Application of Deep Learning for Quality Assessment of Atrial Fibrillation ECG Recordings

Álvaro Huerta[1], Arturo Martinez-Rodrigo[1], Miguel A Arias[2], Philip Langley[3], José J Rieta[4], 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] Faculty of Science and Engineering, University of Hull, UK
[4] BioMIT.org, Electronic Engineering Department, Universitat Politecnica de Valencia, Spain

Abstract

In the last years, atrial fibrillation (AF) has become one of the most remarkable health problems in the developed world. This arrhythmia is associated with an increased risk of cardiovascular events, being its early detection an unresolved challenge. To palliate this issue, long-term wearable electrocardiogram (ECG) recording systems are used, because most of AF episodes are asymptomatic and very short in their initial stages. Unfortunately, portable equipments are very susceptible to be contaminated with different kind of noises, since they work in highly dynamics and ever-changing environments. Within this scenario, the correct identification of free-noise ECG segments results critical for an accurate and robust AF detection. Hence, this work presents a deep learning-based algorithm to identify high-quality intervals in single-lead ECG recordings obtained from patients with paroxysmal AF. The obtained results have provided a remarkable ability to classify between high- and low-quality ECG segments about 92%, only misclassifying around 7% of clean AF intervals as noisy segments. These outcomes have overcome most previous ECG quality assessment algorithms also dealing with AF signals by more than 20%.

1. Introduction

In the last years, atrial fibrillation (AF) has become one of the most remarkable health problems in the developed world, roughly affecting 37.5 million of people [1]. Moreover, its prevalence has doubled in the last decade, being more acute on those aged older than 60 years. Even though this arrhythmia is not life-threatening in itself, it can lead to stroke, heart failure and many other heart-related complications [2]. Moreover, around 90% of people suffering from AF for the first time show no symptoms, thus making their early diagnosis extremely difficult [3]. To palliate this issue, potential patients with paroxysmal AF are often monitored with ECG recording systems. In this way, long-term ECG signals can be acquired, while the patient can lead a normal live. However, when these devices work outside of a controlled hospital environment, they are usually contaminated with different kind of noises, such as power-line interference, motion artifacts, and high-frequency muscular disturbances [4].

For ECG denoising, a broad variety of methodologies have been proposed in the literature [5]. However, their performance have shown to be limited, because most of them are based on identifying and preserving time and frequency ECG features, which usually coincide with noise components [6]. Considering this context, proper identification of highly contaminated and strongly noisy ECG excerpts is a main priority for accurate early AF detection, as well as for reaching precise ECG-based diagnosis and clinical decisions.

Clearly, quality assessment of long-term ECG recordings requires automatic algorithms, since they are more robust and faster than human visual inspection. Thus, different automatic techniques based on time and morphological features can be found in the literature [6, 7]. Nevertheless, these methods have reported a performance closely dependent to their ability to detect ECG boundary points, which notably decreases when ECG signals are contaminated with noise [8]. Unlike these algorithms, the widely-used convolutional neural networks (CNNs) have reported a positive ability to discern between high- and low-quality segments in some very recent works, without requiring delineation of ECG fiducial points, or any other kind of previous preprocessing operations [4, 5, 9].

However, the performance of these algorithms have been validated exclusively using ECG recordings from healthy people, not considering a most challenging sce-

nario where normal sinus rhythm (SR) is alternated with AF episodes [7]. Additionally, these methods have designed from scratch using reduced databases, thus possibly leading to be notably overffited [10]. Within this context, the present work analyzes the ability of the well-known pre-trained AlexNet architecture [11] to discriminate between high- and low-quality ECG excerpts in a challenging scenario where SR and AF episodes are alternated.

2.   Methods

2.1.   Database

To validate the proposed method, the training dataset proposed for the PhysioNet/CinC Challenge 2017 was used. This database is formed by 8,528 single-lead ECG recording lasting between 9 and 60 seconds. The signals were acquired with a sampling rate of 300 Hz and 16 bits of resolution by linking a portable ECG monitoring system (AliveCor™) to a smart phone. Apart from signal recordings, their classification by experts into four different rhythms, i.e., normal sinus rhythm, atrial fibrillation, other rhythms, and noisy recordings, is also available.

No kind of preprocessing was applied to the ECG signals, which were segmented into 5 second-length intervals and rearranged into two groups. More precisely, ECG excerpts obtained from AF, SR and other rhythms (OR) recordings composed the high-quality group, whereas those extracted from noisy recordings formed the low-quality group. Thus, a total of 48,607 ECG excerpts were finally analyzed, such that 47,439 (97.59%) belonged to the high-quality group and 1,168 (2.41%) to the low-quality group. It is worth noting that high-quality group contained 4,329 intervals of AF, 28,413 of SR and 14,697 of OR.

2.2.   Continuous Wavelet Transform

Time and frequency information in every ECG except was exploited by transforming it into a two-dimensional representation using a continuous Wavelet transform (CWT). This tool has proven superior ability to deal with local, transient and intermittent aperiodicities in non-stationary signals, such as the ECG, compared with other existing time-frequency transformation methodologies [12]. From a conceptual point of view, CWT decomposes a signal at different time scales, each one representing a particular frequency range in the time-frequency plane. The approach is based on correlating an initial signal with scaled and shifted versions of a wavelet function, called mother wavelet and referred to as $\psi(t)$. Mathemati-
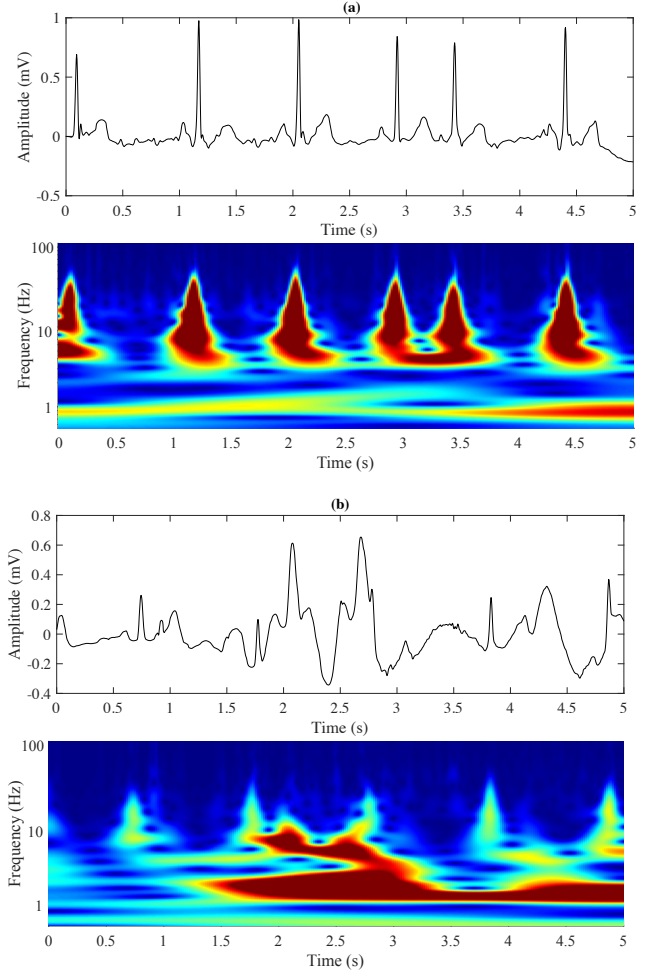

Figure 1.   Typical 5 second-length ECG intervals from the (a) high- and (b) low-quality groups, along with their corresponding scalograms.

cally speaking, CWT of a signal $x(t)$ is obtained as

$$CWT(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)\psi^*\Big(\frac{t-b}{a}\Big)dt, \quad (1)$$

where $a, b \in \Re, a \neq 0$ are the scaling and shifting parameters of the mother wavelet, respectively, and $*$ denotes the complex conjugate operator. The outcome of this operation is a two-dimensional matrix containing wavelet coefficients located according to the analyzed scales and positions. Applying a color map to the absolute value of these coefficients, an image, known as scalogram, is obtained. This kind of representation of the ECG has been widely inputted to many CNN-based algorithms in a variety of scenarios [13]. As an example, Figure 1 shows typical high- and low-quality ECG intervals, along with their corresponding scalograms.

2.3.   The AlexNet architecture

CNNs have been intensively used over the last few decades for solving pattern recognition problems, specially in image classification [14]. The high capability of extracting spatial and temporal features from input data, without requiring extra preprocessing operations, is the strongest point of these algorithms. They are constructed employing convolution, pooling and fully-connected layers [13].

In the present work, the widely-used AlexNet architecture [11] was fine-tuned to classify low- and high-quality ECG segments. From a structural perspective, AlexNet is composed of eight sequential layers, where five are convolutional and three are fully-connected. Additionally, two pooling layers are also used to reduce spatial dimension of the feature map in the previous stages [11]. Moreover, a rectifier linear unit function is used after each convolutional and fully-connected layers. Finally, dropout regularization operations are performed after the two first fully-connected layers to reduce overfitting as much as possible. More details about this CNN scheme can be found in [11].

Initially, AlexNet was created to distinguish among more than 1,000 different classes [11]. To adapt the network to the task proposed in this study, the last fully-connected layer was re-adapted to exclusively work with two classes. Furthermore, the resulting model was also fine-tuned with the database described in Section 2.1, and making use of stochastic gradient descent algorithm with a momentum of 0.9. The initial learning rate was set to a constant value of 0.0001 during all the training time, such that no learn rate drop factor was used during the conducted 10 epochs with mini batch sizes of 10 scalograms.

2.4.   Statistical analysis

Bearing in mind the great imbalance between groups found in the database (see Section 2.1), severals validation cycles were conducted. Thus, 40 training/testing iterations were developed, such that two subsets of 1,150 samples randomly selected from high- and low-quality groups were considered in each one. Within each iteration, a stratified 80/20 (80% for training and 20% for testing) holdout scheme was repeated 5 times, thus obtaining a total of 200 fine-tuned CNN-based models.

The classification performance of every model was evaluated in terms of sensitivity (Se), specificity (Sp) and accuracy (Acc). More precisely, Se was defined as the rate of correctly classified high-quality ECG segments, while Sp was computed as the percentage of properly identified low-quality intervals. Acc was obtained as the percentage of total ECG excerpts rightly detected. Beyond these metrics, the rates of correctly classified SR ($\mathcal{R}_{\mathcal{SR}}$), AF ($\mathcal{R}_{\mathcal{AF}}$) and OR ($\mathcal{R}_{\mathcal{OR}}$) segments within the high-quality group were also obtained. It should be noted that values of

Table 1. Values of mean, std, maximum and minimum obtained for all performance metrics from the 200 conducted validation cycles.

Value	Se	Sp	Acc	$\mathcal{R}_{\mathcal{SR}}$	$\mathcal{R}_{\mathcal{AF}}$	$\mathcal{R}_{\mathcal{OR}}$
Mean	90.4%	93.2%	91.4%	91.5%	92.4%	87.3%
Std.	2.8%	1.3%	2.5%	2.4%	3.4%	4.3%
Max.	96.4%	98.3%	96.7%	96.1%	97.8%	96.1%
Min.	87.6%	93.5%	88.7%	88.8%	86.8%	79.9%

mean, standard deviation (std), maximum, and minimum for these performance metrics were also computed from all validation cycles.

3.   Results

Table 1 shows values of mean, std, maximum and minimum for all performance metrics. As can be observed, mean values of Acc around 91%, with a maximum of almost 97%, were achieved. Moreover, slight differences were only seen between Se and Sp, reaching comparable maximum values. On the other hand, within the high-quality group, mean classification rates of both SR and AF episodes were longer than 91%, but the mean value of $\mathcal{R}_{\mathcal{OR}}$ fell to 87%. Finally, low values of std (lower than 4.5%) were obtained for all performance metrics, thus showing a reduced dispersion among the results obtained for all conducted validation cycles.

4.   Discussion

In the present work, automatic quality assessment of single-lead ECG signals acquired using a portable recording system from patients with paroxysmal AF has been explored through a CNN-based algorithm. Although different databases have been used for the validation of previous algorithms found in the literature and results should be compared with caution, the proposed approach has reached mean classification rates comparable or even higher than those methods [15]. Moreover, it is interesting to note that a more robust validation procedure than in previous works has been here conducted by analyzing more than 48,000 ECG excerpts, rearranged into 200 learning/testing iterations. Indeed, most of the previous studies have only used one validation cycle with a limited set of data [15]. For instance, Zhao & Zhang [16] only analyzed 300 30 second-length ECG intervals (150 for high-quality and 150 for low quality) obtained from a much more larger public database.

It is also interesting to note that only SR signals from healthy subjects have been used for the validation of most previous algorithms dealing with ECG quality assessment [7]. However, considering other rhythms apart from SR, involves a more challenging context, because ECGs

with a wider variety of shapes and morphologies are inputted to the methods [7]. Indeed, when this scenario have been considered to validate some previous algorithms, drastic reductions in their performance have been noticed, only reaching values of Acc between 50 and 70% [7, 15]. In contrast, the method presented in this work has been able to report a good performance even in the presence of AF and other rhythm, correctly classifying noisy signals and AF episodes with an accuracy greater than 90%.

A relevant limitation of the proposed algorithm is the lack of a clear interpretation of the obtained results. In fact, all CNN-based methodologies are characterized by concealing links between input and outputs [17]. Also, despite a wide dataset with ECG samples from a fully realistic context has been analyzed in the present work, further analysis with ECG recordings acquired using different kinds of portable and wearable recording systems could be helpful in contrasting the presented outcomes.

5.   Conclusions

A novel deep learning method able to facilitate automatized diagnosis with great accuracy in long-term ECG monitoring of patients with intermittent AF has been proposed. The algorithm has been throughly validated by making use of many ECG samples and numerous learning/testing cycles, reporting a better performance than most previous techniques proposed for ECG quality assessment. Hence, the use of this tool in portable and wearable recording systems could avoid potential confounding bias of poor-quality ECG intervals, thus leading to more robust and precise diagnosis and clinical decisions about some cardiac pathologies, such AF and other arrhythmias.

Acknowledgment

References

[1] Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge. Int J Stroke Jan 2020;1–5.

[2] Khoo CW, Krishnamoorthy S, Lim HS, Lip GYH. Atrial fibrillation, arrhythmia burden and thrombogenesis. Int J Cardiol Jun 2012;157(3):318–23.

[3] 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.

[4] 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.

[5] Yoon D, Lim HS, Jung K, Kim TY, Lee S. Deep learning-based electrocardiogram signal noise detection and screening model. Healthcare Informatics Research 2019;25(3):201–211.

[6] Oster J, Behar J, Sayadi O, Nemati S, Johnson AE, Clifford GD. Semisupervised ecg ventricular beat classification with novelty detection based on switching kalman filters. IEEE Transactions on Biomedical Engineering 2015;62(9):2125–2134.

[7] Behar J, Oster J, Li Q, Clifford GD. ECG signal quality during arrhythmia and its application to false alarm reduction. IEEE Transactions on Biomedical Engineering 2013; 60(6):1660–1666. ISSN 0018-9294.

[8] Satija U, Ramkumar B, Manikandan MS. Automated ECG noise detection and classification system for unsupervised healthcare monitoring. IEEE J Biomed Health Inform 05 2018;22(3):722–732.

[9] 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. ISSN 1748-670X.

[10] 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.

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

[12] Shoeb A, Cliford G. Chapter 16—wavelets; multiscale activity in physiological signals. Biomedical Signal and Image Processing Springer New York NY USA 2005;.

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

[14] Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: A review. Computer Methods and Programs in Biomedicine 2018;161:1–13. ISSN 0169-2607.

[15] 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.

[16] Zhao Z, Zhang Y. SQI quality evaluation mechanism of single-lead ECG signal based on simple heuristic fusion and fuzzy comprehensive evaluation. Front Physiol 2018;9:727.

[17] Mincholé A, Rodriguez B. Artificial intelligence for the electrocardiogram. Nat Med 01 2019;25(1):22–23.

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