

# ECG Quality Assessment via Deep Learning and Data Augmentation

Álvaro Huerta<sup>1</sup>, Arturo Martínez-Rodrigo<sup>1</sup>, José J Rieta<sup>2</sup>, Raúl Alcaraz<sup>1</sup>

<sup>1</sup> Research Group in Electronic, Biomedical and Telecommunications Engineering,  
University of Castilla La-Mancha, Cuenca, Spain

<sup>2</sup> BioMIT.org, Electronic Engineering Department, Universitat Politècnica de Valencia, Spain

## Abstract

*Quality assessment of ECG signals acquired with wearable devices is essential to avoid misdiagnosis of some cardiac disorders. For that purpose, novel deep learning algorithms have been recently proposed. However, training of these methods require large amount of data and public databases with annotated ECG samples are limited. Hence, the present work aims at validating the usefulness of a well-known data augmentation approach in this context of ECG quality assessment. Precisely, classification between high- and low-quality ECG excerpts achieved by a common convolutional neural network (CNN) trained on two databases has been compared. On the one hand, 2,000 5 second-length ECG excerpts were initially selected from a freely available database. Half of the segments were extracted from noisy ECG recordings and the other half from high-quality signals. On the other hand, using a data augmentation approach based on time-scale modification, noise addition, and pitch shifting of the original noisy ECG experts, 1,000 additional low-quality intervals were generated. These surrogate noisy signals and the original high-quality ones formed the second dataset. The results for both cases were compared using a McNemar test and no statistically significant differences were noticed, thus suggesting that the synthesized noisy signals could be used for reliable training of CNN-based ECG quality indices.*

## 1. Introduction

Recent evolution of wearable devices along with development of new telemedicine and portable systems have popularized very long-term electrocardiographic (ECG) monitoring (for several weeks or months) of patients suffering from different cardiac disorders [1]. This kind of monitoring is highly promising to improve diagnosis of some common cardiovascular diseases, and especially those characterized by an intermittent nature. This is the case of atrial fibrillation (AF), whose initial episodes are mostly asymptomatic and only last for a few seconds or minutes [2]. Thus, the longer the duration of monitoring,

the greater the possibility of early identification of patients suffering from intermittent AF [2].

However, these modern devices often acquire the ECG in free-living conditions, the signal then presenting fluctuating quality [3]. Indeed, ambulatory resting ECG is well-known to be susceptible to several artifacts, such as powerline interference, muscle contractions, and baseline drifts (due to respiration), but wearable and portable systems are additionally sensitive to motion artifacts (as the users are now mobile), electrode contact noise (when sensor loses contact with skin during movement), and impulse noise [3]. Moreover, the change of noise intensity over time and overall non-stationarity of the ECG also complicate further processing of this long-term recording [3].

Regrettably, the presence of large levels of noise in the ECG is detrimental to automated decision support systems, which are imperatively required to process such big data acquired with wearable and portable devices [3]. Thus, ECG quality assessment is paramount, such that potential of very long-term monitoring can be completely exploited, but avoiding automated misdiagnosis and misinterpretation of corrupted ECG intervals [3]. For that purpose, a handful of ECG quality indices have been recently proposed. Most are based on extracting morphological features or fiducial points of the ECG and then constructing decision models via common machine learning techniques, such as support vector machines or  $k$ -nearest neighbors classifiers [4]. Although these ECG quality indices have reported interesting results, they have been overcome by more recent deep learning algorithms [5]. Moreover, these new methods also present some interesting advantages, such as ability to directly deal with the raw ECG without demanding any kind of preprocessing stage, manual or external intervention, and tedious ECG-based or R-peak-based feature computation and selection [5].

However, training of deep learning algorithms requires large amount of data and only reduced and highly unbalanced public ECG databases with reliable annotations from experts are nowadays available [6]. To face this drawback, different data augmentation techniques have been recently introduced, such as oversampling, Gaus-

sian mixture modeling, or generative adversarial networks (GAN) [7]. The most common methodology in the literature is oversampling, which consists of applying different kinds of transformations (e.g., rotation, mirroring, cropping, etc.) to the original samples for expanding the dataset [7]. Hence, the present work aims at validating usefulness of this approach in the context of ECG quality assessment through deep learning techniques.

## 2. Methods

### 2.1. Algorithm for ECG quality assessment

To discern between high- and low-quality ECG intervals, a previously published algorithm based on the well-known pre-trained convolutional neural network (CNN) AlexNet was used [5]. In brief, the architecture of this CNN is composed of eight layers with learning ability, where five are convolutional and three are fully-connected [8]. After these layers, linear activation functions are employed. Moreover, in two intermediate points of the network, pooling layers are included to reduce spatial length of the feature map. Also, two dropout regularization functions are inserted after the two first fully-connected layers to reduce the problem of over-fitting. All details about this architecture can be found in [8].

Because AlexNet is a 2-D CNN, i.e., receives as input a 2-D image, ECG intervals were transformed into a matrix using a Continuous Wavelet transform (CWT) [9]. The resulting wavelet coefficients were then graphically represented with a Jet colormap to obtain a wavelet scalogram. More details about the parameters used in this transformation can be found in [5].

### 2.2. Experimental setup

Although AlexNet is a pre-trained CNN, it firstly had to be fine-tuned for its use in ECG quality assessment [5]. Hence, the layers containing the pre-trained weights (i.e., the convolutional and fully-connected layers) were re-trained to discern between high- and low-quality ECG excerpts. Nonetheless, to assess usefulness of the oversampling approach in this fine-tuning of AlexNet, the network was re-trained and tested in two different ways. On the one hand, 2,000 5 second-length ECG excerpts from the PhysioNet/CinC Challenge 2017 database were initially used as a reference. The mentioned database is freely available and contains more than 8,000 single-lead ECG recordings with annotations from experts into four classes: AF, normal sinus rhythm (NSR), other rhythms (OR) and noise [10]. Given the huge unbalance between clean (three first classes) and noisy (forth class) recordings in the database, only 1,000 high-quality and 1,000 low-quality segments were analyzed, such as in previous

works [5]. Note that similar proportions of the three different rhythms in high-quality ECG signals were maintained, thus analyzing 300 AF, 400 NSR, and 300 OR segments.

On the other hand, AlexNet was re-trained and tested on a surrogate dataset generated from the previous one. Since annotations of noise in public ECG databases are much less frequent than those for clear heart rhythms (from clean signals), the 1,000 high-quality ECG intervals were maintained. Contrarily, a new subset was obtained by randomly applying different transformations to the original 1,000 low-quality ECG excerpts. Precisely, four transformations were simultaneously considered for each segment with an occurrence probability of 0.5. The first transformation was a *time stretching*, where duration of the ECG signal was slightly modified by a factor ranging between 0.75 and 1.5. This operation does not affect the pitch at all, but resampling was needed. The second transformation was *pitch shifting*, such that the ECG pitch was raised or lowered scaling by a factor ranging between -2 and 2. Note that the ECG was considered as a sound and its pitch was measured in semitones. The third transformation was *noise addition*. In this case, Gaussian white noise was added by adjusting a signal to noise ratio between 0 and 3 dB. Finally, the fourth transformation was *amplitude modification*, where the ECG gain was changed between -5 y 5 dB.

As an example, Figure 1 shows the result of each transformation individually applied to a typical 5 second-length ECG excerpt, which presents electrode contact noise (a). After the first transformation (b), a shorter ECG signal can be seen. Notable changes in the ECG morphology and amplitude can also be noticed after pitch shifting (c). The effect of the Gaussian noise is also clearly observed, because the original ECG morphology is completely degraded (d). Contrarily, the original morphology of the ECG signal is totally preserved after the four transformation, but its amplitude is notably altered (d). The ECG-based images obtained for each example are shown in Figure 2. As can be seen, there exists a clear concordance between each ECG and its corresponding image. To this respect, Figure 2(a) shows how QRS complexes correspond with Figure 1(a), in which R-peaks and the artifact are conspicuous. Figures 2(b) and (c) display how time stretching and pitch shifting have modified the signal morphology, completely changing the original hallmark in the wavelet scalogram. Gaussian noise addition also alters the original pattern by blurring R-peaks, as Figure 2(d) shows. Finally, Figure 2(e) depicts how the gain adjustment is the transformation which applies less distortion to the original signal.

### 2.3. Performance analysis

Classification performance achieved by AlexNet after fine-tuning on both the original and surrogate datasets was analyzed through a stratified 5-fold cross-validation ap-

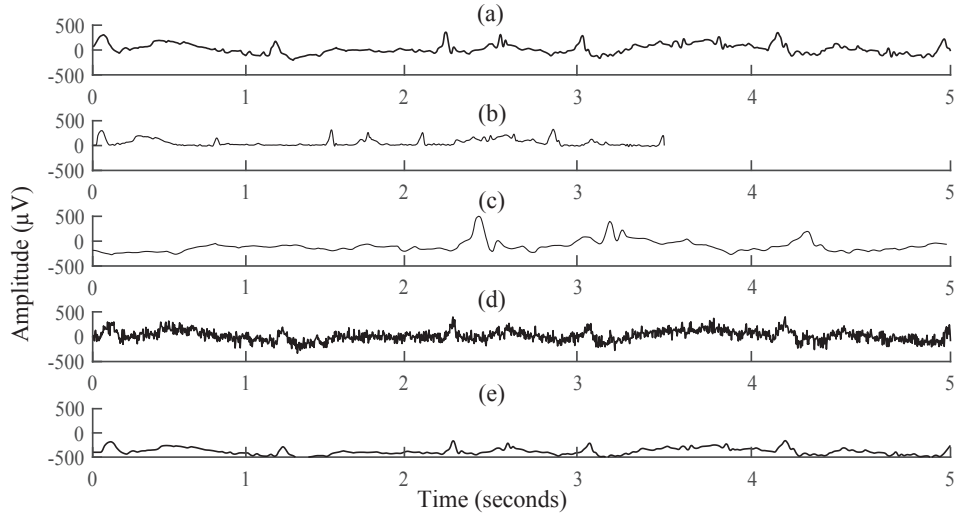


Figure 1. Result of common transformations in the oversampling approach individually applied to a noisy 5 ECG excerpt (a). The transformations were: time stretching (b), pitch shifting (c), noise addition (d) and amplitude modification (e).

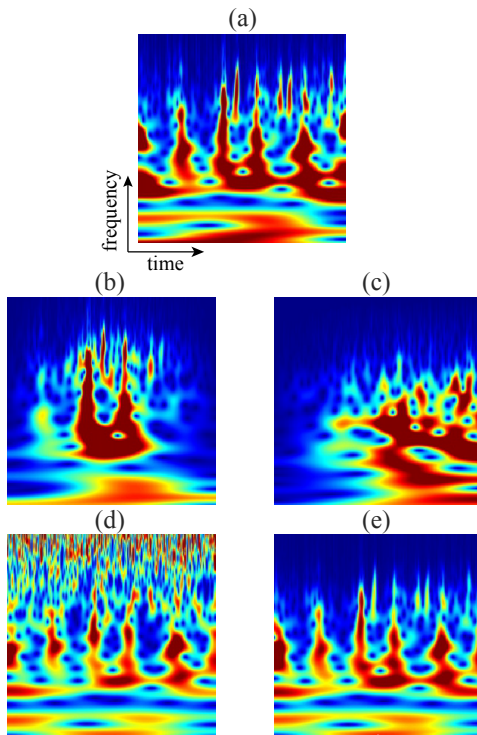


Figure 2. Images obtained from the ECGs presented in Figure 1 by applying CWT and inputted to AlexNet for ECG quality assessment.

proach, and the two cases were compared using a McNemar test. Classical statistics of sensitivity ( $Se$ ), specificity ( $Sp$ ) and accuracy ( $Acc$ ) were computed. Thus,  $Se$  was defined as the percentage of correctly identified high-quality

ECG excerpts,  $Sp$  as the rate of properly classified low-quality segments, and  $Acc$  as the total ratio of all ECG intervals appropriately detected. Finally, the success rates of correctly identified NSR ( $\mathcal{R}_{NSR}$ ), AF ( $\mathcal{R}_{AF}$ ) and OR ( $\mathcal{R}_{OR}$ ) within the high-quality ECG group were also computed for the two experiments.

### 3. Results

A McNemar test reported no statistically significant differences ( $p$ -value  $> 0.05$ ) between the classification results obtained by AlexNet after re-training on the original and surrogate databases. In fact, Table 1 displays very similar values of  $Acc$ ,  $Se$  and  $Sp$  for both cases. Similarly, no differences in the values of  $\mathcal{R}_{NSR}$ ,  $\mathcal{R}_{AF}$  and  $\mathcal{R}_{OR}$  were also observed. Nonetheless, values about 90% were seen for all performance metrics, with standard deviation among the five conducted iterations being lower than 6%.

### 4. Discussion

To the best of our knowledge, this is the first study dealing with the usefulness of a data augmentation approach for the training of CNN-based algorithms in the context of ECG quality assessment. The classification results obtained for the two analyzed databases suggest that common transformations (i.e., time stretching, pitch shifting, noise addition and amplitude modification) used for data augmentation in other fields, e.g., audio deep learning, could also be useful in this case. However, it is worth noting that the present pilot study has only assessed the possibility of generating realistic noisy ECG signals using such transformations, and the effect of oversampling the low-

Table 1. Classification results obtained by the AlexNet-based algorithm re-trained on the original and surrogate databases.

Dataset	Classification Performance					
	$Se$	$Sp$	$Acc$	$\mathcal{R}_{NSR}$	$\mathcal{R}_{AF}$	$\mathcal{R}_{OR}$
Original	0.89±0.02	0.87±0.045	0.88±0.02	0.92±0.03	0.88±0.03	0.87±0.04
Surrogate	0.90±0.03	0.90±0.04	0.90±0.01	0.91±0.02	0.89±0.06	0.89±0.05

quality ECG group on the fine-tuning of CNN-based algorithms will be analyzed in a future work. In fact, the kind of study conducted in the present work has motivated that the original and surrogate datasets shared the same 1,000 high-quality ECG intervals for an unbiased comparison.

Although making use of other transformations, a few recent works have also suggest that data augmentation could be useful in improving the training of CNN-based methods in a variety of ECG-based contexts. To this respect, Alghamdi et al. [11] have reported better detection of myocardial infarction with different CNN architectures when data augmentation was conducted with common image transformation techniques. Similarly, Shaker et al. [6] have proposed a GAN to synthesize artificial heartbeats and then improve their subsequent classification with a CNN-based algorithm. Despite the good results reported by this work and the current trend to use GANs for data augmentation in many fields [6], this alternative could not be appropriate to generate noisy ECG recordings. The great variety of artifacts and nuisance components in the ECG signals could make it difficult to find recurrent patterns, even for the high abstraction levels achieved by the CNNs. Nonetheless, this data augmentation approach will be analyzed in the future.

## 5. Conclusions

The use of common transformations for the well-known data augmentation approach of oversampling could be interesting to increase the number of noisy ECG signals for the fine-tuning of CNN-based ECG quality indices, thus resulting in more robust and reliable methods. However, more studies considering wider databases and using additional data augmentation techniques are still required for validation of the obtained results.

## 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, AICO/2021/286 from Generalitat Valenciana and FEDER 2018/11744.

## References

[1] Sana F, Isselbacher EM, Singh JP, Heist EK, Pathik B, Armondas AA. Wearable devices for ambulatory cardiac

monitoring: JACC state-of-the-art review. *Journal of the American College of Cardiology* 2020;75(13):1582–1592.

[2] Stachon P, Ahrens I, Faber T, Bode C, Zirlik A. Asymptomatic atrial fibrillation and risk of stroke. *Panminerva Med Dec* 2015;57(4):211–5.

[3] Smital L, Haider CR, Vitek M, et. al. Real-time quality assessment of long-term ECG signals recorded by wearables in free-living conditions. *IEEE Trans Biomed Eng* 2020; 67(10):2721–2734.

[4] Satija U, Ramkumar B, Manikandan MS. An automated ECG signal quality assessment method for unsupervised diagnostic systems. *Biocybern Biomed Eng* 2018;38:54–70.

[5] Huerta Herraiz Á, Martínez-Rodrigo A, Bertomeu-González V, Quesada A, Rieta JJ, Alcaraz R. A deep learning approach for featureless robust quality assessment of intermittent atrial fibrillation recordings from portable and wearable devices. *Entropy* 2020;22(7):733.

[6] Shaker AM, Tantawi M, Shedeed HA, Tolba MF. Generalization of convolutional neural networks for ECG classification using generative adversarial networks. *IEEE Access* 2020;8:35592–35605.

[7] Hatamian FN, Ravikumar N, Vesal S, Kemeth FP, Struck M, Maier A. The effect of data augmentation on classification of AF in short single-lead ECG signals using deep neural networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2020; 1264–1268.

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

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

[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. *Comput Cardiol* 2017;44.

[11] Alghamdi A, Hammad M, Ugail H, et. al. Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities. *arXiv preprint arXiv:190609358* 2019;ISSN 1380-7501.

Address for correspondence:

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