# Comparative Study of Convolutional Neural Networks for ECG Quality Assessment

Álvaro Huerta[1], Arturo Martinez-Rodrigo[1], Alberto Puchol[2], Marta I Pachón[2], José J Rieta[3], Raúl Alcaraz[1]

[1] Research Group in Electronic, Biomedical and Telecommunications Engineering,
University of Castilla La-Mancha, Cuenca, Spain
[2] Cardiac Arrhythmia Department, Hospital Virgen de la Salud, Toledo, Spain
[3] BioMIT.org, Electronic Engineering Department, Universitat Politecnica de Valencia, Spain

## Abstract

*In the last years, convolutional neural networks (CNNs) have become popular in ECG analysis, since they do not require pre-processing stages, nor specific pre-training. However, their ability for ECG quality assessment has still not been thoroughly assessed. Hence, this work introduces a comparison about the ability of several CNN algorithms to classify between high and low-quality ECGs. Taking advantage of the concept of transfer learning, five common pre-trained CNNs were analyzed, such as AlexNet, GoogLeNet, VGG16, ResNet18 and InceptionV3. They were fed with 2-D images obtained by turning 5 second-length ECG segments into scalograms through a continuous Wavelet transform. To train and validate the algorithms, 1,168 noisy ECG intervals, along with other 1,200 ECG excerpts with sufficient quality for their further interpretation, were extracted from a public database. The obtained results showed that all CNNs provided mean values of accuracy between 89 and 91%, but notable difference in terms of computational load were noticed. Thus, AlexNet was the fastest algorithm, requiring notably less CPU usage and memory than the remaining methods. Consequently, this CNN exhibited the best trade-off between high-quality ECG identification accuracy and computational load, and it could be considered as the most convenient algorithm for ECG quality assessment.*

## 1. Introduction

In the last decade, convolutional neural networks (CNNs) have evolved significantly. Indeed, they are being today applied in a broad variety of areas, such as sports, health, automotive and robotics, among others [1]. Within the field of medicine, these networks have been widely employed for automatic tasks, such as analysis and interpretation of ECG signals, classification of different kinds of arrhythmias, or biometric identification of subjects [2]. Nonetheless, ECG quality assessment is an interesting and unresolved clinical challenge where CNNs have still not been explored.

In general terms, accurate interpretation of the ECG recording could be blurred by the presence of large levels of noise, even making the signal unusable [3]. This problem is aggravated in the case of wearable or portable ECG monitoring systems. These devices have the ability to acquire ECG signals for extended periods of time in uncontrolled and ever-changing environments, while the patient continues with a normal daily life, and therefore the acquired ECG signal could be strongly affected by different kinds of noises [4]. The most common nuisance interferences found in these ECG recordings are powerline interference, baseline wander, motion artifacts, and high-frequency electromyography disturbances [4]. In this context, automatic identification of contaminated and noisy ECG intervals is essential to only process those excerpts with sufficient quality to avoid further misinterpretations and misdiagnoses [3].

To this end, several algorithms, mostly based on the detection of ECG fiducial points and other morphological events, have been recently proposed [3]. However, the ECG recordings strongly affected by noise and morphological alterations do not allow reliable detection of most of their relevant waves and points [5], thus reducing the ability of those methods for an accurate ECG quality assessment [3]. In fact, some authors are currently demanding other kinds of indicators of ECG quality which are not based on morphological ECG features and events [3].

Recently, the use of CNNs has emerged as a viable option to estimate ECG quality. These networks are able to obtain the most relevant ECG characteristics without the need of detecting and delineating its fiducial points and other waveforms. To date, however, only a few works have introduced algorithms based on CNNs for ECG quality as-

ISSN: 2325-887X DOI: 10.22489/CinC.2020.370

sessment [6, 7]. Moreover, these methods could be notably overfitted, since they have been designed and trained from scratch with a reduced number of ECG samples. In fact, the lack of publicly available datasets with enough ECG samples and annotations makes the use of pre-trained CNN architectures an attractive alternative to discern between high- and low-quality ECG excerpts. Taking advantage of the concept of transfer learning, a pre-trained CNN scheme can be adapted to a new task, thus saving learning time and resources [8]. For instance, the performance of a well-known pre-trained CNN has been recently exhibited a notably better performance in ECG quality assessment than other CNN models designed from scratch [9]. Therefore, the main goal of the present work is to compare several pre-trained CNN architectures to discern between high- and low-quality ECG segments obtained from a portable recording system.

## 2. Database

In the present work, the training dataset proposed for the PhysioNet/CinC Challenge 2017 [10] was used. This database is composed by 8,528 single-lead ECG signals lasting between 9 and 60 seconds. They were acquired with a sampling rate of of 300 Hz and 16 bits of resolution by linking a portable ECG monitoring system (AliveCor™) to a smartphone. In addition to the signals, their classification by experts into four different rhythms, i.e., normal sinus rhythm, atrial fibrillation, other rhythms, and noisy recordings, is also available.

To make training and validation of the CNN-based algorithms possible, the ECG signals were divided into 5 second-length intervals and clustered into two classes. The excerpts belonging to noisy recordings formed the group of low-quality ECGs, whereas those obtained from signal containing sinus rhythm, atrial fibrillation, and other rhythms constituted the group of high-quality ECGs. The result was a dataset with a great imbalance between both groups, such as Table 1 shows. To overcome this problem, a balanced subset was constructed by randomly selecting 1,200 ECG intervals with the a high-quality label, as well as all 1,168 excerpts with a poor-quality one. It should be noted that representativity of all rhythms was maintained in the new high-quality subset, becase it was composed by 400 sinus rhythm excerpts, 400 atrial fibrillation episodes and 400 intervals of other rhythms.

Table 1. Number of recordings and 5 second-length ECG segments found in the training set of the PhysioNet/CinC Challenge 2017 database for each cardiac rhythm.

| Rhythm | # of Recordings | # of Segments |
|---|---|---|
| Sinus Rhythm | 5,154 | 28,413 |
| Atrial Fibrillation | 771 | 4,329 |
| Other Rhythms | 2,557 | 14,697 |
| Noise | 46 | 1,168 |

## 3. Methods

### 3.1. ECG-based scalograms

Many pre-trained CNNs are 2-D models fed with images [11]. Hence, ECG excerpts were firstly transformed into 2-D images to be inputted to the five compared algorithms. For that purpose, time-frequency representation of every 5 second-length ECG interval was turned into a 2-D matrix using a Continuous Wavelet transform [12]. The resulting wavelet coefficients were then graphically represented with a Jet colormap to obtain a wavelet scalogram. This kind of 2-D image has been used as input for many CNN-based algorithms in a wide variety of scenarios [11]. The parameters used in the described transformation were a Morlet function as mother wavelet, and a number of wavelet scales determined by 48 voices per octave.

### 3.2. Pre-trained CNN models

Five well-known CNN models were here compared, i.e., AlexNet, VGG16, GoogLeNet, ResNet18 and InceptionV3. AlexNet became very popular after winning a famous challenge dealing with visual recognition [13]. Since then, it has been used in a huge variety of applications [1]. The original architecture of this network is composed of eight layers with learning ability, where five are convolutional and three are fully-connected. Furthermore, to reduce spatial length of the feature map, two pooling layers are included in the model. Moreover, a linear activation function is employed in all convolutional and fully-connected layers. Finally, to reduce the problem of overfitting, two dropout regularization functions are inserted after the two first fully-connected layers.

Thanks to its architecture based on the utilization of small-sized convolution kernels, VGGNet has also become a popular CNN scheme [14]. This network takes advantage of the idea that highly deep and efficient models can be reached by making use of numerous convolutional layers with small-sized kernels. There exist several variants of this model with different deeper levels, but the most used is VGG16 and thereby it was selected for the present work [14]. This CNN is composed by 13 convolution layers with different number of kernels, and 5 max-pooling layers for feature extraction. At the end of the model, three full-connected layers are also found. As before, activation and dropout regularization functions are included in the model to prevent overfitting.

From a structural point of view, GoogLeNet [15] presents an architecture totally different from the previous networks. While AlexNet and VGG16 are based on a sequential structure of layers, GoogLeNet makes use of a model composed of several branches in parallel. In fact, this network introduces a novelty, i.e., the *Inception Module*. This block is built through three convolutional and one max-poling layers, which are placed in parallel to finally combine their feature maps.

On the other hand, the main characteristic in the ResNet model is the use of residual blocks, which allow to easily train very deep CNN architectures [16]. Hence, given an input, the convolution layers transform the input data into residue feature maps, while the original input is added directly to the transformed signal bypassing the convolution layers. In this way, the output is the sum of the original input data with the residue feature maps. Although several variants of this architecture exist, the popular ResNet18 was used in the present work.

Finally, InceptionV3 consists of a network architecture based on three different inception modules, similar to that previously described for GoogLeNet [17]. The main contribution of this network is the use of different ways for factoring data into smaller convolutions, thus increasing computational efficiency and reducing grid size.

It is worth noting that all the models have been designed to deal with hundreds, or even thousands, of different classes, and hence the output layer of the five networks was adapted to exclusively work with two classes, i.e., high- and low-quality ECGs.

## 3.3. Performance assessment

Training and validation of all CNN-based algorithms were developed by running 5 times a holdout approach, where data were stratifiedly divided into two sets, i.e., 80% of samples were used for training and 20% for testing. Note that, to conduct the fine-tuning of the algorithms, a stochastic gradient descent with a momentum of 0.9 and a learning rate of 0.001 was used during ten epochs with mini-batches of ten samples.

On the one hand, the ability of each method to discern between high- and low-quality ECG excerpts was assessed in terms of sensitivity (Se), specificity (Sp) and accuracy (Acc). Additionally, computational time, CPU usage, and memory consumption required by the algorithms during the validation phase were also computed. All experiments were developed with Matlab R2019a (The MathWorks, Inc), under an all-in-one workstation HP ProOne 600 G2 with a processor Intel i7 at 3.41 GHz. The results obtained for all these metrics in the five conducted training/testing cycles were averaged to provide a more global overview about the performance of each method.

Table 2. Values of mean and standard deviation for Se, Sp and Acc obtained by the analyzed CNN-based algorithms in the validation phase.

| **Model** | **Se (%)** | **Sp (%)** | **Acc (%)** |
|---|---|---|---|
| AlexNet | 88.90±3.27 | 92.50 ±4.00 | 90.70±0.96 |
| VGG16 | 85.60±8.98 | 93.70±3.87 | 89.65±3.42 |
| GoogLeNet | 88.80±2.77 | 92.70±1.72 | 90.75±0.75 |
| ResNet18 | 88.40±2.19 | 93.80±1.30 | 91.10±0.91 |
| InceptionV3 | 89.00±2.24 | 93.50±0.79 | 91.25±0.88 |

Table 3. Values of mean and standard deviation for computational time, CPU consumption and memory usage required by each algorithm during the validation phase.

| **Model** | **Time (s)** | **CPU (%)** | **Mem. (MB)** |
|---|---|---|---|
| AlexNet | 12.76±0.21 | 30.40±0.52 | 2814.8±75.08 |
| VGG16 | 78.84±0.57 | 39.27±0.29 | 4854.2±58.87 |
| GoogLeNet | 25.33±0.45 | 37.03±0.31 | 1589.8±8.87 |
| ResNet18 | 21.07±0.36 | 41.65±0.32 | 3266.0±24.03 |
| InceptionV3 | 76.86±0.25 | 32.51±0.33 | 1814.2±15.97 |

## 4. Results

As can be seen in Table 2, similar mean values of Acc were obtained for all methods, ranging from 89.65% for VGG16 to 91.25% for InceptionV3. Moreover, the mean values of Sp and Se were well-balanced for all algorithms, because differences between 1 and 3% were only noticed. Nonetheless, for all cases a better ability to identify low-quality ECG segments than high-quality ones was observed, since mean values of Sp were about 92% and Se around 88%. Of note is also that a low dispersion was observed for the three performance metrics, presenting values of standard deviation lower than 4% in most cases.

Regarding computational time, CPU consumption, and memory usage required by the CNN models during the testing phase, the results are showed in Table 3. The most remarkable differences among the methods were noticed in computational time. Precisely, this measure ranged from 12.76 to 78.84 seconds per iteration, being AlexNet 6.5 times faster than VGG16. Likewise, the memory required by GoogLeNet was significantly lower than the one needed by VGG16, varying this mesure between 1,589.8 and 4,854.2 MB. Finally, regarding the CPU consumption, slightly differences were only encountered among all methods. Nonetheless, it should be noted that AlexNet was the algorithms requiring less CPU usage.

## 5. Discussion and conclusions

In the present work, five pre-trained CNN models have been compared for quality assessment of single-lead ECG signals acquired using a portable recording system. The obtained results have shown that VGG16 and InceptionV3 required much more time than the remaining algorithms to classify all ECG intervals included in the subset of testing. Furthermore, for a comparable CPU comsumption, VGG16 also demanded substantially more memory than the other methods. The huge number of hyper-parameters required by this CNN scheme could explain this behavior. In fact, VGG16 contains more than 138 million of hyper-parameters, whereas the others less than 65 million [14].

On the other hand, GoogLeNet and ResNet18 have reported computational time and CPU usage notably higher than AlexNet. In fact, both algorithms required nearly twice the time and around 7-10% more of CPU consumption than AlexNet for the same task. In view of these results and having in mind that no great differences in values of Se, Sp, and Acc were noticed for all CNN schemes, it could be concluded that AlexNet has reported the best trade-off between poor-quality ECG identification accuracy and computational load, and therefore it is the most convenient CNN-based approach for quality assessment of ECG signals obtained from wearable and portable recording systems.

## Acknowledgments

This research has been supported by the grants DPI2017–83952–C3 from MINECO/AEI/FEDER EU, SBPLY/17/180501/000411 from Junta de Comunidades de Castilla-La Mancha and AICO/2019/036 from Generalitat Valenciana.

## References

[1] Patil T, Pandey S, Visrani K. A review on basic deep learning technologies and applications. In Kotecha K, Piuri V, Shah H, Patel R (eds.), Data Science and Intelligent Applications, volume 52. Springer, 2021; 565–573.

[2] Hong S, Zhou Y, Shang J, Xiao C, Sun J. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. Comput Biol Med Jul 2020; 122:103801.

[3] Satija U, Ramkumar B, Manikandan MS. A review of signal processing techniques for electrocardiogram signal quality assessment. IEEE Rev Biomed Eng 2018;11:36–52.

[4] Vavrinsky E, Subjak J, Donoval M, Wagner A, Zavodnik T, Svobodova H. Application of modern multi-sensor holter in diagnosis and treatment. Sensors Basel May 2020;20(9).

[5] Martínez A, Alcaraz R, Rieta JJ. Application of the phasor transform for automatic delineation of single-lead ECG fiducial points. Physiol Meas Nov 2010;31(11):1467–85.

[6] Zhang Q, Fu L, Gu L. A cascaded convolutional neural network for assessing signal quality of dynamic ECG. Computational and Mathematical Methods in Medicine 2019; 2019.

[7] Zhao Z, Liu C, Li Y, Li Y, Wang J, Lin BS, Li J. Noise rejection for wearable ECGs using modified frequency slice wavelet transform and convolutional neural networks. IEEE Access 2019;7:34060–34067.

[8] Jadhav P, Rajguru G, Datta D, Mukhopadhyay S. Automatic sleep stage classification using time–frequency images of CWT and transfer learning using convolution neural network. Biocybernetics and Biomedical Engineering 2020; 40(1):494–504.

[9] Huerta A, Martínez-Rodrigo A, González VB, Quesada A, Rieta J, Alcaraz R. Quality assessment of very long-term ECG recordings using a convolutional neural network. In Proceeding of the IEEE 7th International Conference on E-Health and Bioengineering. IEEE, 2019; 1–4.

[10] Clifford GD, Liu C, Moody B, Lehman LWH, Silva I, Li Q, Johnson AE, Mark RG. AF classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017. Computing in Cardiology Sep 2017;44:1–4.

[11] Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: A review. Neurocomputing 2016;187:27–48.

[12] Shoeb A, Cliford G. Chapter 16 – Wavelets; Multiscale activity in physiological signals. Biomedical Signal and Image Processing 2005; Springer: New York, USA;1 –29.

[13] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems. 2012; 1097–1105.

[14] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv14091556 2014;.

[15] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. 2015; 1–9.

[16] 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. 2016; 770–778.

[17] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016; 2818–2826.

Address for correspondence:

Álvaro Huerta Herraiz
ITAV, Campus Universitario S/N., 16071, Cuenca, Spain
Phone: +34-969-179-100
e-mail: alvaro.huerta@uclm.es