



A Two-Stream Model Combining ResNet and Bi-LSTM Networks for Non-contact Dynamic Electrocardiogram Signal Quality Assessment

Guoqiang Zhu¹, Yang Li¹, Yonglin Wu¹, Zhikun Lie¹, Chen Chen²,
and Wei Chen^{1,2}(✉)

¹ Center for Intelligent Medical Electronics, School of Information Science and Technology,
Fudan University, Shanghai 200433, China
{20210720211, w_chen}@fudan.edu.cn
² Human Phenome Institute, Fudan University, Shanghai 201203, China

Abstract. With the rapid advancement of monitoring without contact electrocardiogram (ECG), dynamic and real-time signal quality assessment (SQA) becoming a practical problem. In this paper, a two-stream structure that combines residual network (ResNet) and bidirectional long short-term memory (Bi-LSTM) for dynamic ECG (dECG) signals quality assessment is proposed. The ResNet stream is dedicated to extracting spatial features using time-frequency spectrum images as inputs. Meanwhile, the Bi-LSTM stream is devoted to exploring the temporal features using the ECG time series as the input. Then, these two streams are fused using a decision fusion mechanism and the performance is significantly boosted. As compared to the single stream-based approach, the proposed structure can compensate for either temporal or spatial information effectively. The overall accuracy of 99.69% can be achieved in distinguishing the dECG signal quality into three categories, namely clear ECG signal with clear R waves, blurry ECG signal without clear R waves, and noisy ECG signal with motion artifact. Experimental results show that this method demonstrates superior accuracy in determining the quality of dECG signals measured by the noncontact device. Therefore, the proposed model is expected to be a promising solution for non-contact dECG signal quality assessment in a practical ECG diagnosis.

Keywords: Non-contact ECG · Signal quality assessment · Convolutional neural network (CNN) · Recurrent neural network (RNN)

1 Introduction

With the fast-developing sensing technologies, the progress in non-contact electrocardiogram (ECG) monitoring is significantly promoted [1]. It's an important tool to detect cardiovascular diseases (CVDs) early [2]. However, noncontact measurement methods

are sensitive to noise. These include baseline drift, electromagnetic signals, motion artifacts, etc. Poor signal quality affects reliable manual or automatic ECG analysis; a lack of signal quality makes it difficult to understand the correct diagnosis information [3], raises the probability of false alerts [4], and increases the assignment of doctors [5]. Therefore, it's important to use automatic signal quality assessment (SQA) methods. A reliable SQA method helps to give suggestions of re-take recording when signal quality isn't good [6], to improve the efficiency of information transmission by rejecting signals of bad quality [7], and to deliver reliable cardiac recordings for CVDs scanning [8].

Many studies have been conducted on the quality evaluation of ECG signals in the previous literature. With the development of computer analysis and processing capabilities, the ECG SQA algorithm can be split into three categories approximately: rule-based SQA methods, machine learning-based SQA methods, and deep learning-based SQA methods. Early designed signal algorithms were limited in processing capacity. Thus, low-complexity algorithms were usually designed with rule-based SQA methods combined with waveform and interval features [9, 10]. The accuracy of this method for multi-classification ECG SQA is barely satisfactory. It is also proposed that ECG SQA has been used using classical machine learning (ML) algorithms. However, to identify the available features, such techniques require time-consuming feature engineering. For instance, Y. Zhang [11] input some features from ECG signals into ML models like decision tree (DT) and support vector machine (SVM). Such classical ML models include many intrinsic limitations, such as weak generalization capability and boring feature extraction. Moreover, deep learning [12] doesn't require complicated feature engineering and has been widely applied to many domains such as picture classification [13], voice recognition [14], and Intelligent translation [15]. Non-contact ECG signals are non-stationary millivolt signals with a low signal-to-noise ratio and it is easy to be interfered with by other signals. Therefore, it's hard to classify the dECG signals with high accuracy through a single neural network. In this paper, a dECG classification algorithm based on deep Residual Networks (ResNet) and Bi-directional Long Short-Term Memory (Bi-LSTM) networks is proposed, which achieves a high accuracy three-classes classification of dECG signals.

The rest of the article is on the following: Sect. 2 introduces the Materials and Methods. Section 3 introduces the experimental setup, results, and comparison with other methods. Finally, a conclusion is in Sect. 4.

2 Materials and Methods

In this section, we will introduce the materials and methods involved in our work. Firstly, we introduce the acquisition and annotation of data involved in our work. Secondly, we transform data to time-frequency spectrum images as network input. At last, we propose the SQA method based on ResNet and Bi-LSTM.

2.1 Data Collection and Annotation

The data of this article is collected by a non-contact dECG measurement system published previously by our teammates [16]. Acquisition of dECG is based on the capacitively coupled electrode and it can acquire ECG signal both in contact with skin and through clothes. The system can denoise and store data simultaneously. Forty volunteers (25 men and 15 women, average age 25 ± 10), were enrolled in this study. All subjects have good health and no history of cardiovascular disease. Meanwhile, all of them signed the informed consent. The dECG signals were measured at a sampling frequency of 500 Hz for 4 h. The collected signals were divided into five-second segments. 4859 sampled signals with obvious statistical characteristics were selected. By consulting ECG experts, we define the dECG signal quality categories as follows.

- Level 0 (clean): A good recording with clear R waves.
- Level 1 (noise): A poor recording without clear R waves.
- Level 2 (motion artifact): A high-amplitude noise signal caused by the movement of the testee.

The waveforms of three types of dECG signals are shown in Fig. 1. Four undergraduates of biomedical engineering with training in ECG analysis annotated each segment respectively and an ECG expert reviewed each annotation. Finally, five of them voted to determine the category of dECG signal segments. The number and categories distribution of dECG signal segments taken for the training set, validation set, and test set are shown in Table 1.

2.2 Pre-processing of dECG Signals

Short-time Fourier transform (STFT) [17] is a popular time-frequency analysis tool that explains frequency domain features in a sequence with a fragment of a signal in the time domain. The frequency features of different window functions are different in STFT. So, Time-domain and frequency-domain resolution should all be thought over.

We used the frequency spectrum of our dECG signals into the CNN model in an effort akin to voice classification [18–21]. This study used a symmetric Hamming window for 8 points that decreased spectrum leakage and obtained superior time-frequency particulars after many experiments. Consequently, after STFT, dECG signals (with a dimension of 1×2500) were transformed into the time-frequency spectrum image (with a dimension of $875 \times 656 \times 3$). The time-frequency spectrum images for three categories of dECG signals are shown in Fig. 1.

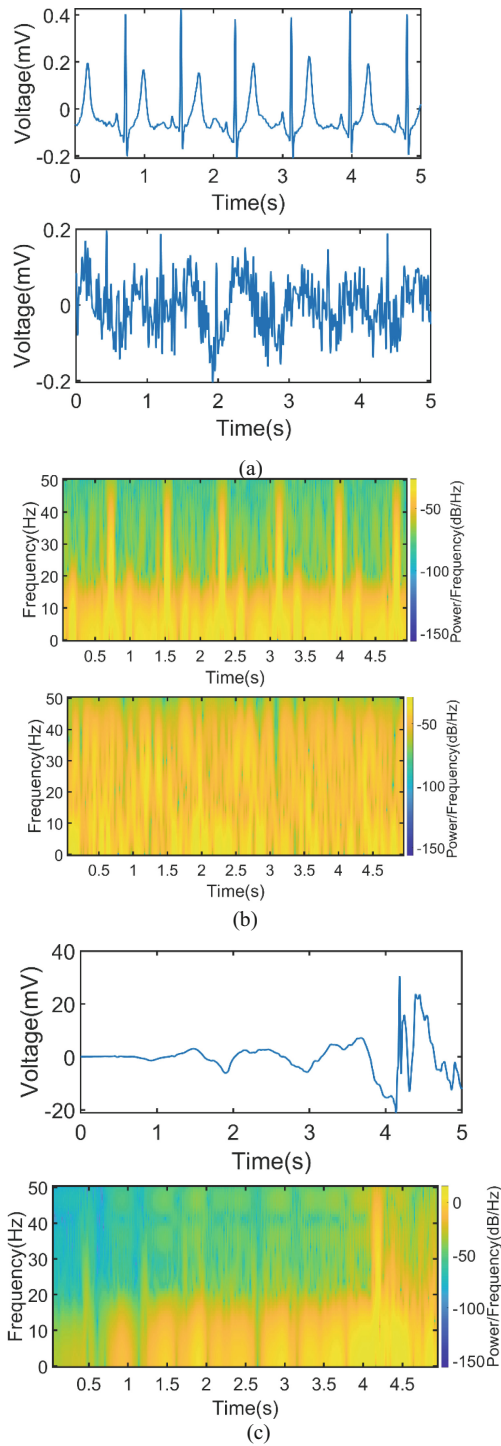


Fig. 1. (a), (b), and (c) show three types of dECG signals and their time-frequency spectrum image respectively

Table 1. The number and categories distribution of dECG signal segments

Category	Training	Validation	Test	Total
Level 0	1000	300	382	1682
Level 1	1000	300	296	1596
Level 2	1000	300	281	1581

2.3 Architecture of Proposed Model

The proposed algorithm model is mainly composed of two parts, one is ResNet, the other is Bi-LSTM. We first resize the time-frequency spectrum image of dECG signals to the

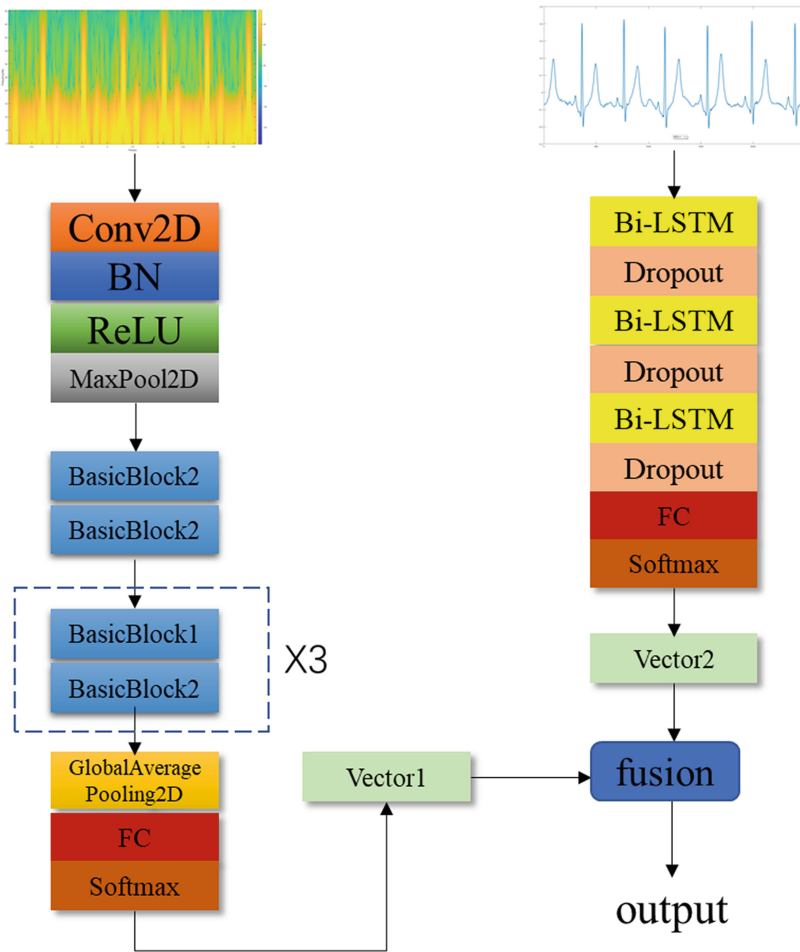


Fig. 2. The architecture of our proposed model.

size of $256 \times 256 \times 3$ by bilinear interpolation and then use them to train the ResNet network. We directly use the dECG signals (with a dimension of 1×2500) to train the Bi-LSTM network. When the test set is input to the two trained networks, we take out the output vectors of the softmax layer of the two networks, add them and multiply them by 0.5. In this way, the prediction probability value of each type is obtained. Finally, the category with the largest probability value is taken as the prediction category. The whole framework of the model is shown in Fig. 2. This model is composed of two networks, ResNet and Bi-LSTM, which improve the total accuracy and robustness of the model by using the idea of ensemble learning.

Description of CNN Layers. The time-frequency spectrum images of dECG signals pass through the two-dimensional convolutional layer (Conv2D) and two-dimensional maximum pooling (Maxpool2D) layer and then get into the residual module. The residual module is composed of 8 basic blocks of 2 kinds, as shown in Fig. 3. The kernel dimension of the Conv2D layer in this network was all 3×3 . The first Conv2D layer had 32 convolution kernels, and the number of kernels for the Conv2D layer among the 8 basic blocks in the latter residual module was taken as 32, 32, 64, 64, 128, 128, 128, 128. When the CNN implemented convolution layer by layer, the acquired characteristics were transformed from particulars to abstractions. The residual module could catch a lot of messages from the characteristics acquired from different Conv2D layers, thus effectively improving features' recognition capabilities and avoiding a substantial increase in network parameters. Moreover, the residual module provided not only characteristic extraction, but also improved gradient flow [22]. ReLU is the activation function of total the Conv2D layers. For better training, we add the BN layer after the convolution layer. Finally, input the output vectors of ResNet to the softmax layer to obtain the vector containing the probability value of each category.

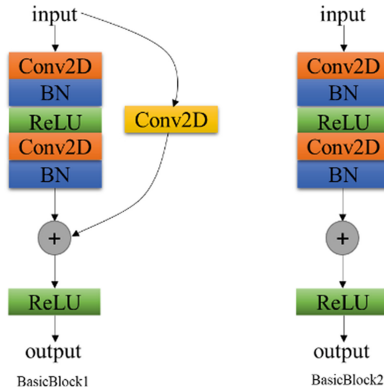


Fig. 3. Structure of two types of residual blocks.

Description of RNN Layers. Because Bi-LSTM has a good effect in processing timing signals, we take the dECG signal as the input of the Bi-LSTM network. Specifically, the module contains 3 Bi-LSTM layers [23], and each layer's size was 64. As seen in Fig. 4, Bi-LSTM, as a particular LSTM, receives information in both forward and backward directions at the same time, which makes the characteristics of time sequence richer, because each signal possesses context message from past and future. In the meantime, In order to prevent overfitting, we add the Dropout layer with a coefficient of 0.5 after each Bi-LSTM layer. Then, the output of the Bi-LSTM was input to the softmax layer to acquire the vectors containing the probability value of each category.

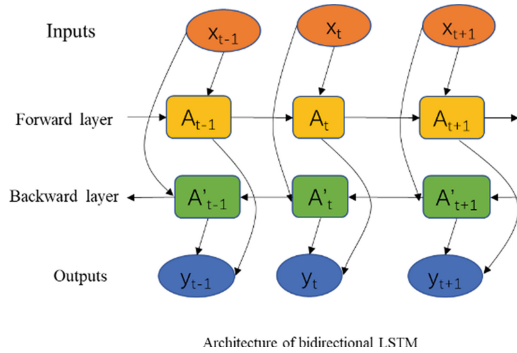


Fig. 4. The architecture of Bi-LSTM.

Training Method. We randomly extract data from the training set in batches to train the model. The weight parameters of the model are continuously updated by the gradient descent method. In this experiment, we use the Tensorflow framework to build and train our network model. We choose the commonly used cross-entropy loss function and Adam optimizer [24] to optimize the parameters. The cross-entropy loss function determines the difference between the resulting and the true values for each round of training samples. The Adam optimizer uses momentum and adaptive learning rate to accelerate model convergence. We set the epoch to 200 and the batch_size to 64.

3 Evaluation Index Experimental Platform and Results

In this section, we first evaluate our model with some evaluation criteria. Then we describe the hardware and software we use to process data and training models. Finally, we discuss the result of our experiments and compare it with some commonly used methods.

3.1 Evaluation Index

We randomly selected 959 segments from 4859 dECG signals as the test set. To evaluate the trained model, we applied the confusion matrix and calculate the sensitivity (Se), precision (P_+), accuracy (ACC), and F-Score of the fusion classifier model. The confusion matrix is a table that includes three indices represented by true-positive (TP), true-negative (TN), false-negative (FN), and false-positive (FP). The calculation formula of each index is as follows:

Accuracy (ACC): The proportion of correct predictions in total predictions.

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

Precision (P_+): The proportion of correct predicted positives to the total predicted positives.

$$P_+ = \frac{TP}{TP + FP} \quad (2)$$

Sensitivity (Se): It is also called Recall. Sensitivity describes the proportion of all positive cases identified in all positive cases.

$$Se = \frac{TP}{TP + FN} \quad (3)$$

F1-Score (F1): It gives a method to merge Sensitivity and Precision.

$$F1 = \frac{2 \times Se \times P_+}{Se + P_+} \quad (4)$$

3.2 Experimental Platform

The proposed network was performed using the Tensorflow framework based on the Python language and MATLAB 2020b was used for data processing. Total experiments were implemented under a Linux OS on a server with CPU Intel(R) Xeon(R) E5-2687 W v4 @ 3.00 GHz, GPU NVIDIA GTX 1080Ti, and 64 GB of RAM.

3.3 Results

Individual ResNet and Bi-LSTM models perform well in the training set, achieving accuracy rates of close to 100% and 95%, respectively. Although their respective accuracy on the validation set is also high, there is a large fluctuation in the accuracy among epochs. The changes of accuracy and loss values of the two networks on the training data and validation data with epoch are shown in Fig. 5. Therefore, we use the idea of ensemble learning to combine the trained ResNet and Bi-LSTM networks to reduce the overall variance of the model and improve the accuracy. The accuracy of using the ResNet model and the Bi-LSTM on the test set was 0.9499 and 0.9541, respectively. The accuracy of using the ResNet + Bi-LSTM model on the test set has been improved to 0.9969. Further comparison of the confusion matrix (ResNet + Bi-LSTM network vs. ResNet, Bi-LSTM) showed that the ResNet + Bi-LSTM network was more robust with higher accuracy (Fig. 6). As shown in Table 2, the ResNet + Bi-LSTM network also performed better on these evaluation indices of precision, sensitivity, and F1 score.

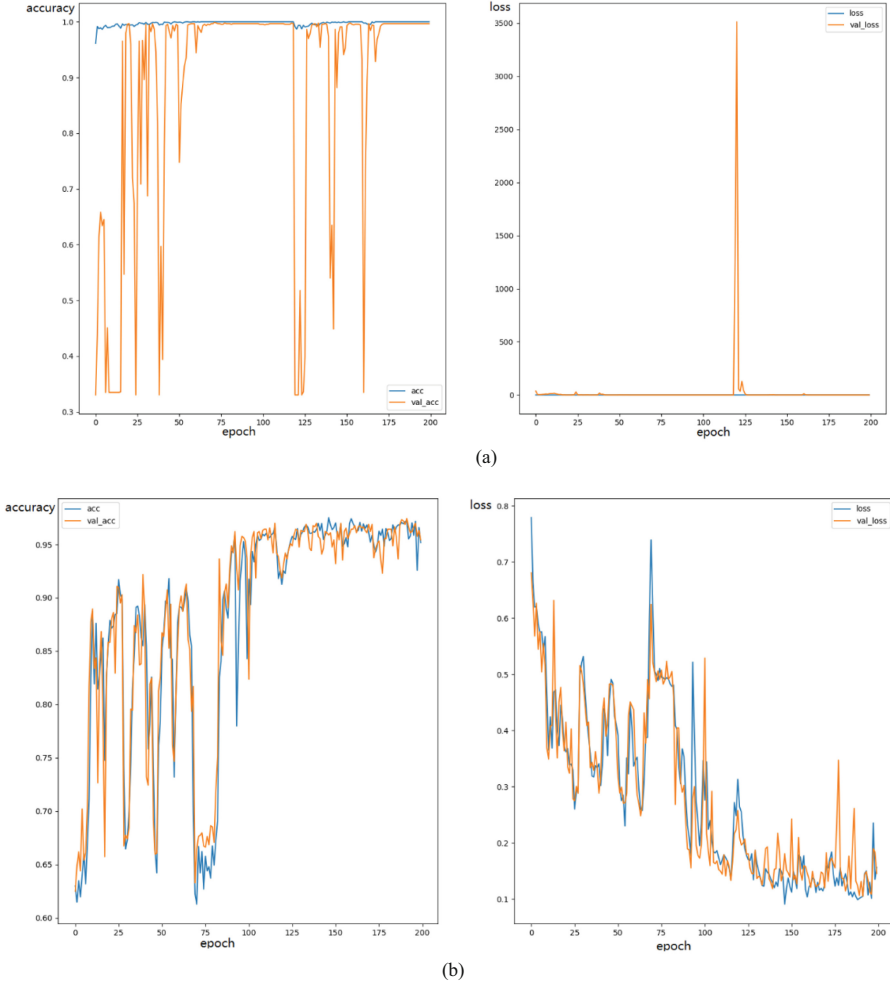


Fig. 5. (a), (b) the accuracy and loss value of each epoch on the training set and validation set of ResNet and Bi-LSTM networks

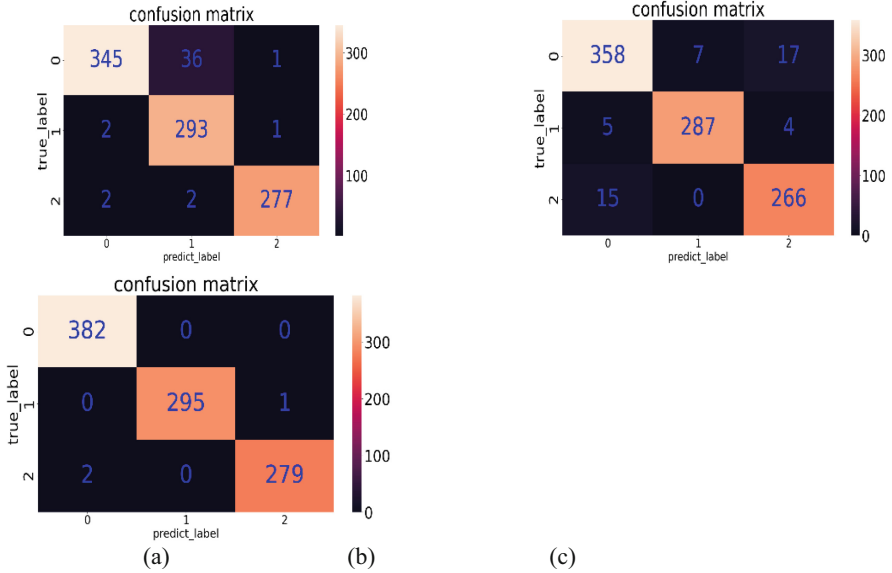


Fig. 6. (a), (b) and (c) are the confusion matrices of Bi-LSTM, ResNet, and ResNet + Bi-LSTM on the test set respectively

Table 2. Comparison of three methods

Method	Accuracy	Precision	Sensitivity	F1
ResNet	0.9541	0.9579	0.9541	0.9543
Bi-LSTM	0.9499	0.9501	0.9499	0.9499
ResNet + Bi-LSTM	0.9969	0.9969	0.9969	0.9969

3.4 Comparison with Methods in Other Papers

Our proposed method improved accuracy in comparison to the existing methods. Xue Zhou et al. [25] used a 1D CNN to classify ECG signals on the database of PhysioNet/Computing in Cardiology Challenge 2011 and the PhysioNet/Computing in Cardiology Challenge 2017 and obtained an accuracy of 94.30%. A. Huerta et al. [26] used 2D CNN to classify ECG signals on the database of PhysioNet/Computing in Cardiology Challenge 2017 and obtained an accuracy of 91.20%. Zeyang Zhu et al. [27] used the Adaboost algorithm to classify ECG signals on the database of PhysioNet/Computing in Cardiology Challenge 2011 and obtained an accuracy of 95.80%. Qifei Zhang et al. [28] used 1D CNN and 2D CNN to classify ECG signals on the MIT-BIH Arrhythmia database and obtained an accuracy of 91.80%. Although we used different datasets, the comparison of these methods is instructive for the evaluation of the model. As shown in Table 3, our algorithm has higher accuracy and robustness.

Table 3. Comparison of the proposed network and other methods

Work	Method	Accuracy	Precision	Sensitivity	F1
X. Zhou et al. [25]	1D CNN	0.9430	0.9550	0.9130	0.9335
A. Huerta et al. [26]	2D CNN	0.9120	0.9030	1	0.9490
Z. Zhu et al. [27]	Adaboost	0.9580	0.8720	0.9860	0.9255
Q. Zhang et al. [28]	1D CNN + 2D CNN	0.9180	0.9583	0.8920	0.9240
Our work	2D CNN + Bi-LSTM	0.9969	0.9969	0.9969	0.9969

4 Conclusion

Non-contact ECG measurement is sensitive to interference and motion artifact, limiting its clinical application. Hence, it greatly increases the need for real-time and dynamic signal quality assessment. In this paper, a ResNet + Bi-LSTM model is proposed to realize a three-level ECG signal quality assessment (clear ECG signal, blurry ECG signal, and noisy ECG signal). The ResNet layers can capture spatial features in the ECG sequences while the Bi-LSTM layer can then learn the temporal features. The proposed model achieves 99.69% accuracy, outperforming the single ResNet - based and Bi-LSTM - based approach. The experimental results demonstrate the superior performance of this ResNet + Bi-LSTM model over the existing methods. By automatic ECG quality classification with high accuracy, wide clinical applications of non-contact ECG measurement may be achieved in the future.

References

- Johannesen, L., Galeotti, L.: Automatic ECG quality scoring methodology: mimicking human annotators. *Physiol. Meas.* **33**(9), 1479 (2012). <https://doi.org/10.1088/0967-3334/33/9/1479>
- Clifford, G.D., Moody, G.B.: Signal quality in cardiorespiratory monitoring. *Physiol. Meas.* **33**(9), 1–6 (2012). <https://doi.org/10.1088/0967-3334/33/9/e01>
- Jekova, I., Krasteva, V., Christov, I., Abächerli, R.: Threshold-based system for noise detection in multilead ECG recordings. *Physiol. Meas.* **33**(9), 1463–1477 (2012). <https://doi.org/10.1088/0967-3334/33/9/1463>
- Liu, C., Li, P., Zhao, L., Liu, F., Wang, R.: Real-time signal quality assessment for ECGs collected using mobile phones. In: *Proceedings of the 2011 Computing in Cardiology*, Hangzhou, China, pp. 357–360. IEEE (2011)
- Allen, J., Murray, A.: Assessing ECG signal quality on a coronary care unit. *Physiol. Meas.* **17**(4), 249–258 (1996)
- Clifford, G.D., Behar, J., Li, Q., Rezek, I.: Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms. *Physiol. Meas.* **33**(9), 1419–1433 (2012). <https://doi.org/10.1088/0967-3334/33/9/1419>
- Silva, I., Moody, G.B., Celi, L.: Improving the quality of ECGs collected using mobile phones: the physionet/computing in cardiology challenge 2011. In: *Proceedings of the 2011 Computing in Cardiology*, Hangzhou, China, pp. 273–276. IEEE (2011)
- Xia, H., Garcia, G.A., Bains, J., Wortham, D.C., Zhao, X.: Matrix of regularity for improving the quality of ECGs. *Physiol. Meas.* **33**(9), 1535–1548 (2012). <https://doi.org/10.1088/0967-3334/33/9/1535>

9. Shi, Y., et al.: Robust assessment of ECG signal quality for wearable devices. In: Proceedings of the IEEE International Conference on Healthcare Informatics (ICHI), Xi'an, China, pp. 1–3 (2019)
10. Satija, U., Ramkumar, B., Manikandan, M.S.: Robust cardiac event change detection method for long-term healthcare monitoring applications. *Healthc. Technol. Lett.* **3**(2), 116–123 (2016)
11. Zhang, Y., Wei, S., Zhang, L., Liu, C.: Comparing the performance of random forest, SVM and their variants for computational and mathematical methods in medicine 11 ECG quality assessment combined with nonlinear features. *J. Med. Biol. Eng.* **39**(3), 381–392 (2019)
12. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
13. Rawat, W., Wang, Z.: Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput.* **29**(9), 2352–2449 (2017)
14. Xiong, W., Wu, L., Allewa, F., Droppo, J., Huang, X., Stolcke, A.: The Microsoft 2017 conversational speech recognition system. In: Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, Canada (2018)
15. Hill, F., Cho, K., Jean, S., Bengio, Y.: The representational geometry of word meanings acquired by neural machine translation models. *Mach. Transl.* **31**(1–2), 3–18 (2017). <https://doi.org/10.1007/s10590-017-9194-2>
16. Peng, S., et al.: Flexible electrodes based smart mattress for monitoring physiological signals of heart and autonomic nerves in a non-contact way. *IEEE Sensors J.* **21**(1), 6–15 (2020)
17. Nawab, S.H., Quatieri, T.F.: Short-time Fourier transform. In: *Advanced Topics in Signal Processing*, pp. 289–337, Prentice-Hall, Upper Saddle River (1987)
18. Liu, Z., Wu, Z., Li, T., Li, J., Shen, C.: GMM and CNN hybrid method for short utterance speaker recognition. *IEEE Trans. Industr. Inform.* **14**(7), 3244–3252 (2018)
19. Zhang, H., McLoughlin, I., Song, Y.: Robust sound event recognition using convolutional neural networks. In: Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 559–563, South Brisbane, Australia (2015)
20. Fu, S.W., Tsao, Y., Lu, X.: SNR-aware convolutional neural network modeling for speech enhancement. In: Proceedings of the Interspeech 2016, San Francisco, CA, pp. 3768–3772 (2016)
21. Satt, A., Rozenberg, S., Hoory, R.: Efficient emotion recognition from speech using deep learning on spectrograms. In: Proceedings of the Interspeech 2017, Stockholm, Sweden, pp. 1089–1093 (2017)
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
23. Cui, Z., Ke, R., Pu, Z., Wang, Y.: Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. *CoRR arXiv preprint arXiv:1801.02143* (2018)
24. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2014). <http://arxiv.org/abs/1412.6980>
25. Zhou, X., Zhu, X., Nakamura, K., Mahito, N.: ECG quality assessment using 1D-convolutional neural network. In: 2018 14th IEEE International Conference on Signal Processing (ICSP), pp. 780–784, IEEE (2018). <https://doi.org/10.1109/ICSP.2018.8652479>
26. Huerta, A., Martínez-Rodrigo, A., González, V.B., Quesada, A., Rieta, J.J., Alcaraz, R.: Quality assessment of very long-term ecg recordings using a convolutional neural network. In: 2019 E-Health and Bioengineering Conference (EHB), pp. 1–4 (2019). <https://doi.org/10.1109/EHB47216.2019.8970077>

27. Zhu, Z., Liu, W., Yao, Y., Chen, X., Sun, Y., Xu, L.: Adaboost based ECG signal quality evaluation. In: 2019 Computing in Cardiology (CinC), pp. 1–4 (2019). <https://doi.org/10.23919/CinC49843.2019.9005515>
28. Zhang, Q., Fu, L., Gu, L.: A cascaded convolutional neural network for assessing signal quality of dynamic ECG. *Comput. Math. Methods Med.* **2019**, 7095137 (2019)