Relevance of Pre-Training in the Development of a Light Convolutional Neural Network for ECG Quality Assessment

Álvaro Huerta[1], Arturo Martinez-Rodrigo[1], José J Rieta[2], Raúl Alcaraz[1]

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

Abstract

Long-term cardiac monitoring with conventional technology requires huge resources. This problem can be addressed using portable devices. However, they acquire ECGs hardly disturbed by noise, and accurate quality assessment (QA) of these recordings is crucial. Recently, pre-trained convolutional neural networks (CNNs) have reported promising performance in that context, but they demand a lot of computational resources, limiting their use in portable systems. Hence, this work aims to explore the ability of a lightweight CNN, with far fewer parameters than well-known pre-trained CNNs. Applying transfer learning is common in many classification tasks, so the performance of the network was compared when trained from scratch and when pre-trained on a freely available set of natural images. Moreover, the well-known GoogLeNet model was also considered for comparison. All CNNs were fine-tuned with a balanced set of 20,000 5 second-length ECG segments and validated externally. Results show that GoogLeNet performed slightly better than the pre-trained lightweight CNN. However, proposed models were about 12 times faster to classify each ECG interval. These findings highlight the suitability of pre-training a CNN using natural images, retaining a comparable performance in ECG QA than much deeper networks but reducing notably the computational cost.

1. Introduction

The resting ECG has become the standard test for diagnosing the most common cardiovascular diseases (CVDs). However, it is insufficient for determine cardiac abnormalities of an intermittent and unpredictable nature, which require continuous monitoring [1]. Thus, the utilization of novel wearable devices that enable continuous ECG recordings over extended periods has recently emerged [1]. Therefore, these new methods for screening diverse CVDs, such as paroxysmal atrial fibrillation (AF) which is the most common type, and it is associated with high morbidity and a risk factor for ischemic stroke [2] represent a short-term breakthrough.

Despite the high possibilities of these novel systems, acquiring the ECG signal during the patient's everyday activity compromises its overall quality [3]. To minimize misinterpretation of the ECG recordings, it is critical to address the issue of automatically evaluating their overall quality, especially under severe disturbances associated with the acquisition environments, i.e. motion artifacts, impulse noise, or electrode contact noise [4]. Not only the environment can corrupt the ECG signal but also typical disturbances inherent to acquisition systems such as powerline interference, muscular contractions, and baseline drifts. Therefore, the first step in any ECG processing system must be to assess the signal quality.

Unfortunately, the vast volume of data registered makes the manual visualization of whole signals unfeasible. For this reason, state-of-the-art methods for automatically assessing the quality of long-term ECG recordings have shown significant advancements making the artificial intelligence systems the latest trend in developing this task. Despite classical signal-processing algorithms were based on extracting fiducial points and morphological characteristics from ECG recordings using machine learning techniques obtaining promising results [5], recent deep learning algorithms have surpassed those approaches. It has been demonstrated that deep learning methods have superior generalization capabilities, hence they present a greater ability to distinguish between high- and low-quality ECG segments. Furthermore, these algorithms may deal with raw ECG data without requiring additional phases of pre-processing or feature extraction [6].

Among the artificial intelligence methods, convolutional neural networks (CNNs) have been established as the most common approach for several tasks such as classification, regression, or detection. The large volume of data necessary for training a CNN from scratch necessitates the use of transfer learning techniques, which is the process of fine-tuning the knowledge obtained by previously pre-trained networks on a new task [7]. Several research teams have developed and shared their own CNN models such as GoogLeNet [8]. However, these models typically have millions of hyperparameters. For this reason, the proposed work aims to investigate how CNNs perform in relation to pre-training for the specific task of assessing the quality of ECG signals [9]. Moreover, a lightweight CNN structure with a limited number of layers was developed to assess its performance in comparison to GoogLeNet.

2. Materials and methods

2.1. Databases

The material selected to develop the proposed work can be divided into two main groups. Firstly, the dataset utilized for training the Light-CNN from scratch is based on the original set from ImageNet [10] but is limited to 70 classes of various objects, each with 500 samples, reaching a total of 35,000 samples. Since computational and time resources required for training a CNN are substantial, this option was selected as the best trade-off between data and resources.

On the other hand, the second group comprises the dataset belonging to ECG recordings. To ensure the generalization capability and to prevent overfitting, two separate ECG databases acquired from different devices were utilized i.e. a proprietary database (PDB) and the public PhysioNet/CinC Challenge 2017 database (PC2017DB) [11]. The acquisition system for the PDB consisted of a textile vest with a sampling rate of 250 Hz with 12 bits of resolution. Two experts labeled the different high- and low-quality ECG segments, identifying as noisy signals those ones where the R-peaks could not be clearly differentiated. This dataset contains 20,000 5 second-length ECG excerpts which were evenly divided across the two groups. It is worth noting that the high-quality class includes both normal sinus rhythm (NSR) and AF episodes. This dataset was used to develop the fine-tuning process in the three different proposed attempts.

The open-source data from PC2017DB is utilized as a testing set for each conducted experiment. The acquisition system consists of a portable device that registers the electrical activity of the heart on demand using fingerprints. The sampling frequency is set as 300 Hz with a resolution of 16 bits. The dataset includes four different labels, i.e. AF, NSR, other rhythms, and noisy signals. This dataset was formed of 48,607 5 second-length ECG segments that were severely unbalanced between the two groups. Specifically, 47,439 belong to the high-quality class and 1,168 belong to the low-quality class.

2.2. ECG transformation

Following the guidelines of previously published papers [6], CNNs were utilized to discern between high- and low-quality segments in long-term ECG recordings [6,12]. Traditionally, CNNs have a special ability to deal with 2-D matrices as inputs, and GoogLeNet is not an exception [8]. For this reason, a previous step that converts ECG recordings into images is mandatory. Hence each 5 second-length raw ECG signal were turned into images using a Continuous Wavelet Transform (CWT) and applying a colormap [13]. The choice of this transformation is deliberated since it has been demonstrated that CWT enhances aspects of quasiperiodic non-stationary signals, such as ECG [13]. The main features of CWT configuration can be found in [6].

2.3. GoogLeNet architecture

Among the pre-trained networks available for the free utilization of transfer learning, GoogLeNet is one of the most extended and utilized [8]. It becomes famous for winning the ImageNet Challenge 2,014 [10] and has the ability to distinguish among 1,000 different image classes. On the other hand, the main novelty of this architecture is the composition of several parallel branches, in contrast to previous structures that stack exclusively a series of layers. This approach allows increasing the depth reducing the computational cost. The *inception module* is the pivotal key of this model, which is a block composed of kernels with different sizes (1×1, 3×3, 5×5) placed in parallel. In this way, the spatial information belonging to different scales is extracted, either fine or coarse grain levels. Basically, GoogLeNet is formed by nine inception modules, besides two convolutional layers, four max-pooling layers, three average pooling layers, five fully-connected (FC) layers, and three soft-max layers. In addition to that, the activation function as rectifier linear units (ReLU) or dropout regularization is part of the architecture. Thus, a depth of 22 learnable layers and 6.7 million parameters are the features of GoogLeNet.

2.4. Proposed Light-CNN architecture

To start with, it is necessary to define what qualifies as a lightweight CNN. For reference, a well-known design is *SqueezeNet*, which contains around 1.24 million parameters [14]. The proposed architecture is composed of four layers with learning ability, where three are convolutional, and the last one is a FC layer. The first convolutional layer has a spatial dimension of 7×7 with 128 kernels with a stride of 5. The second convolutional layer reduces the size of the first one and uses filters of dimension 4×4 with 128 kernels. The last convolutional layer is composed of

Table 1. Description of the full architecture of the proposed Light-CNN. The layer column refers to the given names of each layer. As can be seen in the table, only four layers have learnable parameters.

Layer	Type	Activations	Learnables
Input	Image input	$227 \times 227 \times 3$	
Conv. 1	2-D convolution	$45 \times 45 \times 128$	18,944
ReLU 1	Activation funct.	$45 \times 45 \times 128$	
MP 1	2-D max. pooling	$22 \times 22 \times 128$	
Conv. 2	2-D convolution	$10 \times 10 \times 128$	262,272
ReLU 2	Activation funct.	$10 \times 10 \times 128$	
MP 2	2-D max. pooling	$9 \times 9 \times 128$	
Conv. 3	2-D convolution	$7 \times 7 \times 256$	295,168
ReLU 3	Activation funct.	$7 \times 7 \times 256$	
MP 3	2-D max. pooling	$6 \times 6 \times 256$	
FC	Fully-connected	$1 \times 1 \times 70$	645,190
Softmax	Softmax	$1 \times 1 \times 70$	
Class.	Output	$1 \times 1 \times 70$	

384 kernels with a size of 3×3 and 256 kernels. After that, the last learnable layer is a FC layer with 70 neurons, corresponding with the number of classes able to classify. Moreover, the convolutional layers are followed by ReLU as activation functions and a max pooling layer to reduce the spatial dimensions of the feature map. Behind the FC layer, a softmax classifier assigns the probabilities to each class prior to the output layer. A detailed analysis of the whole architecture along with the activations and learnable parameters in each layer are presented in Table 1. Finally, the proposed architecture has a total amount of 1,221,574 parameters and a depth of 4 layers.

2.5. Performance analysis

In consequence, the experiment has three different configurations: the Light-CNN without pre-training; the Light-CNN trained with the set of ImageNet 70; and finally, GoogLeNet, trained with the whole set of ImageNet. All three approaches were submitted to a fine-tuning process using the PDB dataset. Usually, to apply the transfer learning technique, the architecture of CNNs needs to be adapted to the new task. Whereas the Light-CNN without pre-training was designed to differentiate two classes, the pre-trained Light-CNN was reconfigured to classify from 70 down to only two classes. Moreover, GoogLeNet changed the last part of the network, readjusting the number of neurons from 1,000 to 2 i.e. to distinguish between high- and low-quality ECG excerpts. Note that in the three conducted experiments, the training stage was repeated 10 times using a training batch size of 32 and 10 epochs, while averaging the classification results obtained in the testing phase. Approximately, the testing set comprises 98% of excerpts belonging to the high-quality class whereas only about 2% belong to the low-quality class. For this reason, besides the typical statistics of sensitivity (Se), specificity (Sp), and accuracy (Acc), the balanced accuracy was also obtained ($BAcc$). The execution time of processing each 5 second-length ECG interval was measured to assess the performance.

3. Results

Table 2 depicts the outcomes for each conducted experiment along with the classification time for the three CNN-based models. As shown, GoogLeNet achieves slightly better classification rates than the Light-CNN when the latter is pre-trained. However, it has been demonstrated that the lightweight models are about 12 times faster in classifying each 5 second-length ECG excerpt than GoogLeNet. Conversely, the model with empty weights has reported a drop in all values of Se, Sp, Acc, and $BAcc$. Balanced values of Sp and Se are always desired, hence only a difference of about 5% was observed among the pre-trained models whereas this difference increases to 16% in the model without pre-training.

4. Discussion

The outcomes have shown that networks that have been pre-trained using a set of diverse natural images improve their ability to generalize when used for different classification tasks. In the case of GoogLeNet, the ECG quality assessment significantly improved the classification performance compared with pre-trained Light-CNN. There is a strong relationship between the dataset utilized for the pre-training and the final results. Whereas GoogLeNet was originally designed to distinguish between 1,000 different classes, the Light-CNN model was designed to classify between only 70 classes. For this reason, better values of Se were provided by GooglLeNet since the high-quality class is more difficult to detect than the low-quality class. Consequently, the Light-CNN without pre-training has a lower ability to generalize and trends to achieve better values of Sp, since the classifier is prone to assigning ECG excerpts as low-quality class due to the random nature of ECG segments affected by noise or artifacts. Moreover, another interesting finding is that the Light-CNN can perform in a similar way as deeper networks. The proposed architecture has only 4 learnable layers, whereas GoogLeNet has 22 layers; similarly, the suggested structure has 5 times fewer parameters than GoogLeNet, which as expected, means a much shorter processing time (12 times).

Even though it is uncommon to find works with the same aim in the literature, the results obtained in previous studies proposing lightweight CNN architectures are aligned with the present research. For instance, Garg and Singh compared the performance of a proprietary lightweight

Table 2. Classification outcomes obtained by the three different CNN models along with the time assessing each 5 second-length ECG interval.

Network	Se	Sp	Acc	BAcc	Time/ECG interval (ms)
Light CNN (Empty weights)	0.776	0.934	0.780	0.855	0.65
Light CNN (Pre-trained on ImageNet 70)	0.836	0.884	0.836	0.860	0.64
GoogLeNet (ImageNet)	0.887	0.841	0.886	0.864	7.94

CNN model with previous pre-trained CNN architectures such as ResNet-50, Inception-V4, or MobileNet-V2 [15]. They achieved similar output results, even outperforming previous works in the context of breast cancer classification. Additionally, the execution time was reduced since fewer parameters were involved. In the same way, Huang and Liao developed a lightweight CNN structure concerning AlexNet and EfficientNetV2 aimed to detect COVID-19 on X-ray and CT images [16]. The algorithm was tested on several image datasets, and they obtained comparable classification values than previous works, using a CNN architecture composed of less than 800,000 parameters.

5. Conclusions

Developing and training a lightweight CNN from scratch for assessing the quality of long-term ECG recordings could be both a very challenging and time-consuming. Nonetheless, it is worth because of the favorable trade-off between speed and performance compared to heavier well-known complexes pre-trained CNNs.

Acknowledgments

This research has received financial support from public grants PID2021-00X128525-IV0 and PID2021-123804OB-I00 of the Spanish Government 10.13039/ 501100011033 jointly with the European Regional Devel-opment Fund (EU), SBPLY/17/180501/000411 and SB-PLY/21/180501/000186 from Junta de Comunidades de Castilla-La Mancha, and AICO/2021/286 from Generalitat Valenciana, as well as from the company Daiichi Sankyo S.A.U.

References

[1] Sana F, Isselbacher EM, Singh JP, Heist EK, Pathik B, Armoundas AA. Wearable devices for ambulatory cardiac monitoring: JACC state-of-the-art review. J of the Amer College of Card 2020;75(13):1582–1592.

[2] 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.

[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] Nagai S, Anzai D, Wang J. Motion artefact removals for wearable ECG using stationary wavelet transform. Healthc Technol Lett 2017;4(4):138–141.

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

[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] Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J of Big data 2016;3(1):1–40.

[8] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In Proc. of the IEEE conf. on comput. vision and pattern recogn. 2015; 1–9.

[9] Krishna ST, Kalluri HK. Deep learning and transfer learning approaches for image classification. Int J of Recent Tech and Eng IJRTE 2019;7(5S4):427–432.

[10] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L. ImageNet Large Scale Visual Recognition Challenge. Int J of Comput Vision IJCV 2015;115(3):211–252.

[11] 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.

[12] 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.

[13] 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.

[14] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. arXiv preprint arXiv160207360 2016;.

[15] Garg S, Singh P. Transfer learning based lightweight ensemble model for imbalanced breast cancer classification. IEEEACM Trans on Comput Biol and Bioinf 2022; 20(2):1529–1539.

[16] Huang ML, Liao YC. A lightweight cnn-based network on COVID-19 detection using X-ray and CT images. Comput in Biol and Medicine 2022;146:105604.

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