Assessing ECG Signal Quality Using a Pre-trained Audio Network

Álvaro Huerta[1], Oscar-AyoMartín[2], José J Rieta[3], Raúl Alcaraz[1]

[1] Research Group in Electronic, Biomedical and Telecom. Eng., Univ. of Castilla-La Mancha, Spain
[2] Department of Neurology, University Hospital of Albacete, Spain
[3] BioMIT.org, Electronic Engineering Department, Universitat Politecnica de Valencia, Spain

Abstract

Novel ECG wearable systems offer the possibility of continuous cardiac monitoring by acquiring long periods of ECG signal uninterruptedly. However, ECG signals obtained from these wearable devices often suffer from severe noise, requiring automated quality assessment. Recent advancements have demonstrated the effectiveness of convolutional neural networks (CNNs) for that purpose. The transfer learning approach has been widely used in vision and natural language domains, however its application to an ECG signal interpreted as audio data remains relatively unexplored. Therefore, this study aims to analyze the applicability of this technique in the assessment of ECG signal quality. The ECG recordings were conditioned to be inputted to the pre-trained audio network YAMNet, which was fine-tuned to discern between low- and high-quality ECG excerpts. Briefly, after the training process with a proprietary balanced dataset, the model was validated and externally tested using a publicly available database. Results showed a slight unbalance between sensitivity and specificity values, obtaining rates of about 82% and 74% respectively. This performance is similar to that obtained by previous algorithms based on wider and deeper CNN architectures. Hence, audio-based transfer learning seems to work successfully in ECG quality assessment and its use in other healthcare applications could also be explored.

1. Introduction

The diagnosis of many cardiovascular diseases (CVDs) is usually performed through the resting electrocardiogram (ECG), still today one of the most used cardiac tests [1]. Nonetheless, the short ECG recording lasting for approximately 10 to 20 seconds is insufficient for detecting intermittent heart conditions like paroxysmal atrial fibrillation (AF). Although traditional ambulatory systems can record ECG signals over several days, recent studies have shown that the success rate of detection of this arrhythmia is related to the length of the recording [2]. According to this scenario, novel wearable devices able to continuously monitor ECG for weeks or even months without interfering with the patient's daily activities offer a significant opportunity. They can enhance the early detection of hidden AF and reduce its rapidly increasing prevalence [3].

Although these novel technologies have presented great potential, the quality of the ECG signal is compromised when it is obtained during the patient's daily activities [4]. Thus it is imperative to address the problem of assessing the quality of the ECG recordings in order to reduce misinterpretation and the large number of false alarms of AF [5]. Nevertheless, the vast amount of data makes it unfeasible to manually visualize entire signals. Because of this, cutting-edge techniques for automatically assessing the quality of long-term ECG recordings have achieved substantial advances, making artificial intelligence systems the newest approach in this task's development.

The most recent works have utilized pre-trained convolutional neural networks (CNNs) to assess the ECG signal quality automatically [6], as opposed to more traditional methods based on manual feature extraction or fiducial point detection [7]. One of the most common ways to use CNNs is under the paradigm of transfer learning, i.e. using the knowledge of training CNNs and applying it to a new task. Well-known CNN models were pre-trained by inputting huge datasets from natural images, and the fine-tuning process was usually developed by turning the ECG excerpts into images. In this sense, several time-frequency transformations such as continuous wavelet transform or phase space plots have been proposed [8, 9].

Nonetheless, some very recent CNN models have been pre-trained by using large audio datasets instead of images. Usually, these networks have been fed by 2-D transformations of audio data and previously acquired knowledge during pre-training could be suitable for assessing the ECG signal quality, but this alternative remains unexplored. Recently, a new model of a pre-trained audio network called YAMNet achieved promising results dealing with audio signals in healthcare applications [10]. Therefore, the proposed work aims to explore the abilities of YAMNet to assess automatically the quality of ECG recordings once this signal has been properly conditioned.

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.449

2. Materials and methods

2.1. Databases and data pre-processing

The performance of a new algorithm needs a robust testing framework to obtain realistic outcomes and avoid biased results attached to training and testing procedures [11]. Thus, during the training stage, a proprietary database (PDB) was initially used for the fine-tuning of YAMNet. The ECG signals in this database were obtained from 36 patients (16 men and 20 women) using an electrode-equipped vest able to capture signals with a sampling frequency of 250 Hz and a resolution of 12 bits over a dynamic range of ± 5 mV. Labeling of ECG segments involved specialized clinicians who categorized the data into two classes: high- and low-quality. Thus, ECG intervals were tagged as noisy when all R-peaks in a segment could not be confidently distinguished. Additionally, it was acknowledged that the high-quality class included AF episodes, normal sinus rhythm (NSR) and other rhythms (OR). Consequently, the PDB dataset comprised 20,000 5-second-length ECG intervals, evenly distributed between classes. In the high-quality group there are 7,650, 1,750 and 600 5-second-length ECG excerpts corresponding with NSR, AF and OR respectively.

For external testing of the proposed algorithm, the well-known and publicly available PhysioNet/CinC Challenge 2017 database (PC2017DB) was used [12]. Significant differences between the acquisition system used in this database exist with respect to PDB. The main dissimilarity is that an under-demand device was used to collect PC2017DB, i.e., it only acquires a signal when the patient requests it. Moreover, this recording instrument captures the electrical activity of the fingertips at a frequency of 300 Hz and with a resolution of 16 bits. These differences in the acquisition of ECGs and consequently in their morphologies make the training and testing process of the proposed algorithm very challenging and result in a very realistic view of the generalisability of the algorithm. Additionally, the available data was categorized into two groups: noisy signals belonged to the low-quality class, while AF, NSR, and OR were included in the the high-quality class as above with PDB. Consequently, after segmenting and processing, 48,607 5-second-length ECG excerpts were considered for analysis.

On the other hand, as YAMNet is a network designed to extract data from audio signals, all 5-second-length ECG intervals were properly pre-processed. First, the amplitude of every ECG excerpt was normalized through a rescaling process clipping the input data to the range -1 and 1. In this way, the shape of the ECG signals remains the same. Second, the ECG excerpts were up-sampled to 16 kHz to fit audio requirements.

2.2. YAMNet architecture

YAMNet is a deep neural network model trained on the AudioSet-YouTube corpus to discern between 521 audio event classes [13, 14]. It is based on the version 2 of the architecture Mobilenet [15], which is built on depth-wise separable convolutions, with the exception of the initial layer that is a standard convolution. This simplistic network definition facilitates the exploration of new topologies to identify an optimal network configuration. Moreover, all layers are followed by a batch normalization and rectifier linear units nonlinearity, except for the last fully connected layer which lacks nonlinearity and feeds into a softmax layer for classification purposes [14]. Therefore, the architecture of YAMNet has a convolution layer after the input layer. These two layers involve a previous step to a cascade of point-wise and depth-wise separable convolution layers. This new approach deal with audio waveforms with an arbitrary length, sampled at 16 kHz.

Despite YAMNet is designed to extract information from audio signals, this audio data is transformed into a spectrogram using a Short-Time Fourier Transform (STFT) with a periodic Hann window, a window hop of 10 ms, and a window size of 25 ms. Subsequently, a mel-spectrogram is computed with the information of the STFT, mapping to 64 mel bins, which span the frequency range 125-7500 Hz. Thus, this mel-spectrogram is time-frequency representation on the mel-scale, which is a perceptual scale of pitches judged by listeners to be equal in distance from one another. In other words, this mel-spectrogram is a 2-D representation where one axis represents time, the other axis represents frequency, and the color represents the magnitude (amplitude) of the observed frequency at a particular time [16]. A final matrix with a dimension of 96×64 is obtained and used to feed the CNN. An example of a typical high-quality 5-second-length ECG segment along with its mel-spectrogram is depicted in Figure 1 (a) and (c) respectively. In the same way, the Figure 1 (c) shows a 5-second-length ECG interval belonging to the low-quality class alongside its corresponding mel-spetrogram (d).

2.3. Experimental setup

For ECG quality assessment, the knowledge acquired by YAMNet in its pre-training was maintained [13]. Nonetheless, its last fully-connected layer was modified for providing a two-classes classification. Moreover, the network was fine-tuned on the PDB.

In short, following a typical holdout approach, 80% of the training set was utilized for the fine-tuning process and the remaining 20% for the validation and early stoping when accuracy was maintained or decreased for five consecutive epochs. The fine-tuning stage was configured uti-

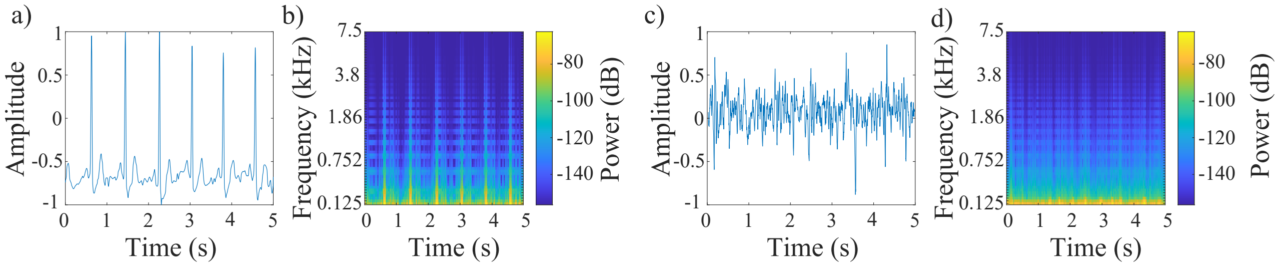

Figure 1. A 5-second-length ECG excerpt labeled as high-quality class (a) along with its mel-spectrogram (b).

lizing a batch size of 128 images and 20 epochs, applying a shuffle of the samples at every epoch. Finally, the model was tested on the PC2017DB. Besides classical statistical values of sensitivity (Se), specificity (Sp), balanced accuracy ($BAcc$), and F_1-score. Moreover, within the high-quality group, the correct classification rate of the different rhythms was also obtained. Thus, the index $\mathcal{R}_{\mathcal{NSR}}$ computed the number of ECG excerpts belonging to NSR properly assessed, and the indicator $\mathcal{R}_{\mathcal{AF}}$ evaluated the percentage of ECG intervals from AF rightly detected, and finally the value $\mathcal{R}_{\mathcal{OR}}$ measured the success rate of ECG branches included in OR. This last information was particularly important to understand how the classifier behaved when there were different rhythms within a database.

3. Results

Table 1 presents classification results obtained by YAMNet in validation and testing phases. As can be seen, whereas the validation outcomes exceeded 93% in all cases, even reaching 99% in Se, Sp, $BAcc$, F_1-score, and $\mathcal{R}_{\mathcal{AF}}$ values, the testing results were slightly lower. Although the metric F_1-score that combines precision and recall surpassed 90%, a notable unbalance of about 8% was observed between Se and Sp metrics, which involves a performance fall about 18% and 25% respectively. Moreover, related to the outcomes within the high-quality class, the well-detected ECG segments belonging to AF rhythm obtained more than 83% of success rate. This value is modestly superior to those obtained when the rhythms belong to NSR and OR in which the values were about 5% and 6% lower respectively.

4. Discussion

The utilization of a new audio-based pre-trained network to assess the quality of ECG signals has proven to be an effective alternative. Thus, the knowledge of YAMNet acquired by using the large AudioSet-YouTube corpus [14] has allowed this new approach to reach Se and Sp values above 99% in the validation phase, and Se and Sp of about 82% and 74% respectively in the testing stage. It was also noted that the model is prone to classify the ECG segments as high-quality class, possibly because the network was able to identify a common pattern on high-quality ECG excerpts. In contrast, noisy segments are more chaotic and unpredictable. Regarding the results within the high-quality group, the same trend was observed in the validation and testing sets since the $\mathcal{R}_{\mathcal{AF}}$ index performed better in both cases, whereas the $\mathcal{R}_{\mathcal{OR}}$ value dropped its output slightly. This is a very valuable result since AF segments usually have common morphological and frequential features with some kinds of noise present in ECG recordings.

Notwithstanding the lack of previously published literature since YAMNet has been recently designed, the present outcomes are in line with previous work dealing with ECG signal quality assessment when tested with the same database, as can be seen in Table 2. In the work proposed by [9] et al., a phase space plot (Poincaré) was generated to be the input of a pre-trained CNN model, reaching Se and Sp values of about 81% and 85% respectively, showing the opposite trend than the proposed work, but retaining F_1-score outcomes with a difference below 0.5%. On the other hand, when the quality assessment was performed with traditional machine learning techniques instead of CNNs, comparable findings in the testing stage were encountered. For instance, following the approach proposed by Albaba et al. [17], the obtained values of Se and Sp were about 73% and 80% respectively. However, the F_1-score falls by around 5%. Likewise, according to the guidelines suggested in the paper of [18] the outcomes presented a large imbalance of Se and Sp about a 20%, also showing a lower value of F_1-score than previous approaches.

5. Conclusions

A new audio-based pre-trained network has been proposed as an effective alternative for assessing the quality of the ECG signals, obtaining comparable results to previous works. The utilization of pre-trained networks under the paradigm of transfer learning could be an efficient alternative in other health applications.

Table 1. Statistical values for the validation and testing process using YAMNet.

Set	Se	Sp	$BAcc$	F_1	$\mathcal{R}_{\mathcal{NSR}}$	$\mathcal{R}_{\mathcal{AF}}$	$\mathcal{R}_{\mathcal{OR}}$
Validation	0.991	0.999	0.995	0.995	0.942	0.998	0.938
Testing	0.825	0.741	0.783	0.901	0.817	0.834	0.811

Table 2. Comparative results with reference methods.

Approach	Se	Sp	$BAcc$	F_1
Poincaré plot [9]	0.813	0.854	0.833	0.895
Albaba et al. [17]	0.739	0.804	0.771	0.847
Behar et al. [18]	0.712	0.890	0.801	0.830

Acknowledgments

This research has received financial support from the company Daiichi Sankyo S.A.U. and from public grants PID2021-128525OB-I00, PID2021-123804OB-I00, and TED2021-130935B-I00 funded by MICIU/AEI/10.13039/501100011033, FEDER EU, and NextGeneration EU/PRTR, grant SBPLY/21/180501/000186 funded by Junta de Comunidades de Castilla-La Mancha, and grand AICO/2021/286 funded by Generalitat Valenciana.

References

[1] Niu J, Deng C, Zheng R, Xu M, Lu J, et. al. The association and predictive ability of ECG abnormalities with cardiovascular diseases: A prospective analysis. Glob Heart Sep 2020;15(1):59.

[2] Kalarus Z, Mairesse GH, Sokal A, Boriani G, Średniawa B, et. al. Searching for atrial fibrillation: looking harder, looking longer, and in increasingly sophisticated ways. An EHRA position paper. Europace Feb 2023;25(1):185–198.

[3] Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge. Int J Stroke Feb 2021;16(2):217–221.

[4] 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.

[5] Ding EY, Marcus GM, McManus DD. Emerging technologies for identifying atrial fibrillation. Circ Res Jun 2020; 127(1):128–142.

[6] 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.

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

[8] 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.

[9] Huerta Á, Martinez-Rodrigo A, Bertomeu-González V, Ayo-Martin Ó, Rieta JJ, Alcaraz R. Single-lead electrocardiogram quality assessment in the context of paroxysmal atrial fibrillation through phase space plots. Biomed Signal Process Control 2024;91:105920.

[10] Grooby E, Sitaula C, Fattahi D, Sameni R, Tan K, Zhou L, King A, Ramanathan A, Malhotra A, Dumont GA, et al. Real-time multi-level neonatal heart and lung sound quality assessment for telehealth applications. IEEE Access 2022; 10:10934–10948.

[11] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. Circ 2015;131(2):211–219.

[12] 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.

[13] Plakal M, Ellis D. Yet Another Audio Mobilenet Network (YAMNet). GitHub repository. Accessed on 08th july 2024.

[14] Gemmeke JF, Ellis DP, Freedman D, Jansen A, Lawrence W, Moore RC, Plakal M, Ritter M. Audio set: An ontology and human-labeled dataset for audio events. In 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2017; 776–780.

[15] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv170404861 2017;.

[16] Rabiner L, Schafer R. Theory and Applications of Digital Speech Processing. 1st edition. USA: Prentice Hall Press, 2010. ISBN 0136034284.

[17] Albaba A, Simões-Capela N, Wang Y, Hendriks RC, De Raedt W, Van Hoof C. Assessing the signal quality of electrocardiograms from varied acquisition sources: A generic machine learning pipeline for model generation. Comput Biol Med Mar 2021;130:104164.

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

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