Multiple Cardiac Disease Detection from Minimal-Lead ECG Combining Feedforward Neural Networks with a One-vs-Rest Approach

Santiago Jiménez-Serrano[1], Miguel Rodrigo[2], Conrado J. Calvo[3], Francisco Castells[1], José Millet[1]

[1]Instituto ITACA, Universitat Politècnica de València, València, Spain
[2]Electronics Engineering Department, Universitat de València, València, Spain
[3]Universitat Politècnica de València, València, Spain

Abstract

Although standard 12-lead ECG is the primary technique in cardiac diagnostic, detecting different cardiac diseases using single or reduced number of leads is still challenging. The purpose of our team, itaca-UPV, is to provide a method able to classify ECG records using minimal lead information in the context of the 2021 PhysioNet/Computing in Cardiology Challenge, also using only a single-lead.

*Firstly, we resampled and filtered the ECG signals. R-waves detected from each available lead using a custom-made algorithm allowed to extract 109 features mostly based on Hearth Rhythm Variability (HRV) such as R-R interval stats or Lorenz plot dispersion descriptors. We used selected features to train one feed-forward neural network (FFNN) with one hidden layer for each class using a One-vs-Rest approach, thus allowing each ECG been classified as belonging to none or more than one class. Furthermore, since each FFNN provided a continuous output in range [-1,1], we used the G-metric (G=sqrt(sensibility * specificity)) to identify the output threshold that best fitted its binary classification. Finally, we performed a hold-out validation to assess the whole model performance.*

Our classifiers received scores of 0.402, 0.402, 0.363, 0.380, and 0.402 (ranked 161th, 148th, 182th, 171th, and 151th out of 256 entries) for the 12-lead, 6-lead, 4-lead, 3-lead, and 2-lead versions of the hidden validation set with the Challenge evaluation metric. Furthermore, we obtained a G metric values of 0.75, 0.73, 0.68, 0.70, 0.74 and 0.71 on the public training set for the 12, 6, 4, 3, 2 and 1-lead versions of our classifier.

Our minimal-lead approach may be highly beneficial for novel portable or wearable ECG devices used as screening tools, as it can also detect multiple and concurrent cardiac conditions. Accuracy in detection can be improved adding more disease-specific features.

1. Introduction

The clinical importance of cardiac arrhythmias is increasing along with their incidence and prevalence mostly associated with population aging [1]. Besides this, nowadays wearable devices are gaining great interest as monitoring devices in both research and clinical ambits [2]. Although standard 12-lead ECG is the primary technique in cardiac diagnostic, detecting different cardiac diseases using single or reduced number of leads is still challenging [3].

The aim of this study is to is to provide and evaluate methods able to classify ECG records using minimal lead information in the context of the *2021 PhysioNet/Computing in Cardiology Challenge* [4, 5], using also only a single-lead.

2. Materials

As database for this study we used the 88,253 12-lead ECG registers provided by the competition as training set containing also the age and gender of the patient for each record. Deeper explanation of the database can be found in [4, 5].

3. Methods

This section describes the signal feature extraction and selection processes, plus the models validation methodology used during this work. All these stages were carried on through MATLAB (R2020b, The MathWorks). In addition to the official leads sets of the challenge, we report this methodology results using the single lead ‘I’.

3.1. Signal preprocessing

First, provided signals were resampled to 500Hz if its sampling frequency were different than that. Next, a 50Hz notch filter plus a band-pass filter between 0.5Hz and 40Hz were applied. Finally, we removed the first and last second of signal in order to leave out the filtering stabilization stage. Lastly we performed an artifacts filtering; to do so, we used a 0.5 second sliding window in order to calculate aberrant maximum and minimum values, and sections surrounded by outliers were set to zero.

3.2. Feature Extraction

We automatically extracted 109 signal features mostly derived from ventricular activity from each ECG lead, most of them previously used in [6]. To carry out this task, initially, we extracted the *RR* sequence using a *QRS* detector based on the first derivative of the ECG. Then we filtered the outliers from the *RR* sequence, and obtained the first and second derivatives of that sequence (*RRd1*, *RRd2*). Also, we created a T-wave detector in order to obtain the QT interval and other related features.

Finally, we got both the QRS and T wave patterns for each lead using a ±100ms window over all the QRS and T wave detections.

Furthermore, we got the Welch's power spectral density estimation for each lead in order to obtain some frequency-based features.

Using the above information, the extracted signal features for each lead can be grouped as follows:

Group 1. Basic statistics over the R and T waves voltages (mean, standard deviation). 4 features.

Group 2. Basic statistics over the QT interval in milliseconds (mean, standard deviation). 2 features.

Group 3. Features based on the QRS and T patterns: Percentage of amplitude of T wave respect the R wave, sign of the R and T waves (positive or negatives), percentage of waves discards and RMSE during the R and T pattern definition, and maximum values for first and second derivatives of both patterns. 11 features.

Group 4. Spectral features: Dominant frequency (*fdom*) using the Welch spectral density estimation method, percentage of the area in *fdom*±0.5Hz in the periodogram normalized in the range [0, 1] and the sum of the normalized periodogram in steps of 2Hz in the range [0, 30] Hz. 17 features.

Group 5. Basic statistics over the *RR*, *RRd1* and *RRd2* sequences (mean, standard deviation, kurtosis, skewness). 9 features.

Group 6. Features based on *RRd1*: *RMSSD*, *pNN25*, *pNN50*, *pNN75*, where *pNNxx* [7] denotes the percentage of intervals between normal beats exceeding *xx* ms. 4 features.

Group 7. *Poincaré* plot-based features using *RRd1*: Maximum, minimum, mean, standard deviation, kurtosis and skewness of the distances among all the points plus the absolute difference between the maximum and minimum distance values. 7 features.

Group 8. Lorenz plot-based features using *RRd2*: Angular variability, dispersion of the distance between points to origin, and differences between 2 and 3 consecutive beats. 8 features.

Group 9. Same statistics as in points 5 and 6, but using an 8 seconds sliding window and a step size of 2 seconds. Once the matrix of values is obtained using each signal interval, we extracted the minimum, maximum, mean and standard deviation for each feature, appending all this values in a 44 features vector.

Group 10. Other features: Shannon entropy of the *RR* sequence, Lempel-Ziv complexity of the *RR* time series after binarization using the median as threshold, and ratio between the number of different *QRS* patterns found and the total number of waves detected. 3 features.

3.3. Feature dataset preprocessing

First, for each feature, outliers exceeding 3 times the standard deviation above or below the median were replaced by these same limits.

Next, if some sample contained a *NaN* value due to a feature extraction error or the impossibility of obtaining such value for a given sample, we replaced that value for the median value in the dataset for such feature. According to this rule, and taking into account 1310 features in the whole dataset using 12 leads (age, sex, and 109 features for each lead), finally the 0.61% of values were replaced for the corresponding median.

Lastly, we performed a z-score using the training set to rescale the whole dataset.

3.4. Scoring

2021 PhysioNet/Computing in Cardiology Challenge scoring rules are described in [5], where only 26 classes are taken into account. Also, we report the G metric ($G = sqrt(Sensitivity * Specificity)$) in this work since it was used in order to select the binary classifiers with best performance during the training and validation stage.

3.5. Feature Selection

Previously to the training of each feed-forward neural network (FFNN) mentioned below, a feature selection was performed for each class using both supervised and unsupervised statistical filtering methods.

Age and sex always were used in order to avoid an empty set of features. Next, we perform a two-sample *ttest* with an alpha value of 0.05 for each feature taking into account if the sample belongs or not to the specified class, and all the features that did not pass the significance test were removed. Finally, we get the correlation coefficient among the lasting features for each pair of features, and we removed the last feature of the pair where their correlation coefficient was greater or equal than 0.9. The remaining features were used as inputs for the corresponding binary classifier.

3.6. One-vs-Rest Classification Approach

In this work we used a One-vs-Rest classification approach, where for each class in the training set, a classifier was trained and used in order to give a binary response indicating if an unseen sample belongs or not to the corresponding class. Thus, each classifier solves an independent problem in the whole classification model, been possible to assign to none or more than one class a new sample.

Each binary classifier uses its selected set of features as inputs that best fits its own classification problem. Next, each binary classifier corresponds to a FFNN made of 18 or 32 hidden units, and a threshold for the output to give the binary response. All the FFNN where trained with the default objects and parameters in the Matlab R2020b Deep Learning Toolbox, using the *trainscg* (Scaled Conjugate Gradient) as training function, the *useGPU* flag switched on in order to use the available GPUs to speed up the training and the *showResources* flag switched off.

Using as inputs the selected features for a given class, the 75% of training data was used to train the FFNN and the resting 25% to select the output threshold in the range [-1, 1] that achieves a higher G value, both with 18 and 32 hidden units. Finally, among the two trained models, we choose the one that presented a higher G value to be used in the whole One-vs-Rest classification model.

3.7. Model validation

Since the number of samples in the database is large enough and the training time could become unnecessarily high for a cross-validation, we used a hold-out validation approach (66.67%-33.33% train/validation split) with 88,253 available training samples with no bias among the distinct databases used. Thus, the results reported in this work (Challenge score and G metric) in the training section corresponds to 29,417 samples never used to train neither select the parameters of the models.

4. Results

Best results in the hidden validation set using the Challenge score have a value of 0.402 using 12, 6 and 2 leads indistinctly. Table 1 shows the whole results set and ranking using the Challenge score.

Table 2 shows the mean of different performance metrics in the classification of the 26 scored classes in the challenge on the public training set, where higger G value of 0.753 was achieved using 12 leads, followed by a G value of 0.74 using only 2 leads.

Finally, Table 3 shows the results achieved for individual binary classifiers where G metric is greater than 0.8 in some of the lead combinations. Thus, ten different cardiac conditions reach this classification performance.

#Leads	Training	Validation	*Ranking*
12	0.364	0.402	161
6	0.354	0.402	148
4	0.271	0.363	182
3	0.313	0.380	171
2	0.371	0.402	151
1	0.330	-	-

Table 1. Challenge scores for our final selected entry (team itaca-UPV) using hold-out validation on the public training set, scoring on the hidden validation set as well as the ranking on the hidden validation set.

#Leads	AUROC	Sens.	Spec.	*G*
12	0.825	0.811	0.735	0.753
6	0.802	0.787	0.710	0.729
4	0.775	0.753	0.675	0.683
3	0.790	0.794	0.667	0.702
2	0.814	0.793	0.727	0.740
1	0.782	0.768	0.689	0.709

Table 2. Mean of different performance metrics in the classification of the 26 scored classes in the challenge for our final selected entry using hold-out validation on the public training set: Area Under the ROC Curve, Sensitivity, Specificity and G metric.

5. Discussion

Results obtained in this work showed few differences among the results obtained in the G values among the classifiers that uses 12 leads and the ones that only uses minimal leads information. Since the signals from leads I and II share many characteristics with those offered by wearables devices, the resulting classification models could be good candidates to be implemented in wearable patient management systems being as this approach is computationally low consuming during classification. Nevertheless, this models could be used safety only in cases where the G value is higher than 0.8

However, poor results in the Challenge score metric shows that this approach must be used carefully in order to detect cardiac conditions with low performance in our results. It must be improved mostly by adding more disease-specific features and/or modifying the binary classification strategy.

Class	G (12-leads)	G (6-leads)	G (4-leads)	G (3-leads)	G (2-leads)	G (1-leads)
Sinus tachycardia	0.931	0.926	0.910	0.891	0.932	0.931
Sinus bradycardia	0.910	0.903	0.865	0.912	0.920	0.923
Left anterior fascicular block	0.893	0.896	0.897	0.853	0.908	0.674
Atrial flutter	0.862	0.860	0.843	0.842	0.857	0.848
Atrial fibrillation	0.865	0.873	0.862	0.848	0.844	0.859
Sinus arrhythmia	0.847	0.828	0.809	0.819	0.840	0.843
Pacing rhythm	0.865	0.857	0.793	0.830	0.836	0.812
Left axis deviation	0.848	0.845	0.812	0.645	0.820	0.641
Sinus rhythm	0.790	0.791	0.620	0.783	0.798	0.793
Complete left bundle branch block	0.814	0.773	0.750	0.634	0.785	0.813
Right bundle branch block	0.871	0.750	0.744	0.733	0.776	0.719

Table 3. G metric values for single FFNN models that exceeds 0.8 in some lead combination, plus Sinus Rhythm results during the hold-out validation on the public training set.

6. Conclusion

We presented and evaluated a robust methodology for multiple cardiac disease detection through ECG registers that combines feature extraction and selection, and a One-vs-Rest classification approach using FFNN as binary classifiers. The results include low computationally consuming classification models for one or two leads suitable for wearable monitoring devices. Improving the identification of some cardiac rhythms incorporating more specific features for those cases where the performance was low, is an interesting direction to explore in the future.

References

[1] Chow GV, Marine JE, Fleg JL (2012). Epidemiology of Arrhythmias and Conduction Disorders in Older Adults. Clinics in Geriatric Medicine, 28(4), 539–553.

[2] Kumari P, Mathew L, Syal P (2017). Increasing trend of wearables and multimodal interface for human activity monitoring: A review. Biosensors and Bioelectronics, 90: 298-307.

[3] Kligfield P. The Centennial of the Einthoven Electrocardiogram. Journal of Electrocardiology 2002;35(4):123–129.

[4] Perez Alday EA, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, et al. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiological Measurement 2020;41.

[5] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, et al. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. Computing in Cardiology 2021; 48:1–4.

[6] S Jiménez-Serrano, J Yagüe Mayans, E Simarro-Mondéjar, CJ Calvo, F Castells, J Millet. Atrial Fibrillation Detection Using Feedforward Neural Networks and Automatically Extracted Signal Features. Computing in Cardiology 2017; 44:131–134.

[7] Mietus JE, Peng C-K, Henry I, Goldsmith RL, Goldberger AL. (2002). The pNNx files: re-examining a widely used heart rate variability measure. Heart, 88: 378-380.

Address for correspondence.

Santiago Jiménez Serrano
Universitat Politècnica de València, Instituto ITACA, Edif. 8G
Camino de Vera s/n, 46022-València (Spain)
sanjiser@upv.es