A Deep Learning Solution for Automatized Interpretation of 12-Lead ECGs

Álvaro Huerta1, Arturo Martinez-Rodrigo1, José J. Rieta2, Raúl Alcaraz1

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

A broad variety of algorithms for detection and classification of rhythm and morphology abnormalities in ECG recordings have been proposed in the last years. Although some of them have reported very promising results, they have been mostly validated on short and non-public datasets, thus making their comparison extremely difficult. PhysioNet/CinC Challenge 2020 provides an interesting opportunity to compare these and other algorithms on a wide set of ECG recordings. Hence, the goal of this work is to introduce a novel method based on deep learning and beat-to-beat classification to participate in this competition. The algorithm is based on segmenting all beats in the 12-lead ECG recording and generating a new signal for each one by concatenating sequentially the information found in each lead. The resulting signal is then transformed into a 2-D image through a continuous Wavelet transform and inputted to an 8-layer convolutional neural network (CNN). Analyzing more than 130,000 images from the dataset proposed for training, nine different rhythm or morphology abnormalities were classified. According to the competition guidelines, classification results were evaluated in terms of a class-weighted F-score ($F_β$) and a generalization of the Jaccard measure ($G_β$). In average for all training signals, these metrics were 0.933 and 0.811, respectively. Regarding validation on the testing set, mean values for both performance indices were 0.654 and 0.372, respectively.

1. Introduction

Cardiovascular diseases (CVDs) embrace a set of disorders affecting blood vessels and the heart, and usually appear in the form of cerebrovascular disease, peripheral artery disease, hypertension, coronary heart disease, heart failure, and rheumatic heart disease. Nowadays, they are still the most common cause of morbidity and mortality worldwide [1]. In fact, more than 30% of all deaths are attributable to them [1]. Thus, their early diagnosis may be helpful in preventing premature deaths by enabling successful treatments [2]. To this end, the standard 12-lead ECG signal is an important tool, because it provides information on the heart’s behavior [3]. However, manual interpretation of the ECG is not a trivial task, because a high level of training is required [4]. Additionally, this task can be very time-consuming, especially when patients are monitored for long hours [5]. In fact, it is often difficult to find compelling symptoms of most cardiovascular diseases and arrhythmias in short ECG recordings, thus requiring monitoring for several hours or days [6]. Current wearable systems are even able to obtain ECG recordings during several weeks and months [7], thus making their manual revision unfeasible and requiring computer-based interpretation [5].

In the last years, a broad variety of automatic algorithms have been proposed to detect and classify different rhythm and morphology abnormalities in ECG signals [8]. Although some of them have reported very promising results, they have been mostly validated on short and non-public datasets, thus making their comparison extremely difficult [8]. To palliate this issue, PhysioNet/CinC Challenge 2020 have been carefully designed [9]. This competition offers an attractive opportunity to compare algorithms on a wide set of ECG recordings. Hence, with the aim of participating in this challenge, a novel method for ECG interpretation based on deep learning and beat-to-beat classification is introduced in the present work.

The algorithm takes advantage of the high learning capability presented by the convolutional neural networks (CNN) to extract the most relevant ECG features without delineation of its fiducial points, as well as without any other kind of manual or external intervention [10]. In fact, these networks have been widely used in ECG-based classification tasks, such as identification of atrial fibrillation, classification of arrhythmias, detection of sleep apnea, biometric identification of individuals, or detection of diabetic subjects, among others, providing excellent performance [10–13].

2. Materials

To conduct the competition, 12-lead ECG recordings lasting from a few seconds to several minutes were col-
lected from different databases, which are briefly described below. Thus, the dataset proposed for the China Physiological Signal Challenge 2018 [8] is composed of 10,330 ECG recordings sampled at 500 Hz and with a duration between 6 and 60 seconds. The St Petersburg INCART 12-lead Arrhythmia Database [14] contains 75 30 minute-length ECG signals extracted from longer Holter records, which were acquired with a sampling rate of 257 Hz. The PTB Diagnostic ECG Database [14] consists of 549 10 second-length ECG signals obtained with a sampling frequency of 1 kHz from patients with different CVDs. Similarly, the PTB-XL dataset contains 21,837 10 second-length signals sampled at 500 Hz. Finally, a Georgia database presents 10,344 10 second-length ECG recordings, obtained with a sampling rate of 500 Hz from patients exclusively from the Southeastern United States.

All these ECG recordings were annotated with one or more labels from different types of rhythm and morphology abnormalities, and mainly constituted the training group. The test set was not freely available, but it was composed of samples from the described databases, as well as from one entire new dataset.

3. Methods

The proposed method to detect and classify different abnormalities in 12-lead ECG recordings involved three steps. More precisely, every beat in the signal was first segmented, its time-frequency representation was next obtained, and finally the resulting 2-D image was inputted to a customized CNN. More details about each stage are below provided.

3.1. Beat segmentation

The well-known Pan & Tompkins algorithm [15] was initially used to detect R-peaks in raw lead I and the mean RR interval was then computed to define a time window in which every single beat was get. Next, this time window was applied to the remaining available leads, consecutively concatenating single beats to generate a new signal, where the same beat was represented through all leads. An example visually describing this approach for a sinus rhythm beat is displayed in Fig. 1(a). It should be noted that the considered time window for beat segmentation was centered on each R-peak found in each lead.

3.2. Continuous Wavelet transform

The resulting signal in the previous stage for each beat was transformed into a 2-D image through a continuous Wavelet transform (CWT). The result of this transformation was a matrix composed of wavelet coefficients containing time-frequency information of the original signal. Graphic representation of the absolute value of these coefficients using a proper colormap is known as wavelet scalogram, and has been widely used to feed numerous CNN-based algorithms [10]. For this transformation, a Morlet function was used as mother wavelet and 48 voices per octave were chosen to determine the number of wavelet scales. Fig. 1(b) displays the scalogram obtained for a typical sinus rhythm beat.

3.3. CNN-based classification

To classify each beat, a 8-layer CNN scheme was specifically designed and trained from scratch. For this last purpose, a total of 130,627 beats were randomly extracted from the training subset, among which 9 different classes were noticed. As Table 1 displays, an uneven number of beats was found for each class, but more than 3,000 samples were always available.

Regarding the proposed CNN architecture, its input layer accepted images with a size of 200 × 1000 × 3 pixels. Next, a convolution layer with filter size 7 × 7 was considered, along with a normalization layer which allow the use of higher learning rates giving less consideration to the initialization. Note that a rectifier linear unit was used as activation function. Next, a pooling layer reduced spatial dimension of the feature map without lost relevant information. Subsequently, a fully-connected layer ensured that the nine neurons in the previous layer were connected to all neurons in the current layer. Finally, a softmax function was responsible for assigning probabilities of belonging to each output class.

After a training of 20 epochs with mini-batches of 100 samples, conducted by a stochastic gradient descent approach with a momentum of 0.9 and a learning rate of 0.001, the proposed CNN was ready to classify rhythm and morphology abnormalities in an entire ECG recording. Thus, the probabilities of belonging to the available classes were averaged for all beats found in the signal. Then, the dominant class, i.e., that exhibiting the highest

<table>
<thead>
<tr>
<th>Class</th>
<th># of Beats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial Fibrillation</td>
<td>26,891</td>
</tr>
<tr>
<td>First degree AV block</td>
<td>10,785</td>
</tr>
<tr>
<td>Left bundle branch block</td>
<td>3,482</td>
</tr>
<tr>
<td>Normal sinus rhythm</td>
<td>15,997</td>
</tr>
<tr>
<td>Premature Atrial Contraction</td>
<td>12,671</td>
</tr>
<tr>
<td>Premature Ventricular Contraction</td>
<td>15,464</td>
</tr>
<tr>
<td>Right bundle branch block</td>
<td>27,856</td>
</tr>
<tr>
<td>ST depression</td>
<td>14,291</td>
</tr>
<tr>
<td>ST elevation</td>
<td>3,190</td>
</tr>
</tbody>
</table>
Figure 1. Example of how the proposed method performs on a normal sinus rhythm beat. (a) Concatenation of the segmented beat for each one of the 12 leads, and (b) its time-frequency representation obtained via CWT.

4. Results

The validation of the proposed approach was developed according to the competition guidelines. Thus, two performance metrics were computed, i.e., a class-weighted $F^\beta$-score ($F^\beta$) and a generalization of the Jaccard measure ($G^\beta$). The $F^\beta$ index was a weighted harmonic of precision and recall described in [16]. The $G^\beta$ index was a classical similarity measure on sets with a lot of practical applications in information retrieval, data mining, machine learning, and many others [17].

For an unofficial validation of the proposed method, 3,000 ECG signals from the training subset were randomly selected. In this way, values of $F^\beta$ and $G^\beta$ of 0.933 and 0.811 were respectively obtained. Regarding the results obtained for the official validation (using the testing subset hidden to the challengers), the performance of the algorithm was notably more limited, because values of $F^\beta$ and $G^\beta$ were 0.654 and 0.372, respectively.

5. Discussion

In general terms, recent methodologies proposed to detect rhythm and morphology abnormalities in the ECG signal have reported a promising performance. However, they have been poorly validated by making use of too reduced proprietary datasets. For instance, Li et al. [18] achieved a discriminant ability greater than 97% when a CNN-based method was trained and tested with databases only composed of 24 and 14 subjects, respectively. Similarly, Yildirim et al. [13] also developed another CNN-based algorithm able to classify seventeen different ECG abnormalities with an accuracy of about 90%. However, in this case the method was only validated on a dataset comprising 150 ECG recordings. To obtain a more realistic view of the true capability of these and other methods to identify ECG abnormalities, the PhysioNet/CinC Challenge 2020 [9] has collected and shared a database with more than 50,000 ECG signals, which were obtained under diverse conditions and from different contexts and geographic regions.

To participate in this competition, the present work has introduced an algorithm able to automatically classify normal sinus rhythm and eight types of ECG abnormalities. The method is based on the idea of analyzing an entire ECG recording in a beat-to-beat fashion. In this way, the likelihood of detecting secondary abnormalities, besides the dominant one, could be maximized. However, the obtained classification results are not as good as initially expected. In fact, although values of $F^\beta$ and $G^\beta$ greater than 90 and 80% were respectively reached by analyzing 3,000 ECG signals randomly selected from the training subset, they were lower than 70 and 40% in the official validation phase. This reduction in the performance suggests that the proposed method presents a notable ability to associate the dominant class to each ECG, but its ability to detect secondary abnormalities has still to be significantly improved.

Additionally, the obtained results also point to the fact that the proposed CNN model was not sufficiently trained. In fact, the ECG recordings available in the initial phase of the competition were only used for that purpose. Moreover, in these signals only 9 classes were detected, whereas 27 different diagnoses were used to evaluate the perfor-
mance of the participant methods [9]. Therefore, a wider training of the proposed algorithm could still improve notably its performance.

Another weakness of the algorithm is that its performance relies excessively on the R-peak detection, and therefore is too sensitivity to the presence of noise and artifacts. The incorporation of some components with ability to denoise the signal, as well as to identify those intervals strongly contaminated with artifacts, could also improve the method’s performance. Likewise, the use of more than one technique to detect the R-peaks and only enable further processing when their results were consistent could also prevent erroneous beat segmentation, thus leading to a better classification of ECG abnormalities. These improvements will be addressed in a future work.

6. Conclusions

A novel CNN-based technique to classify different abnormalities in 12-lead ECG signals have been proposed in the framework of the PhysioNet/CinC Challenge 2020. Although the idea of analyzing an entire ECG signal in a beat-to-beat way is interesting to detect more than one abnormality, the method’s performance has proven to be limited. Nonetheless, a variety of improvements could still be considered in the future to increase its classification capability.

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/2019/036 from Generalitat Valenciana and FEDER 2018/11744.

References


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