

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

Álvaro Huerta¹, Arturo Martínez-Rodrigo¹, José J Rieta², Raúl Alcaraz¹

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

² BioMIT.org, Electronic Engineering Department, Universitat Politècnica de València, 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. The present model was created by “ELBIT” team. The algorithm is based on deep learning, and the segmentation of all beats in the 12-lead ECG recording, 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 a convolutional neural network. 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 from the first phase of the challenge, 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 collected from different databases, which are briefly described below. Thus, the dataset proposed for the China Physiolog-

ical 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 scalo-

gram, 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 $200 \times 1000 \times 3$ 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 probability, was assigned to the ECG recording as primary diagnosis. In the case that other classes presented averaged

Table 1. Number of beats for each class found in the subset selected for training the proposed CNN-based method.

Class	# of Beats
Atrial Fibrillation	26,891
First degree AV block	10,785
Left bundle branch block	3,482
Normal sinus rhythm	15,997
Premature Atrial Contraction	12,671
Premature Ventricular Contraction	15,464
Right bundle branch block	27,856
ST depression	14,291
ST elevation	3,190

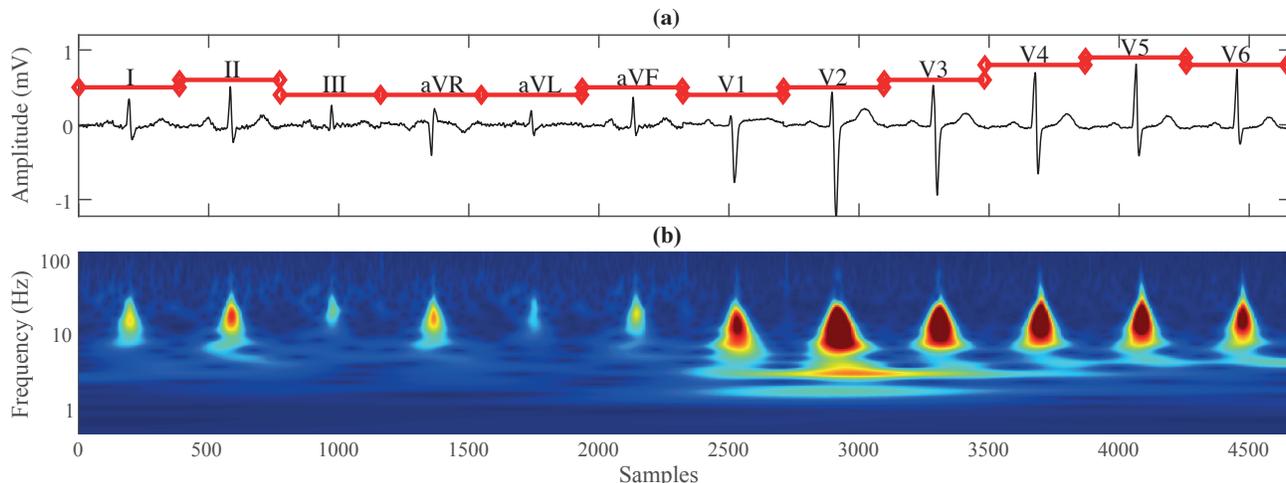


Figure 1. Example of how the proposed method performs on a normal sinus rhythm beat. (a) Concatenation of the segmented beat for each one of the 12 leads, and (b) its time-frequency representation obtained via CWT.

probabilities larger than 10%, they were also considered in descending order as secondary diagnoses.

4. Results

The validation of the proposed approach was developed according to the competition guidelines for the first phase of the Challenge in which only 9 different classes were available. Thus, two performance metrics were computed, i.e., a class-weighted F -score (F_β) and a generalization of the Jaccard measure (G_β). The F_β index was a weighted harmonic of precision and recall described in [16]. The G_β 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_β and G_β of 0.933 and 0.811 were respectively obtained. Although we were unable to participate in the final phase of the Challenge, we obtained official validation results in the first phase (using the testing subset hidden to the challengers), the performance of the algorithm was notably more limited, because values of F_β and G_β 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_β and G_β 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 points 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

- [1] Benjamin EJ, Virani SS, Callaway CW, et al. Heart disease and stroke statistics-2018 update: A report from the American Heart Association. *Circulation* 03 2018;137(12):e67–e492.
- [2] Adams Jr HP, del Zoppo G, Alberts MJ, et al. Guidelines for the early management of adults with ischemic stroke. *Stroke* May 2007;38(5):1655–711.
- [3] Waldmann V, Combes N, Ladouceur M, Celermajer DS, Islerin L, Gatzoulis MA, Khairy P, Marijon E. Understanding electrocardiography in adult patients with congenital heart disease: A review. *JAMA Cardiol* Aug 2020;.
- [4] Bickerton M, Pooler A. Misplaced ECG electrodes and the need for continuing training. *British Journal of Cardiac Nursing* 2019;14(3):123–132.
- [5] Schläpfer J, Wellens HJ. Computer-interpreted electrocardiograms: Benefits and limitations. *J Am Coll Cardiol* Aug 2017;70(9):1183–1192.
- [6] Vavrinsky E, Subjak J, Donoval M, Wagner A, Zavodnik T, Svobodova H. Application of modern multi-sensor holter in diagnosis and treatment. *Sensors Basel* May 2020;20(9).
- [7] Pagola J, Juega J, Francisco-Pascual J, et al. Yield of atrial fibrillation detection with textile wearable holter from the acute phase of stroke: Pilot study of Crypto-AF registry. *Int J Cardiol* Jan 2018;251:45–50.
- [8] Liu F, Liu C, Zhao L, Zhang X, Wu X, Xu X, Liu Y, Ma C, Wei S, He Z, et al. An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection. *Journal of Medical Imaging and Health Informatics* 2018;8(7):1368–1373.
- [9] Alday EAP, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, Liu F, Rad AB, Elola A, Seyedi S, Li Q, Sharma A, Clifford GD, Reyna MA. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. *Physiol Meas* 2020;.
- [10] Ebrahimi Z, Loni M, Daneshlatab M, Gharehbaghi A. A review on deep learning methods for ECG arrhythmia classification. *Expert Systems with Applications X* 2020;100033.
- [11] Acharya UR, Fujita H, Lih OS, Hagiwara Y, Tan JH, Adam M. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Information Sciences* 2017;405:81–90.
- [12] Al Rahhal MM, Bazi Y, Al Zuair M, Othman E, BenJdira B. Convolutional neural networks for electrocardiogram classification. *Journal of Medical and Biological Engineering* 2018;38(6):1014–1025.
- [13] Yıldırım Ö, Pławiak P, Tan RS, Acharya UR. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Computers in Biology and Medicine* 2018;102:411–420.
- [14] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. Physiobank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation* 2000;101(23):e215–e220.
- [15] Pan J, Tompkins WJ. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering* 1985; 32(3):230–236.
- [16] Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European Conference on Information Retrieval*. Springer, 2005; 345–359.
- [17] Kosub S. A note on the triangle inequality for the Jaccard distance. *Pattern Recognition Letters* 2019;120:36–38.
- [18] Li Y, Pang Y, Wang J, Li X. Patient-specific ECG classification by deeper CNN from generic to dedicated. *Neurocomputing* 2018;314:336–346.

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