	LS-SVM	100%	39.91%	89.66%	90.33%
	LSTM	99.2%	59.78%	92.78%	92.85%
2	SVM	99.88%	29.66%	88.09%	88.58%
	LS-SVM	100%	29.19%	88.04%	88.6%
	LSTM	98.33%	35.71%	88.85%	88.25%
3	SVM	98.66%	59.63%	92.72%	92.38%
	LS-SVM	99.97%	39.44%	89.59%	90.22%
	LSTM	98.9%	60.25%	92.84%	92.67%
4	SVM	96.48%	87.27%	97.53%	95%
	LS-SVM	99.55%	56.83%	92.32%	92.67%
	LSTM	97.2%	85.71%	97.26%	95.35%
5	SVM	96.42%	87.73%	97.62%	95.03%
	LS-SVM	98.36%	55.75%	92.05%	91.5%
	LSTM	96.99%	86.80%	97.46%	95.35%
6	SVM	96.75%	86.65%	97.42%	95.13%
	LS-SVM	98%	59.32%	92.62%	91.77%
	LSTM	98.57%	81.06%	96.44%	95.75%
7	SVM	96.48%	89.6%	97.97%	95.38%
	LS-SVM	98%	56.06%	92.08%	91.25%
	LSTM	96.93%	83.7%	96.87%	94.8%
8	SVM	97.38%	86.49%	97.41%	95.63%
	LS-SVM	99.4%	57.45%	92.41%	92.65%
	LSTM	98.6%	81.21%	96.47%	95.8%

The bold and italic values indicate the highest Acc for each classifier

Table 5 Classification performance on the non-morphological indices (statistics and PSD features)

Classifier	Features	Results on the test set			
		PR	NR	Se	Acc
SVM	Group 8 + PSD	97.14%	90.68%	98.19%	96.1%
SVM	Group 8	97.38%	86.49%	97.41%	95.63%
LS-SVM	Group 4 + PSD	98.42%	57.45%	92.34%	91.83%
LS-SVM	Group 4	99.55%	56.83%	92.32%	92.67%
LSTM	Group 8 + PSD	98.66%	86.65%	97.47%	96.73%
LSTM	Group 8	98.6%	81.21%	96.47%	95.8%

The bold and italic values indicate the highest classification accuracy

It should be mentioned that in contrast to many recently proposed studies [12, 18] that used ECG signals from non-cardiac patients, our study uses dynamic wearable ECG data from 200 patients with heart disease, making a small step towards application in basic clinical environments.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40846-020-00588-7>.

Acknowledgements The authors would like to thank the editor and reviewers for their valuable comments and suggestions. The authors would also like to thank Lenovo for their equipment production and data transmission services regarding the ECG devices.

Funding This research was funded by the National Key Research and Development Program of China (2017YFB1303200), the Jiangsu Provincial Key Research and Development Program (BE2017735), the National Natural Science Foundation of China (61571113, 81871444, 62071241, and 62001240), and the Zhejiang Provincial Natural Science Foundation of China (LY17F010003).

Compliance with Ethical Standards

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Liu, C., Yang, M., Di, J., Xing, Y., Li, Y., & Li, J. (2019). Wearable ECG: History, key technologies and future challenges. *Chinese Journal of Biomedical Engineering*, 38(6), 641–652.
2. Fouassier, D., Roy, X., Blanchard, A., & Hulot, J. S. (2020). Assessment of signal quality measured with a smart 12-lead ECG acquisition T-shirt. *Annals of Noninvasive Electrocardiology*, 25(1), e12682.
3. Li, Q., & Clifford, G. D. (2012). Signal quality and data fusion for false alarm reduction in the intensive care unit. *Journal of Electrocardiology*, 45(6), 596–603.
4. Redmond, S. J., Xie, Y., Chang, D., Basilakis, J., & Lovell, N. H. (2012). Electrocardiogram signal quality measures for unsupervised telehealth environments. *Physiological Measurement*, 33(9), 1517–1533.
5. Naseri, H., & Homaeinezhad, M. (2014). Electrocardiogram signal quality assessment using an artificially reconstructed target lead. *Computer methods in Biomechanics and Biomedical Engineering*, 18(10), 1126–1141.
6. Clifford, G. D., Behar, J., Li, Q., & Rezek, I. (2012). Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. *Physiological Measurement*, 33(9), 1419–1433.
7. Kužílek, J., Huptych, M., Chudáček, V., Spilka, J., & Lhotská, L. (2011). Data driven approach to ECG signal quality assessment using multistep SVM classification. In *Computing in Cardiology* (pp. 453–455).

8. Behar, J., Oster, J., Li, Q., & Clifford, G. D. (2013). ECG signal quality during arrhythmia and its application to false alarm reduction. *IEEE Transactions on Biomedical Engineering*, *60*(6), 1660–1666.
9. Li, Q., Rajagopalan, C., & Clifford, G. D. (2014). A machine learning approach to multi-level ECG signal quality classification. *Computer Methods and Programs in Biomedicine*, *117*(3), 435–447.
10. Zhang, Y., Liu, C., Wei, S., Wei, C., & Liu, F. (2014). ECG quality assessment based on a kernel support vector machine and genetic algorithm with a feature matrix. *Journal of Zhejiang University Science C*, *15*(7), 564–573.
11. Zhang, Y., Wei, S., Zhang, L., & Liu, C. (2019). Comparing the performance of random forest, SVM and their variants for ECG quality assessment combined with nonlinear features. *Journal of Medical and Biological Engineering*, *39*(3), 381–392.
12. Liu, C., Zhang, X., Zhao, L., Liu, F., Chen, X., Yao, Y., & Li, J. (2019). Signal quality assessment and lightweight QRS detection for wearable ECG SmartVest system. *IEEE Internet of Things Journal*, *6*(2), 1363–1374.
13. Saadatnejad, S., Oveisi, M., & Hashemi, M. (2020). LSTM-based ECG classification for continuous monitoring on personal wearable devices. *IEEE Journal of Biomedical and Health Informatics*, *24*(2), 515–523.
14. Liu, C., Li, P., Di Maria, C., Zhao, L., Zhang, H., & Chen, Z. (2014). A multi-step method with signal quality assessment and fine-tuning procedure to locate maternal and fetal QRS complexes from abdominal ECG recordings. *Physiological Measurement*, *35*(8), 1665–1683.
15. Zhu, S., Xu, Z., Yin, K., & Xu, Y. (2011). Effects of quantization on detrended fluctuation analysis. *Chinese Physics B*, *5*, 153–158.
16. Li, Q., Mark, R. G., & Clifford, G. D. (2008). Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter. *Physiological Measurement*, *29*(1), 15–32.
17. Wang, L., Wang, Y., & Chang, Q. (2016). Feature selection methods for big data bioinformatics: A survey from the search perspective. *Methods*, *111*, 21–31.
18. Orphanidou, C., & Drobnjak, I. (2017). Quality assessment of ambulatory ECG using wavelet entropy of the HRV signal. *IEEE Journal of Biomedical and Health Informatics*, *21*(5), 1216–1223.
19. Hayn, D., Jammerbund, B., & Schreier, G. (2012). QRS detection based ECG quality assessment. *Physiological Measurement*, *33*(9), 1449–1461.
20. Calabria, A. (2012). Understanding lead-off detection in ECG. *Texas Instruments, Tech. Rep.*
21. Wachowiak, M. P., Hay, D. C., & Johnson, M. J. (2016). Assessing heart rate variability through wavelet-based statistical measures. *Computers in Biology and Medicine*, *77*, 222–230.
22. Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, *278*(6), 2039–2049.
23. Zhao, L., Wei, S., Zhang, C., Zhang, Y., Jiang, X., Liu, F., & Liu, C. (2015). Determination of sample entropy and fuzzy measure entropy parameters for distinguishing congestive heart failure from normal sinus rhythm subjects. *Entropy*, *17*(9), 6270–6288.
24. Liu, C., Li, K., Zhao, L., Liu, F., Zheng, D., Liu, C., & Liu, S. (2013). Analysis of heart rate variability using fuzzy measure entropy. *Computers in Biology and Medicine*, *43*(2), 100–108.
25. Tulppo, M., Hughson, R., Timo, H., Airaksinen, K., & Huikuri, H. (2001). Effects of exercise and passive head-up tilt on fractal and complexity properties of heart rate dynamics. *AJP Heart and Circulatory Physiology*, *280*(3), H1081–H1087.
26. He, T., Clifford, G., & Tarassenko, L. (2006). Application of independent component analysis in removing artefacts from the electrocardiogram. *Neural Computing & Applications*, *15*(2), 105–116.
27. Clifford, G. D., Azuaje, F., & McSharry, P. (2006). *Advanced methods and tools for ECG data analysis*. Boston: Artech House.
28. Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and Their Applications*, *13*(4), 18–28.
29. Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40–79.
30. Zhao, Z., Liu, C., Li, Y., Li, Y., Wang, J., Lin, B. S., & Li, J. (2019). Noise rejection for wearable ECGs using modified frequency slice wavelet transform and convolutional neural networks. *IEEE Access*, *7*, 34060–34067.