2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Signal Quality Assessment of PPG Signals using STFT Time-Frequency Spectra and Deep Learning Approaches

Jianzhong Chen[1,2,3], Ke Sun[1], Yi Sun[1] and Xinxin Li[1,2,3,]*

***Abstract*—Photoplethysmography (PPG) is an important signal which contains much physiological information like heart rate and cardiovascular health etc. However, PPG signals are easily corrupted by motion artifacts and body movements during their recordings, which may lead to poor quality. In order to accurately extract cardiovascular information, it is necessary to ensure high PPG quality in these applications. Although there are several existed methods to get the PPG signal quality, those algorithms are complex and the accuracies are not very high. Thus, this work proposes a deep learning network for the signal quality assessment using the STFT time-frequency spectra. A total of 5804 10s signals are preprocessed and transformed into 2D STFT spectra with 250 × 334 pixels. The STFT figures are as the input of the CNN networks, and the model gives the result as good or bad quality. The model accuracy is 98.3% with 98.9% sensitivity, 96.7% specificity, and 98.8% F1-score. And the heart rate error is much reduced after classification with the reference of ECG signals. Thus, the proposed deep learning approaches can be useful in the classification of good and bad PPG signals. As far as we know, this is the first article using deep learning methods combined with STFT time-frequency spectra to get the signal quality assessment of PPG signals.**

I. Introduction

The photoplethysmography (PPG) is an important physiological signal which can provide heart rate (HR) [1], blood oxygen saturation [2] directly, and also obtain blood pressure [3], respiration rate [4], arterial stiffness information, vascular aging [5], and biometric recognition.

The quality of PPG will affect the accuracy of parameter extraction. Take heart rate as an example, poor quality signals affect HR acquisition. As shown in Fig.1, the maximum frequency peak of PPG signal is typically considered as the HR after removing the DC part. The maximum frequencies in both PPG and ECG spectrums in Fig.1 (a) are 1.4 Hz, which leads to 84 bpm in HR estimation. However, the PPG signal in Fig.1 (b) is corrupted by motion artifacts and its maximum frequency is 0.6Hz in the PPG spectrum rather than 1.4 Hz in the ECG spectrum, which leads to the wrong result in HR estimation. In more rigorous conditions, such as respiratory signal extraction or arterial stiffness diagnosis, more attention is paid to the quality of the waveform.. In those conditions, many parameters are extracted from the PPG waveforms. These papers also show the signal quality is important when extracting respiration rate [6].

Several methods have been proposed for PPG signal quality detection. However, the accuracy is not very high, and most of the methods require complex algorithm designs or feature extractions. J. A. Sukor et al. proposed an automatic rejection method for artifact-contaminated PPG waveforms based on waveform morphology analysis and a decision tree classifier [7]. Q Li et al. introduced a dynamic time warping (DTW) method to stretch the single-cycle pulse waveform to match a running template and combine it with several other signal quality features, which are then presented to a multilayer perceptron (MLP) neural network to indicate good and bad quality [8]. R Couceiro et al. proposed a motion artifact detection algorithm based on the variation of time-domain and period-domain features of PPG signals and using a support vector machine model to distinguish between clean and corrupted signals [9]. S. Cherif et al. described a new method based on waveform morphology for detecting artifacts in PPG signals using a random distortion test to perform adaptive thresholding [10]. C. H. Goh et al. designed a 1-D Convolutional Neural Networks (CNN) model to classify five-second PPG segments into clean or artifact-affected segments [11]. S.-H. Liu et al. evaluated the qualities of PPG signals from a wearable forehead pulse oximeter with CNN and SVM approaches and achieved lower error ratios of oxygen saturation ratio (SpO2) after classification [12].

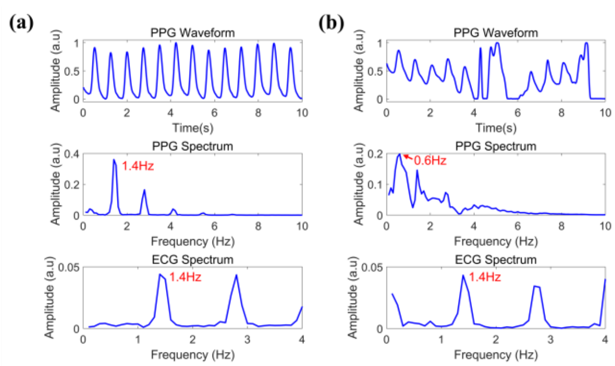

Figure 1. Different PPG signal qualities for HR estimation.

Deep learning has gained widespread attention due to its powerful ability to automatically learn from data. CNN was first proposed by LeCun et al. [13] and was developed through a project to recognize handwritten digits. With the advent of CNN models, correlations of spatially adjacent pixels can be extracted by applying nonlinear filters, and various local

This research is supported by Shanghai Municipal Science and Technology Commission (project 18dz1100600).

Jianzhong Chen, Ke Sun, Yi Sun and Xinxin Li are with the State Key Laboratory of Transducer Technology, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China (*Correspondence: xxli@mail.sim.ac.cn; Tel.: +86-21-6213-1794).

Jianzhong Chen and Xinxin Li are with School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China.

Jianzhong Chen and Xinxin Li are with Unversity of Chinese Academy of Sciences, Beijing 100049, China.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

features of images can be extracted. In recent years, deep learning has been successfully applied to the biomedical field. CNN methods have been widely used in the ECG classification of arrhythmia [14], the blood pressure and heart rate estimation from PPG signals [3], and the heart rate extraction and biometric Identification from PPG signals in the ambulant environment [15] et al.

Consider the limitations of the above methods, this paper proposes a deep learning network for the signal quality assessment using STFT time-frequency spectra. Both of the time and frequency domain information are contributed to getting signal quality using the STFT images. The 1D PPG signals are first transformed into 2D STFT images, and the CNN model takes those images as the inputs and gives the classification result of good or bad quality. The model is validated on a dataset of 102 people and performed with high accuracy.

II. Materials and Methods

The whole processing flow is described as follows. The PPG data are from 102 individuals. Each PPG signal is 600s, the PPG signals are then divided into 10s segments. Then we convert the one-dimensional PPG signal into a two-dimensional STFT spectrum and puts it into the CNN network. Finally, the model divides the corresponding signal quality into good or poor.

A. Data

The data sources, 102 individuals in total, were derived from the arterial data (ABP) of the VitalDB database, which is an open-access public dataset of intraoperative vital signs and biosignals collected by the Department of Anesthesia at Seoul National University Hospital using the Vital Recorder program [16]. The collected PPG waveforms were measured from the finger by a device called SNUADC with a sampling rate of 100 Hz.

B. Signal quality annotation

Each PPG segment was labeled as ‘Good’, ‘Bad’, or ‘Not sure’ manually by the expert engineers. The total annotation uses ECG signals as a reference. A good PPG signal is defined as: (1). the PPG signal has a clear and undisturbed waveform. (2). the reflection points of the waveform are relatively consistent. (3). the heart rate information is consistent with the corresponding ECG signals.

As there are few PPG signals with ‘Not sure’ quality, therefore, the signals of this category are discarded. The remaining signals are divided into good quality (1) and poor quality (0). Finally, a total of 5804 segments of signals are selected, of which 3969 segments are high-quality signals, and 1835 segments are low-quality signals. The distribution of the data set uses 80% of individual data (82 individuals) as the training set and 20% of the data (20 individuals) as the test set.

C. Short-time Fourier transform (STFT)

STFT has been widely used in the field of digital signal processing. Compared with the classical FFT theory that loses time-domain information, STFT uses a suitable window size to obtain the corresponding time-frequency information. When analyzing a non-smooth signal, it is assumed to be approximately smooth over a finitely supported range of time windows [14]. For a discretized digital signal, the mathematical formulation is shown in Equation 1.

$$STFT\{x[n]\} = X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]\omega[n-m]e^{-j\omega n} \quad (1)$$

where x[n] denotes the PPG signal with a sampling rate of 100 Hz and w[n] is the window function. In this paper, a Hanning window with a window size of 256 is used.

Since the effective frequency range of the PPG signal is 0 ~ 10 Hz, and most of the energy is within 0 ~ 5 Hz, the frequency spectrum of the STFT selected in this paper is 0 ~ 8 Hz. Fig.2 (a) is a clean and high-quality PPG signal and its STFT spectrum, Fig.2 (b) is a partially polluted PPG signal with low quality and its STFT spectrum. The STFT transformation is processed by MATLAB 2019a and then resized to 250 × 334 pixels.

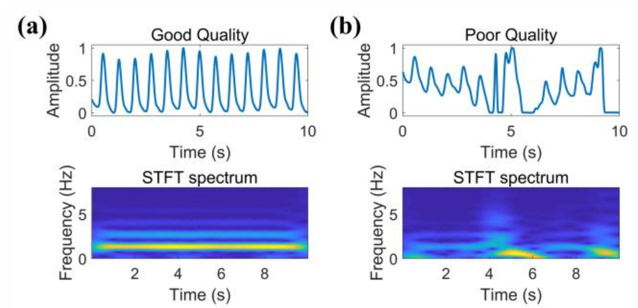

Figure 2. Different PPG signal quality and STFT Spectra.

D. Deep learning methods

The proposed CNN architecture for PPG quality assessment is shown in Fig.3. The input of the network is 2D STFT of PPG signals at 250 × 334 pixels. The convolutional layers are used for extracting features from the STFT images. The pooling layers are placed behind the convolutional layer to reduce the size of the feature maps. The dropout layers are used for avoiding overfitting and the activation function is RELU (Rectified Linear Unit). Finally, a full connection (FC) layer with a sigmoid activation function is used to output the probability of good or bad classes. The code is available on https://github.com/shangjianshizhe/PPGSQI.

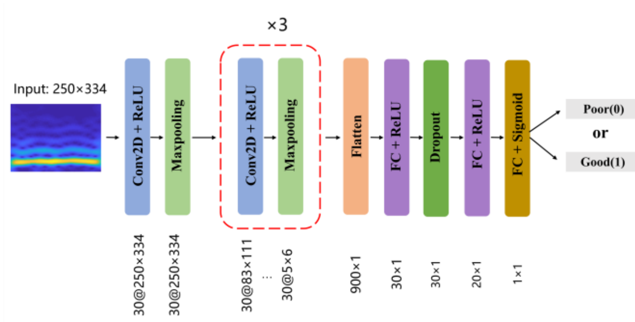

Figure 3. The proposed network.

The model is built on the Keras 2.3.1 platform, the running background is Tensorflow 1.15.0 and Python 3.7.6 is used for programming in the Jupyter notebook. Finally, the model runs on an AMD Ryzen5 2400G computer with 32G DDR4 memory, the operating system is win10, and the graphics card is GTX-1660 with 6G memory for acceleration. The model uses the adam optimizer and binary cross-entropy as the loss function. The training epochs are 90, and the batch size is 32 with a learning rate of 0.00005.

III. RESULT

A. Evaluation Metrics

Many metrics are used to evaluate the proposed method [11]. Accuracy, sensitivity, specificity and F1 scores are calculated as defined in equations (2) - (5). The receiver operating characteristic (ROC) curves and the area under the receiver operating characteristic curve (AUC) are also considered.

Accuracy:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Sensitivity:

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

Specificity:

$$Sp = \frac{TN}{FP+TN} \quad (4)$$

F1-score:

$$F1 = \frac{2TP}{2TP+FP+FN} \quad (5)$$

where TP stands for true positive; TN stands for true negative; FP stands for false positive; FN represents false negative.

B. Test Results

The model is validated in 1131 fragments, and the test results are shown in Table I. Table I presents accuracy, sensitivity, specificity, F1-score and support values of the two classes. Fig. 4 shows the confusion matrix of the proposed model. Of all the results, 1112 are considered correct and 19 are misclassified. The ROC is shown in Fig. 5. The AUC of 0.997 represents a reliable performance.

TABLE I. TEST RESULTS

Metrics	Se	Sp	F1-score	ACC	Support
Values	98.9%	96.7%	98.8%	98.3%	1131

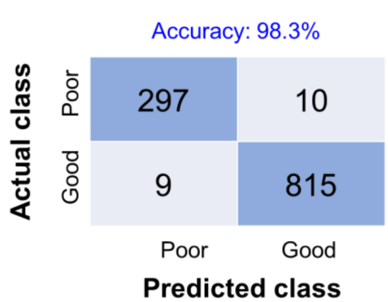

Figure 4. The confusion matrix.

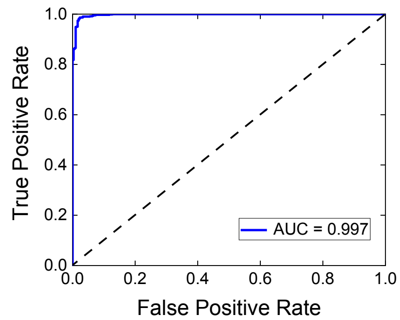

Figure 5. The ROC curve.

C. HR performance

The FFT-based heart rate estimation algorithm was used by finding the maximum peak of the spectrum within 2 Hz. Prior to classification, the HRs of 1131 PPG segments are reported as 79.7 ± 16.3 (mean ± standard) BPM (beats per minute). After classification using the proposed model, the PPG segments were re-evaluated using the same algorithm; the heart rates of good quality segments are 81.7 ± 13.0 BPM while heart rates of poor quality segments are 74.7 ± 22.1 BPM.

The errors before classification are 6.18 ± 0.29 (mean ± std) BPM compared to the HRs of the ECG signals. After model classification, the errors for the good PPG segments are 3.20 ± 0.21 BPM, while the errors for the poor PPG segment are 14.07 ± 0.35 BPM. Details are listed in Table II. The bolded sections show the lower error and standard deviation values, indicating that classification as clean PPG segments yielded more consistent results in heart rate assessment.

TABLE II. HR PERFORMANCE

Parameter	Unclassified	Good	Poor
PPG HR (BPM)	79.7 ± 16.3	**81.7 ± 13.0**	74.7 ± 22.1
Error (BPM)	6.18 ± 0.29	**3.20 ± 0.21**	14.07 ± 0.35

D. Compare with other baseline methods

The method also compares with several machine learning algorithms. In traditional machine learning models often use Histogram of Oriented Gradients (HOG) to extract information from pedestrian trajectories or ECG images, the method is chosen in this paper to extract image features from STFT images of PPG and train three classifiers for PPG quality evaluation. In this paper, three baseline algorithms, multilayer perceptron (MLP), support vector machine (SVM), and random forest (RF) are also applied. A total of 1764 HOG features were extracted from the STFT spectrum of PPG and put into the machine learning algorithms of MLP, RF, and SVM. HOG feature extraction was processed by MATLAB 2019a. For a better performance comparison, we also applied a 1D CNN network, which uses the original PPG waveform as input to the model.

The performance comparisons are shown in Table III and all the optimal parameters are bolded. As can be seen from the Table, comparing with the machine learning algorithms such as MLP, RF, and SVM, the 2D CNN model proposed in this paper has much better sensitivity, specificity, F1-score, and accuracy. The performance of the 2D CNN combined with the STFT approach also improved slightly on the 1D CNN, which proves that the method proposed in this paper is very effective.

TABLE III. PERFORMANCE COMPARISONS WITH DIFFERENT METHODS

Methods	Se	Sp	F1-score	ACC
MLP + HOG	91.5%	83.1%	92.5%	89.2%
RF + HOG	91.4%	87.6%	93.3%	90.4%
SVM + HOG	93%	68.1%	90.8%	86.2%
1D CNN	97.1%	93.5%	97.6%	96.6%
2D CNN + STFT	**98.9%**	**96.7%**	**98.8%**	**98.3%**

E. *Compare with Other works*

Table IV shows the comparison results with other methods, it can be seen that the proposed model in this paper has higher accuracy, and its value can reach 98.3%, while the corresponding value of other references is 83% - 95.2%. In addition, the sensitivity and specificity are higher than most methods. What's more, the model in this paper is verified on the data of 102 people and 600s per person, which will be slightly larger than the data in most papers. In Ref. [8], the accuracy is 95.2%, but the data size is 1055 segments.

TABLE IV. PERFORMANCE COMPARISONS WITH OTHER WORKS

Papers	Year	Dataset	Subjects	Se	Sp	ACC
Ref. [7]	2011	104, 60s	13	89%	77%	83%
Ref. [8]	2012	1055, 6s	104	**99%**	80.6%	95.2%
Ref. [9]	2014	/	15	84.3%	91.5%	88.5%
Ref. [10]	2016	104, 60s	/	84%	83%	83%
Ref. [12]	2020	12876, 7s	20	91.8%	87.3%	89.9%
Proposed	2021	5804, 10s	102	98.9%	**96.7%**	**98.3%**

IV. DISCUSSION AND CONCLUSION

As far as we know, this is the first paper to use deep learning methods combined with STFT time-frequency spectra for signal quality assessment of PPG signals. The CNN model takes STFT images as input and gives good or bad quality results. The proposed method is validated on a dataset of 102 people and 600 seconds of PPG data for each person. The model achieved good results with an accuracy of 98.3%, a sensitivity of 98.9%, a specificity of 96.7% and an F1 score of 98.8%.

The model has better performance than machine learning algorithms such as MLP, RF and SVM, as well as 1D CNN. The model also has higher accuracy compared to the algorithms proposed in other papers. In terms of HR evaluation, the standard deviation of the high-quality PPG signals is lower than that of the poor-quality signal. The HR error is also much lower with reference to the ECG signal. This proves that the proposed network architecture can help researchers to obtain reliable estimates of physiological parameters from PPG signals. More importantly, the method can also be applied to similar time series conditions, such as ECG signals and arterial blood pressure (ABP) signals.

ACKNOWLEDGMENT

This research is supported by Shanghai Municipal Science and Technology Commission (project 18dz1100600). We are really grateful for the open source contributions of the VitalDB workers.

REFERENCES

[1] Temko, "Accurate Heart Rate Monitoring during Physical Exercises Using PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2016–2024, 2017.

[2] Wan-Young Chung, Young-Dong Lee, and Sang-Joong Jung, "A Wireless Sensor Network Compatible Wearable U-healthcare Monitoring System Using Integrated ECG, Accelerometer and SpO2," *2008 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 1529–1532, 2008.

[3] M. Panwar, A. Gautam, D. Biswas, and A. Acharyya, "PP-Net: A Deep Learning Framework for PPG-Based Blood Pressure and Heart Rate Estimation," IEEE Sens. J., vol. 20, no. 17, pp. 10000–10011, 2020.

[4] H. Sharma, "Extraction of respiration from PPG signals using Hilbert vibration decomposition," *ACM Int. Conf. Proceeding Ser.*, pp. 48–52, 2019.

[5] R. M. Rozi, S. Usman, M. A. Mohd Ali, and M. B. I. Reaz, "Second derivatives of photoplethysmography (PPG) for estimating vascular aging of atherosclerotic patients," *2012 IEEE-EMBS Conf. Biomed. Eng. Sci. IECBES 2012*, no. December, pp. 256–259, 2012.

[6] P. H. Charlton *et al.*, "Breathing Rate Estimation from the Electrocardiogram and Photoplethysmogram: A Review," *IEEE Rev. Biomed. Eng.*, vol. 11, pp. 2–20, 2018.

[7] J. A. Sukor, S. J. Redmond, and N. H. Lovell, "Signal quality measures for pulse oximetry through waveform morphology analysis," *Physiol. Meas.*, vol. 32, no. 3, pp. 369–384, 2011.

[8] Q. Li and G. D. Clifford, "Dynamic time warping and machine learning for signal quality assessment of pulsatile signals," *Physiol. Meas.*, vol. 33, no. 9, pp. 1491–1501, 2012.

[9] R. Couceiro, P. Carvalho, R. P. Paiva, J. Henriques, and J. Muehlsteff, "Detection of motion artifact patterns in photoplethysmographic signals based on time and period domain analysis," *Physiol. Meas.*, vol. 35, no. 12, pp. 2369–2388, 2014.

[10] S. Cherif, D. Pastor, Q. T. Nguyen, and E. L'Her, "Detection of artifacts on photoplethysmography signals using random distortion testing," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2016-Octob, pp. 6214–6217, 2016.

[11] C. H. Goh, L. K. Tan, N. H. Lovell, S. C. Ng, M. P. Tan, and E. Lim, "Robust PPG motion artifact detection using a 1-D convolution neural network," *Comput. Methods Programs Biomed.*, vol. 196, p. 105596, 2020.

[12] S.-H. Liu, H.-C. Liu, W. Chen, and T.-H. Tan, "Evaluating Quality of Photoplethymographic Signal on Wearable Forehead Pulse Oximeter With Supervised Classification Approaches," *IEEE Access*, vol. 8, no. Dc, pp. 185121–185135, 2020.

[13] R. E. H. Y. LeCun, B. Boser, J. S. Denker, D. Henderson and and L. D. J. W. Hubbard, "Backpropagation applied to handwritten zip code recognition," *Neural Comput*, vol. 1, pp. 541–551, 1989.

[14] J. Huang, B. Chen, B. Yao, and W. He, "ECG Arrhythmia Classification Using STFT-Based Spectrogram and Convolutional Neural Network," *IEEE Access*, vol. 7, pp. 92871–92880, 2019.

[15] D. Biswas et al., "CorNET: Deep Learning framework for PPG based Heart Rate Estimation and Biometric Identification in Ambulant Environment," IEEE Trans. Biomed. Circuits Syst., vol. 4545, no. c, pp. 1–1, 2019.

[16] H. C. Lee and C. W. Jung, "Vital Recorder- A free research tool for automatic recording of high-resolution time-synchronised physiological data from multiple anaesthesia devices," *Sci. Rep.*, vol. 8, no. 1, pp. 1–8, 2018.