Rule-Based Method and Deep Learning Networks for Automatic Classification of ECG

Giovanni Bortolan[1], Ivaylo Christov[2], Iana Simova[3]

[1] Institute of Neuroscience IN-CNR, Padova, Italy
[2] Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria
[3] Heart and Brain Center of Excellence, University Hospital, Pleven, Bulgaria

Abstract

The objective of the study is to explore the potentiality of combining a classical rule-based method with a Deep Learning method for automatic classification of ECG for participation in PhysioNet/Computing in Cardiology Challenge 2020. Six databases are considered for training set. They consist 43101 12-leads ECG recording, lasting from 6 to 60 seconds considering 24 diagnostic classes.

The rule-based method is using morphological and time-frequency ECG descriptors, characterizing each diagnostic labels. These rules have been extracted from the knowledge-base of a physician, with no direct learning procedure in the first phase, while a refinement have been tested in the second phase. The Deep Learning method consider both raw ECG signals and median beat signals. These data are processed by continuous wavelet transform analysis obtaining a time-frequency domain representtation, with the generation of specific images. These images are used for training Convolutional Neural Networks for ECG diagnostic classification. Official result of the classification accuracy of the ECGs Test set of our team named 'Gio_Ivo' produced a challenge validation score of 0.325 for the rule based method, and a 0.426 for the Deep learning methodology with GoogleNet, which was chosen for the final score, obtaining a full test score of 0.298, placing us 12th out of 41 in the official ranking.

1. Introduction

Over the years, we have developed methods for detection, accurate localization of ECG waves and automatic diagnosis. For example, different methods have been tested for detection and identification QRS and T-waves and QT interval [1 - 5], for the classification of the heart beats [6, 7] and for the classification of diagnoses of cardiac abnormalities such as atrial fibrillation or myocardial infarction [8 - 10]

Similarly, several techniques of learning with Neural Networks approach or direct learning process from a set of extracted parameters from the analysis of ECG signal were performed and tested [11 - 13].

The main objective of this study is to test these two different techniques of automatic classification of ECG signals with the participation to the PhysioNet/CinC_2020 Challenge. Then, to the classical rule-based system method, a new technique based on direct learning from ECG raw data through the Deep Learning methods are explored and compared in the same framework.

2. Challenge database

The Challenge provided a dataset with annotated 43101 12-leads ECG recordings lasting from 6 to 60 seconds. [14]. The initial 107 diagnoses were further reduced to the 27 diagnostic classes considered in the Physionet/Challenge scoring system (see Table 3 in [14] for a full list of the diagnoses and codes). They reduce to 24 considering 3 equivalent classes. The six considered dataset for the learning phase consists of:

1. DB1- 6,877 recordings from China Physiological Signal Challenge in 2018
2. DB2- 3,453 recordings from China 12-Lead ECG Challenge.
3. DB3- 74 recordings from the St Petersburg INCART 12-lead Arrhythmia
4. DB4- 516 recordings from the PTB Diagnostic ECG.
5. DB5- 21,837 recordings from the PTB-XL electrocardiography Database
6. DB6- 10,344 recordings from a Georgia 12-Lead ECG Challenge

All ECG data are resampled at 500 Hz (if necessary) for compatibility purposes.

This large dataset is consisting for a total of 43101 ECG recordings and 60373 diagnostic instances.

A random selection of ECG records with at most N_max instances for all the 24 considered diagnostic classes were determined for a more equilibrated distribution of their consistency and for a more efficient learning phase,

Table 1. Distribution of the 60373 diagnostic instances (INST) present in the entire database, and weighted number of records (w_S) of the subset S16K.

Code	INST	w_S	Code	INST	w_S
01 IAVB	2394	1536	13 PAC*	1944	884
02 AF	3475	1626	14 PVC*	553	333
03 AFL	314	275	15 LQT	1513	713
04 Brady	288	269	16 LPR	340	140
05CRBBB*	3085	737	17 QAb	1013	523
06 IRBBB	1611	1022	18 RAD	427	207
07 LAnFB	1806	972	19 SA	1240	641
08 LAD	6086	1477	20 SB	2359	568
09 LBBB	1041	468	21 NSR	20846	1000
10 LQRSV	556	441	22 STach	2402	640
11 NSIVCB	997	575	23 Tab	4673	455
12 PR	299	286	24 TInv	1112	214
			TOTAL	60373	16002

(* Equiv: CRBBB & RBBB, PAC & SVPB and PVC & VEB)

obtaining the following subsets:
S16K: N_max=1000 produces 16002 ECG records
S20K: N_max=1500 produces 20.044 ECG records
S11K: N_max=600 produces 11210 ECG records.
Table 1 reports the distribution of the 60373 diagnostic instances present in the entire database, and the weighted number of records of the subset S16K, consisting of 16002 ECG records.

3.   Method

The ECG recordings are filtered to eliminate the power-line interference, the drift of the zero-line and the electromyographic noise (EMG). QRS detection using combined adaptive threshold [1] is performed. Then a robust average beat is calculated, with reference to the most distinguished positive (R), or negative (S) peaks, by signal-averaging of the sustained beats in the record. The rejected outliers are suspected as artifacts or abnormal beats with non-sustained amplitudes.

3.1   Rule-based method

The rule-based method uses morphological and time-frequency ECG –VCG descriptors, characterizing each diagnostic labels. Orthogonal (X, Y, Z) leads were derived by transverse formulas of Dower, [15]. These rules have been extracted from the knowledge-base of a physician, with no direct learning procedure in the first phase, while a refinement has been tested in the second phase.

The way to determine some diagnosis is shown below. In case P-wave does not exist, the number of waves between two QRS in lead II is computed, by counting the number of zero line crossings of the first derivative as shown in Fig. 1. In case of Count>9, AFL=1 (atrial flutter), and in case Count>15, AF=1 (atrial fibrillation),

The QRS-angle in frontal plane was calculated. LAD (left axis deviation) is characterized by QRS angle in the range [-45, 60], while RAD (right axis deviation) is characterized by the QRS angle in the range [180, -44]. In addition, LAnFB (Left anterior fascicular block) is characterized by RAD=1 or LAD =1 and by the presence of small Q-waves with tall R-waves in lead I and small R waves with deep S waves in leads II, III.
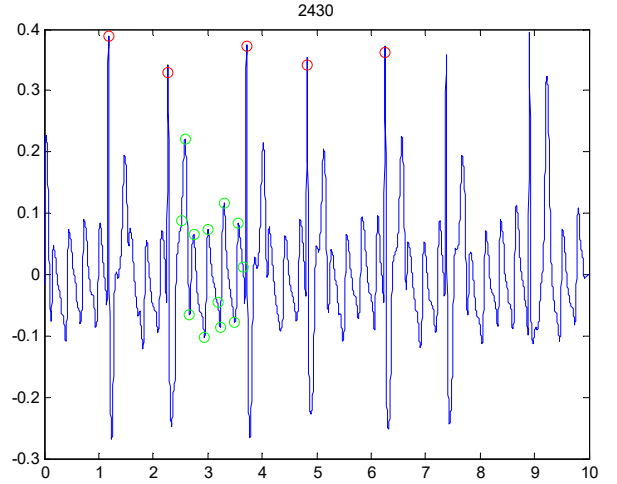

Figure 1. Detection of AF

3.2   Deep Learning Networks

The Deep Learning method considers both raw ECG signals and median beat signals. These data are processed by continuous wavelet transform analysis obtaining a time-frequency domain representation, with the generation of specific images. These images are then used for training Convolutional Neural Networks for ECG diagnostic classification. Pretrained image CNN classification network that has already learned to extract powerful and informative features from natural images has been used as a starting point to learn a new classification task [16].

Two pre-trained CNN for image classification have been initially used, GoogleNet and SqueezeNet networks. These are models pretrained on a subset of the ImageNet database, which is used in the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [16 - 18]. Both networks have been tested in the first phase, while the GoogleNet was used in the second phase considering the better performance.

GoogleNet is a convolutional neural network characterized by 22 layers, and it is pretrained to classify images into 1000 object categories Each layer can be considered as a filter, consequently the first ones characterize more common features while the deeper ones characterize more specific features in order to differentiate the considered diagnostic classes.

In order to capture time-frequency representations of a particular window of the ECG signal with an image, the Continuous Wavelet Tranform CWT have been used. In

particular, the absolute values of the CWT coefficients of the considered ECG signal have been considered, obtaining the so called scalogram [16].

Two examples of scalogram are reported in Fig. 2 (Premature Atrial Contraction PAC) and in Fig. 3 (Normal Sinus Rhythm NSR), where the x axis represent the time, the y axis the frequency, and the color is the magnitude.
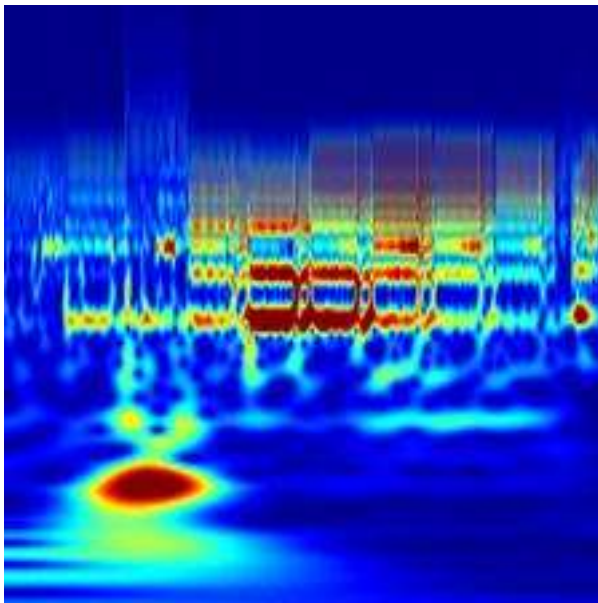

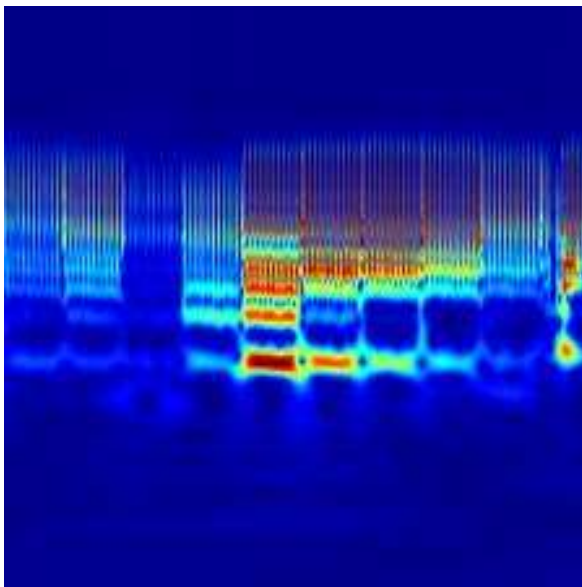Fig 2. Example of ECG scalogram (A00275 - PAC)

Fig. 3 Example of ECG scalogram (A00419 - NSR)

For improving the computation efficiency, the technique of precomputing filter bank of CWT was used for a more efficient computation technique, considering the computation for a large training database [16].

Particular techniques have been developed in order to adapt the general task to this particular problem:
- duplication of records in the learning set in order to have a more uniform and balanced distribution of the diagnostic classes
- the learning process of the Deep Learning method was adapted in order to cope with multiple diagnoses classification (comorbidity)
- the ECG signal is defined by two components of 8 independent leads:
   a. 10 seconds of ECG signal
   b. Average beats computed previously in the rule-bases block.

4. Results and Discussion

The score indices of the first and second phase of the Challenge (validation scores) are defined and reported in [14], and F-score (F_2) and G-score (G_2) and F-measure (F_meas) are considered. Our team named 'Gio_Ivo' participated successfully to the unofficial and official phases of the Challenge.

In the first phase the learning process was based only of the database DB1 consisting of 6877 ECG records with 9 possible diagnostic classes. Table 2. Reports the official indices. For the Deep Learning method, a matlab platform-related error was present and for this reason cross-validation indices on the database DB1 are reported.

Table 3 reports cross validation indices of the tested algorithms considering the learning database DB1.

Table 2. Results of the unofficial phase

	F_2	G_2
Rule-based	0.530	0.283
DL GoogleNet	0.575*	0.300*

*from Cross-validation techniques

Table 3. First phase cross-validation results on DB1

	F_2	G_2	F_meas
Rule-based	0.5110	0.269	0.461
DL GoogleNet 1	0.636	0.386	0.623
DL GoogleNet 2	0.632	0.381	0.614
DL Googlenet 3	0.634	0.390	0.618

In the second phase, the official Challenge, the number of the diagnostic classes indicated increased to the number of 110, but the scoring system was essentially concentrated to a subset of 27 classes, which reduced to 24 considering 3 equivalent classes. In order to increase the efficiency of the learning process, the subset S16K (16002 records, Table 1) was mainly used.

The Deep Learning process was performed and tested by cross-validation techniques. However, in the submitted

Table 4. Second Phase – Challenge Validation score

	score	cpu time	#it	N_max
DL GoogleNet 6	0.426	27:44:00	2(+10)	1000
DL GoogleNet 8	0.420*	61:54:00*	20	1000
DL GoogleNet 7	0.400	44:40:00	10	1000
DL GoogleNet 9	0.422	67:09:00	20	600
DL GoogleNet10	0.415	47:55:00	18	1500
Rule based 2	0.325	00:33:00	-	-
Rule-based 1	0.324	01:02:00	-	-

Table 5. Cross Validation results in subset S16K

	score	F_2	G:_2
DL GoogleNet 6*	0.49	0.47	0.19
DL GoogleNet 8	0.499	0.50	0.222
DL GoogleNet 7	0.497	0.47	0.199
DL GoogleNet10 *	0.480	0.466	0.194
Rule based 2	0.363	0.315	0.138
Rule-based 1	0.325	0.278	0.119

* With a shorter learning process

algorithm, the learning was performed with one fold for the relevant duration of the learning process and feedback phase (for the presence of several platform-related problems).

Table 4 reports the official Challenge Validation score of the submitted algorithms. The rule-based method RB1 essentially did not use any learning process from the database S16K and the score is in agreement with the behaviour of the first phase., while the second version (RB2) try to use extract some information from S16K. For example, it tried to differentiate AF from AFL by the AF-waves frequency and amplitude, but the consequent improvement was not significant.

Different Deep Learning algorithm were submitted, with different learning subsets (S16K, S20K, S11K) and different number of iterations, but the score (Table 4) were all in the range [0.400, 0.426], indicating that all these algorithms have a similar behaviour. In particular, GoogleNet 6 resume the training from a previously saved pre-trained network which comes from a 3-fold cross-validation technique on S16K and 10 iterations.

Table 5 reports Cross validation indices tested in the database S16K. It is interesting to note that the reported indices F_2, G_2 and the normalized score in agreement with the official results, with some more optimistic results, probably depending by the composition of the unknown test set.

The final official results of our team named 'Gio_Ivo' performed on the Deep Learning method GoogleNet 6, achieved a challenge validation score of 0.426 and a full test score of 0.298 placing us 12th out of 41 in the official ranking.

References

[1] I.I. Christov, Real time electrocardiogram QRS detection using combined adaptive threshold, *Biomed. Eng. Online*, 3;28:2004, www.biomedical-engineering-online.com/content/3/1/28

[2] I.K. Daskalov, I.I. Christov, Electrocardiogram signal preprocessing for automatic detection of QRS boundaries, *Med. Eng. & Phys*., 1999;21:37-44.

[3] I.K. Daskalov, I.I. Christov, Automatic detection of the electrocardiogram T-wave end, *Med. & Biol. Eng. & Comp*. 1999;37:348-53.

[4] I.I. Christov, I. Simova, Fully Automated Method for QT Interval Measurement in ECG. *Comp. in Card.,* 2006;33:321-4

[5] G. Bortolan, I,I. Christov I. Simova , I. Dotsinsky, Noise processing in exercise ECG stress test for the analysis and the clinical characterization of QRS and T wave alternans, *Biomedical Signal Processing and Control*. 18:378-85

[6] I. Jekova, G. Bortolan, I.I. Christov, Assessment and comparison of different methods for heartbeat classification. *Med. Eng. & Phy*., 2008;30:248-57

[7] I.I. Christov, I. Jekova , G. Bortolan, Premature ventricular contraction classification by the Kth nearest neighbors rule, *Phys. Meas.,* 2005;26:123-30.

[8] G. Bortolan, I.I. Christov, Myocardial infarction and ischemia characterization from T-loop morphology in VCG, *Comp. in Card*., 2001;28,633-6

[9] I.I. Christov, G. Bortolan, I. Daskalov. Sequential analysis for automatic detection of atrial fibrillation and flutter. *Comp in Card.,* 2001;28:293−6

[10] I.I. Christov, V. Krasteva, I. Simova, T. Neycheva, R. Schmid , Multi-parametric analyses for atrial fibrillation classification in ECG. *Comp. in Card*., 2017;44:1-4.

[11] G. Bortolan, C. Brohet, S. Fusaro, Possibilities of using neural networks for ECG classification. *J. of Electrocard*. 1996;29 (suppl):10-6.

[12] G Bortolan, W Pedrycz, An interactive framework for an analysis of ECG signals. *Artif Int. in Med*. 2002;24:109-32.

[13] I. Jekova, G. Bortolan, I. Christov, Assessment and comparison of different methods for heart beat classification. *Medical Engineering & Physics*, 2008;30:248-57.

[14] E.A.P. Alday, A. Gu, A. Shah, C. Robichaux, A.K.I. Wong, C. Liu, F. Liu, A.B. Rad, A. Elola, S. Seyedi, Q. Li, A. Sharma, G.D. Clifford, M.A. Reyna. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. *Physiol. Meas*. 2020 (under review)

[15] G.E. Dower, A lead synthesizer for the Frank system to simulate the standard 12-lead electrocardiogram. *J. of Electrocard.,* 1968;1:101-16.

[16] The Deep Learning Toolbox, User's guide, R2019b, The Matworks, 2019. (www.mathworks.com

[17] O, Russakovsky, J. Deng, H. Su et al. "ImageNet Large Scale Visual Recognition Challenge." International *Journal of Computer Vision*. 115, 2015;115:211–52

[18] ImageNet. http://www.image-net.org

Address for correspondence.
Giovanni Bortolan
IN-CNR
Corso Stati Uniti 4, - 35129 – Padova, Italy
e-mail: giovanni.bortolan@cnr.it