

Document downloaded from:

<http://hdl.handle.net/10251/180560>

This paper must be cited as:

Huerta, A.; Martínez-Rodrigo, A.; Bertomeu González, V.; Quesada, A.; Rieta, JJ.; Alcaraz, R. (2019). Quality Assessment of Very Long-Term ECG Recordings Using a Convolutional Neural Network. IEEE. 1-4. <https://doi.org/10.1109/EHB47216.2019.8970077>



The final publication is available at

<https://doi.org/10.1109/EHB47216.2019.8970077>

Copyright IEEE

Additional Information

---

# Quality Assessment of Very Long-Term ECG Recordings Using a Convolutional Neural Network

A. Huerta<sup>1</sup>, A. Martínez-Rodrigo<sup>1</sup>, V. Bertomeu González<sup>2</sup>, A. Quesada<sup>3</sup>, J.J. Rieta<sup>4</sup>, R. Alcaraz<sup>1</sup>

<sup>1</sup>Research Group in Electronic, Biomedical and Telecommunication Engineering, University of Castilla-La Mancha, Cuenca, Spain, {alvaro.huerta, arturo.martinez, raul.alcaraz}@uclm.es

<sup>2</sup>Cardiology Department, University Hospital San Juan, Alicante, Spain, vbertog@gmail.com

<sup>3</sup>Cardiac Arrhythmia Department, General University Hospital Consortium of Valencia, Spain, quesada\_aur@gva.es

<sup>4</sup>BioMIT.org, Electronic Engineering Depart., Universitat Politècnica de Valencia, Valencia, Spain, jjrieta@upv.es

**Abstract**—The electrocardiogram (ECG) is a physiological signal highly sensitive to disturbances during its acquisition. To palliate this issue, many works have described preprocessing algorithms operating in 12-lead, short-term ECG recordings. However, only a few methods have been introduced to detect noisy segments in single-lead, long-term ECG signals, this being a pending challenge to be resolved. Hence, this work proposes a novel technique to automatically detect low-quality segments in single-lead, long-term ECG recordings. The method is based on the high learning capability of a convolutional neural network (CNN), which has been trained with 2D images obtained when turning ECG recordings into scalograms using a continuous Wavelet transform (CWT). To validate the method, a publicly available dataset containing single-lead, long-term ECG intervals from patients with different cardiac rhythms has been used. These signals have been annotated by experts, who identified noisy intervals and those with sufficient quality to be clinically interpreted. The results have shown that the proposed method discriminates correctly between low and high-quality ECG segments with an accuracy greater than 90%, and with sensitivity slightly larger than specificity.

**Keywords**—*Electrocardiogram; Convolutional Neural Network; Noise; Quality Assessment*

## I. INTRODUCTION

In general terms, most physiological signals are highly sensitive to disturbances and noise, which makes their subsequent analysis and interpretation difficult [1]. This limitation is especially relevant for the particular case of electrocardiographic (ECG) recordings, because the presence of perturbations can hinder accurate detection and diagnosis of most cardiovascular disorders, which provokes about 30% of all deaths worldwide [2], [3]. Although a wide variety of algorithms to reduce noise and interferences can be nowadays found in the scientific literature, they are unable to successfully work in every context. For instance, this is the case of long-term ECG recordings, because they are often acquired within dynamic and ever-changing environments, and nuisance signals are then extensively stronger than for resting ECG signals [4].

Hence, to take advantage of current wearable devices which enable very long-term monitoring of a patient's daily life [5], automatic identification of low and high-quality ECG intervals plays a pivotal role [6]. In this way, real-time analysis of only noise-free segments could be analyzed.

In last years, some authors have proposed several methods to assess quality of ambulatory ECG recordings. To this respect, an interesting and complete review summarizing these approaches can be found in [6]. Briefly, all methods are divided into five categories, i.e.: (1) those based on extracting fiducial features and applying heuristic rules; (2) those based on detecting fiducial features and applying machine learning algorithms; (3) those working with non-fiducial features and applying heuristic rules; (4) those working with non-fiducial features and applying machine learning algorithms; and finally (5) those based on filtering. Nonetheless, these algorithms are characterized by estimating ECG global quality from the 12 standard leads. Thus, even when only a few leads do not present sufficient quality for their interpretation and analysis, the whole 12-lead ECG interval is discarded [6].

Although these algorithms have shown a good performance when dealing with 12-lead, short-term ECG recordings [6], their ability to deal with long-term signals acquired from wearable systems, where the number of available leads is restricted, has not been demonstrated. Hence, a novel algorithm able to discriminate between high and low-quality segments in single-lead, long-term ECG recordings is proposed in this work. The method is based on the high learning capability of the convolutional neuronal networks (CNNs) for discriminating between those ECG intervals fully noise-free and those segments where artifacts are significant enough to mask partially or completely clinical information. It should be noted that CNNs have experienced a great evolution, such that a number of methodologies have been developed with the aim of transforming and processing a wide range of physiological variables analyzed from different points of view [7].

Nevertheless, although these CNNs have shown inadequate characteristics for dealing with 12-leads ECG recordings [8], they could be relevant when single-lead ECG signals are used. Indeed, previous works have demonstrated the capability of several CNNs to discriminate between single-lead ECG recordings with different characteristics [7].

## II. MATERIALS AND METHODS

### A. Database

To ensure reproducibility of the proposed methodology, a freely available database from *PhysioNet* was used. More precisely, the *PhysioNet/Challenge 2017* [9] dataset was chosen due to two reasons. On the one hand, this database contains a set of single-lead ECG recordings acquired from a modern heart monitoring device. On the other hand, although the database has not been specifically created for ECG quality assessment, it contains annotations from different clinicians about whether an ECG signal has sufficient quality to identify the patient's rhythm. Indeed, the signals were labeled into four groups: (1) normal sinus rhythm, (2) atrial fibrillation, (3) other rhythms and (4) noisy, such that the first three groups formed the "acceptable quality" set and the remaining one constituted the "unacceptable quality" group. Precisely, this grouping was later used to obtain the two classes used in the validation method.

Another interesting feature of this database consists of the large number of ECG signals, which are divided into two subsets for testing the methodology. Thus, while the first set is proposed for training, the another one is given for validation, as shown in Table I. However, it is worth noting the recordings present a different duration. To deal with this inconvenient and accurately limit noisy segments, all the recordings were segmented into intervals with the same duration of 10 seconds. In total, 25.834 ECG segments of acceptable quality and 506 of unacceptable quality were available in the training stage. Similarly, 831 and 83 intervals of acceptable and unacceptable quality, respectively, were used in the validation stage.

### B. ECG quality assessment

The proposed algorithm to discern between low and high-quality ECG intervals is based on three steps, such as Figure 1 shows. In the first one, the ECG segment is preprocessed using classic methodologies to remove as much noise as possible and, thus, facilitate the subsequent classification. More precisely, baseline wander (BW) is removed following the algorithm proposed by Meyer et al., where the BW is estimated and then subtracted from the signal without adding phase distortion [10]. Similarly, powerline and high frequency interferences are removed making use of a novel wavelet shrinkage denoising approach based on stationary Wavelet transform [11].

In the next step, the preprocessed ECG segment is transformed into a two dimensions and three channels color image. This transformation is based on the use of Continuous Wavelet

TABLE I. DATABASE DESCRIPTION: NUMBER AND AVERAGE LENGTH OF THE RECORDINGS INCLUDED IN THE PHYSIONET/CHALLENGE 2017 DATABASE.

| Class               | Training set | Validation set | Average length(s) |
|---------------------|--------------|----------------|-------------------|
| Normal sinus rhythm | 5154         | 148            | 31.9              |
| AF                  | 771          | 47             | 31.6              |
| Other rhythms       | 2557         | 65             | 34.1              |
| Noisy               | 279          | 40             | 27.1              |
| <b>Total</b>        | <b>8761</b>  | <b>300</b>     | <b>32.5</b>       |

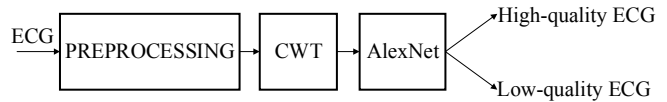


Figure 1. Block diagram describing the operation of the proposed method to discern between high and low-quality ECG intervals.

Transform (CWT), which applies a bank of fixed-bandwidth filters to obtain time-frequency information of the ECG signal [12]. In fact, the resulting picture is known as scalogram and represents absolute value of wavelet coefficients as a function of time and frequency. It should be noted that CWT has been here used because, unlike other time-frequency transformations, it presents a superior ability to accurately detect local, transient and intermittent aperiodicities in non-stationary signals [13]. As an example, Figure 2 shows the existing differences between the scalograms obtained from two common ECG segments, with acceptable and unacceptable quality, respectively. As can be observed, the scalogram from the acceptable ECG segment presents a characteristic pattern, which is repeated along the time. On the other hand, the scalogram from the unacceptable ECG segment shows a picture with a high degree of randomness and without an identifiable pattern.

Finally, in the last stage of the proposed methodology a CNN, known as AlexNet, is used to discriminate between low and high-quality ECG segments. Although a wide variety of CNNs can be nowadays found in the literature [14], AlexNet has been chosen in this work because it is previously pre-trained, thus minimizing computational cost for its specific training in identifying noisy ECG intervals. This feature is also essential in the context of this work, since the database under study does not allow to train adequately a CNN from the beginning. Nonetheless, it is worth noting that the use of pre-trained neural networks in conceptually similar tasks is a current common practice, which is known as transfer learning [15]. Nevertheless, it should be noted that in this work AlexNet software architecture has been modified to adequate their layer structure to deal with two classes. More precisely, the layers conforming the original architecture were designed to discern among 1000 common objects and have been here reprogrammed to deal exclusively with ECG recordings of two

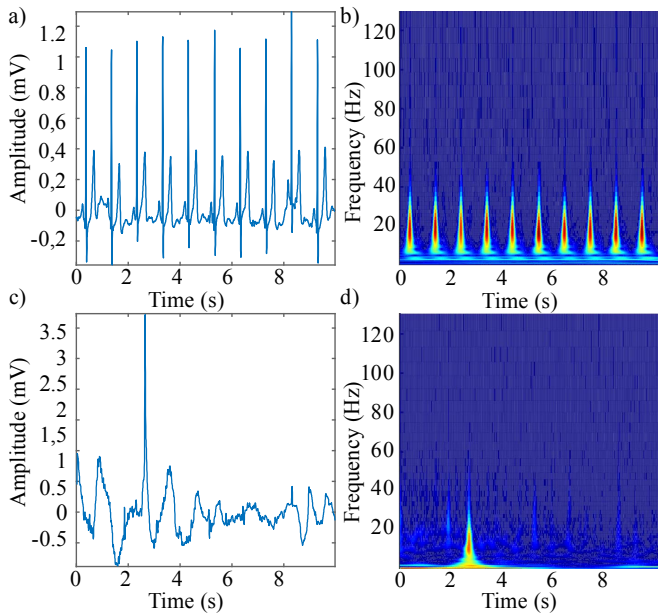


Figure 2. Typical ECG intervals with high (top) and low (bottom) quality. Each signal is represented together with its scalogram obtained via CWT.

classes, i.e., those with acceptable and unacceptable quality, respectively [16].

### III. RESULTS

Given the existing imbalance between high and low-quality ECG segments described in Section II.A for the training set, they were rearranged to obtain several stratified and balanced training sets, which ensured a robust and unbiased validation of the proposed algorithm. Thus, 10 stratified groups with a similar proportion of both types of signals were chosen. More precisely, each group was composed by 500 (out of the 506) low-quality ECG segments and 500 high-quality segments, randomly chosen from the 25.834 available ones. Note that high-quality ECG intervals were not repeated in different groups.

Next, the proposed method was exposed to 10 training cycles, employing in each case a different group. The resulting classification model in each iteration was then tested using the validation set described in Section II.A. The results were computed in terms of sensitivity, specificity and accuracy of the model under study in each case [17]. In this context, sensitivity (Se) was defined as the rate of low-quality ECG segments correctly classified. Similarly, specificity (Sp) was defined as the percentage of high-quality ECG segments properly identified. Finally, accuracy (Acc) was computed as the total proportion of ECG segments rightly detected. The values obtained on the validation set for each training/testing iteration, as well as mean value, are shown in Table II. As can be seen, sensitivity was always 100%, but specificity was around 90% in most cases.

TABLE II. CLASSIFICATION RESULTS FOR EACH TRAINING/TESTING ITERATION OF THE PROPOSED ALGORITHM. THE RESULTS ARE EXPRESSED IN TERMS OF SENSITIVITY (SE), SPECIFICITY (SP) AND ACCURACY (ACC). COMPUTATIONAL TIME FOR CLASSIFICATION OF THE VALIDATION SIGNALS IN EACH ITERATION IS ALSO INCLUDED (T).

| Iteration         | Se. | Sp.   | Acc.  | T (s) |
|-------------------|-----|-------|-------|-------|
| Group #1          | 1   | 0.964 | 0.967 | 17.48 |
| Group #2          | 1   | 0.912 | 0.920 | 17.22 |
| Group #3          | 1   | 0.908 | 0.917 | 16.80 |
| Group #4          | 1   | 0.924 | 0.931 | 17.57 |
| Group #5          | 1   | 0.899 | 0.908 | 18.01 |
| Group #6          | 1   | 0.882 | 0.893 | 18.06 |
| Group #7          | 1   | 0.877 | 0.888 | 17.87 |
| Group #8          | 1   | 0.876 | 0.887 | 16.92 |
| Group #9          | 1   | 0.911 | 0.919 | 16.38 |
| Group #10         | 1   | 0.876 | 0.887 | 16.60 |
| <b>Mean value</b> | 1   | 0.903 | 0.912 | 17.29 |

Finally, remark that Table II also presents the time spent by the proposed method to classify the signals included in the validation set for each training iteration. In all cases, this time ranged from 16 to 18 seconds, when experiments were conducted on a PC with a 3.40 GHz Intel i7 processor, 16 GB of RAM, and running Windows 10 operating system.

### IV. DISCUSSION

A new methodology able to discern between high and low-quality single-lead, long-term ECG recordings has been presented in this work, obtaining a global accuracy greater than 90%. Moreover, the method has also proven ability to correctly identify 100% of the low-quality segments, such that they could be rejected, and any misinterpretation of their clinical information could also be prevented. On the contrary, the method has presented a slightly more limited capability of correctly identifying high-quality segments. To this respect, about 10% of the high-quality ECG intervals have been inappropriately considered as noisy ones. Although this circumstance means a limitation that could be corrected in future versions of the method, in the specific context of long-term ECG monitoring, it would not avoid the possibility of improving the diagnosis and follow-up of some patients suffering from cardiovascular disorders. In fact, a lot of information would be still available for real-time analysis.

Nonetheless, it is mandatory to note that this method has obtained a similar global performance compared with other previous works. Thus, Satija et al. developed a method based on decision-making trees, obtaining an accuracy of 89% when noisy ECG segments acquired from a wearable device were analyzed [18]. However, that method was designed to work in a context of 12-lead ECG recordings and, moreover, it was not validated on a realistic environment. Indeed, clean ECG signals were disturbed by artificial noise synthetically generated. On the other hand, Zhao et al. have recently presented a work with

a similar procedure [4], but intended for dealing with 12-lead ECG recordings. Despite much more information is available when 12-lead ECG signal are considered, that method only reached a global accuracy about 85%.

Another relevant difference regarding previous works, is that the proposed method has been validated considering a more realistic context. Thus, besides intervals from normal sinus rhythm, ECG segments with cardiac rhythm disorders have been considered. The presence of arrhythmias and other rhythm alterations is very common when patients suffering from cardiac diseases are long-term monitored [19]. Moreover, although a more detailed analysis must be performed, the presence of atrial fibrillation and other rhythm alterations does not seem to cause specificity reduction noticed when using this algorithm. Indeed, among the high-quality ECG segments incorrectly classified as noisy ones, the same proportion of all ECG rhythms (i.e., sinus rhythm, atrial fibrillation and other rhythms) was found.

Finally, remark that although the total duration of the segments included in the validation set was approximately 150 minutes, once the proposed method was trained, it only needed around 17 seconds to process all information. This observation suggests that the method could be implemented in wearable devices and process the information in real time, thus offering the possibility of only recording and processing ECG segments with enough quality. In this way, energy consumption and memory could be minimized.

## V. CONCLUSION

A method able to discern between high and low-quality intervals in single-lead, long-term ECG recordings with an accuracy higher than 90% has been proposed for first time in this work. Despite using much less information, its accuracy is comparable to other methods dealing with 12-lead ECG signals. Moreover, the method also offers the possibility of working in real time. Nevertheless, it presents a slightly lower specificity than sensitivity, so that future versions will have to address this limitation.

## ACKNOWLEDGMENT

This work has been funded by projects DPI2007-83952-C3 from MINECO/AEI/FEDER EU, SBPLY/17/180501000411 from “Junta de Castilla la Mancha”, AICO/2019/036 from “Generalitat Valenciana”, and grand 2018/11744 from European Regional Development Fund (FEDER, UE).

## REFERENCES

[1] Singh, B.N.; Tiwari, A.K. Optimal selection of wavelet basis function applied to ECG signal denoising. *Digital Signal Processing* **2006**, *16*, 275-287.

[2] Nichols, M.; Townsend, N.; Scarborough, P.; Rayner, M. Cardiovascular disease in Europe: epidemiological update. *European Heart Journal* **2013**, *34*, 3028-3034.

[3] Berkaya, S.K.; Uysal, A.K.; Gunal, E.S.; Ergin, S.; Gunal, S.; Gulmezoglu, M.B. A survey on ECG analysis. *Biomedical Signal Processing and Control* **2018**, *43*, 216-235.

[4] Zhao, Z.; Liu, C.; Li, Y.; Li, Y.; Wang, J.; Lin, B.-S.; Li, J. Noise Rejection for Wearable ECGs Using Modified Frequency Slice Wavelet Transform and Convolutional Neural Networks. *IEEE Access* **2019**, *7*, 34060-34067.

[5] Nagai, S.; Anzai, D.; Wang, J. Motion artefact removals for wearable ECG using stationary wavelet transform. *Healthcare Technology Letters* **2017**, *4*, 138-141.

[6] Satija, U.; Ramkumar, B.; Manikandan, M.S. A review of signal processing techniques for electrocardiogram signal quality assessment. *IEEE Reviews in Biomedical Engineering* **2018**, *11*, 36-52.

[7] Faust, O.; Hagiwara, Y.; Hong, T.J.; Lih, O.S.; Acharya, U.R. Deep learning for healthcare applications based on physiological signals: A review. *Computer Methods and Programs in Biomedicine* **2018**, *161*, 1-13.

[8] Jambukia, S.H.; Dabhi, V.K.; Prajapati, H.B. Classification of ECG signals using machine learning techniques: A survey. In Proceedings of 2015 International Conference on Advances in Computer Engineering and Applications; pp. 714-721.

[9] Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.-K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, e215-e220.

[10] Meyer, C.; Keiser, H. Electrocardiogram baseline noise estimation and removal using cubic splines and state-space computation techniques. *Computers and Biomedical Research* **1977**, *10*, 459-470.

[11] García, M.; Martínez-Iniesta, M.; Ródenas, J.; Rieta, J.J.; Alcaraz, R. A novel wavelet-based filtering strategy to remove powerline interference from electrocardiograms with atrial fibrillation. *Physiological Measurement* **2018**, *39*, 115006.

[12] Orini, M.; Bailon, R.; Mainardi, L.T.; Laguna, P.; Flandrin, P. Characterization of dynamic interactions between cardiovascular signals by time-frequency coherence. *IEEE Transactions on Bbio-medical Engineering* **2012**, *59*, 663-673.

[13] Shoenb, A.; Cliord, G. Chapter 16—wavelets; multiscale activity in physiological signals. *Biomedical Signal and Image Processing; Springer: New York, NY, USA* **2005**.

[14] Cheplygina V, de Bruijne M, Pluim JP. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical Image Analysis* **2019**;54:280–296.

[15] Hwang, S.; Kim, H.-E. Self-transfer learning for weakly supervised lesion localization. In Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention; pp. 239-246.

[16] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of Advances in Neural Information Processing Systems; pp. 1097-1105.

[17] Zhu, W.; Zeng, N.; Wang, N. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *NESUG Proceedings: Health Care and Life Sciences, Baltimore, Maryland* **2010**, *19*, 67.

[18] Satija, U.; Ramkumar, B.; Manikandan, M.S. A simple method for detection and classification of ECG noises for wearable ECG monitoring devices. In Proceedings of 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN); pp. 164-169.

[19] Lake, DE; Moorman, RJ. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices, *Am J Physiol Heart Circ Physiol* **300** (1) (2011) H319-H325.