

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

Santiago Jiménez-Serrano<sup>1</sup>, Miguel Rodrigo<sup>2</sup>, Conrado J. Calvo<sup>3</sup>, Francisco Castells<sup>1</sup>, José Millet<sup>1</sup>

<sup>1</sup>Instituto ITACA, Universitat Politècnica de València, València, Spain

<sup>2</sup>Electronics Engineering Department, Universitat de València, València, Spain

<sup>3</sup>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.*

*We resampled and filtered the ECG signals, and extracted 109 features mostly based on Heart Rate Variability (HRV). Then, 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 to be classified as belonging to none or more than one class. Finally, we performed a 3-fold cross validation to assess the model performance.*

*Our classifiers received scores of 0.34, 0.34, 0.27, 0.30, and 0.34 (ranked 26th, 21th, 29th, 25th, and 22th out of 39 teams) for the 12, 6, 4, 3 and 2-lead versions of the hidden test set with the Challenge evaluation metric.*

*Our minimal-lead approach may be 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 settings [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 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 performed using 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, all ECG signals were resampled to 500Hz if necessary. 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 each signal in order to leave out the filtering stabilization stage. Lastly we removed artifacts using a 0.5 second sliding window in order to calculate aberrant maximum and minimum values, where 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  $\pm 100$ ms 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  $fdom \pm 0.5$ Hz 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 ( $G = \sqrt{\text{Sensitivity} * \text{Specificity}}$ ) 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 (see Figure 1), 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 were 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.

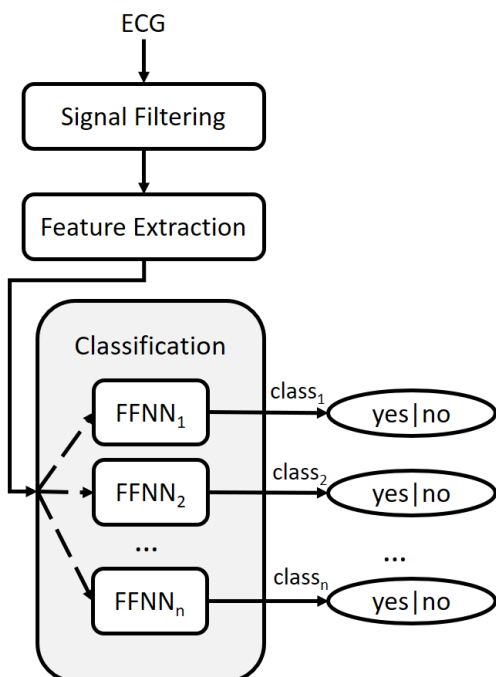


Figure 1. Summary diagram of the One-vs-Rest classification approach used during this work.

### 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 with a large number of folds, we used a 3-fold cross validation with the 88,253 training samples. Furthermore, we selected the samples for each fold with no bias among all the distinct databases available.

## 4. Results

Best results in the hidden test set using the Challenge score have a value of 0.34 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 higher G value of 0.76 was achieved using 12 leads, followed by a G value of 0.74 using both 6 and 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 during the validation of the first training fold, where 10 different cardiac conditions reach this classification performance threshold.

#Leads	Training	Validation	Test	Ranking
12	0.38±0.01	0.402	0.34	26/39
6	0.36±0.01	0.402	0.34	21/39
4	0.28±0.01	0.363	0.27	29/39
3	0.31±0.01	0.380	0.30	25/39
2	0.36±0.01	0.402	0.34	22/39
1	0.34±0.01	-	-	-

Table 1. Challenge scores for our final selected entry (team itaca-UPV) using 3-fold cross validation on the public training set, repeated scoring on the hidden validation set, and one time scoring on the hidden test set as well as the ranking on the hidden test set.

#Leads	AUROC	Sens.	Spec.	G
12	0.83	0.81	0.74	0.76
6	0.81	0.79	0.72	0.74
4	0.78	0.76	0.68	0.69
3	0.79	0.78	0.67	0.69
2	0.81	0.79	0.73	0.74
1	0.79	0.78	0.70	0.71

Table 2. Mean of other performance metrics among the classification of the 26 scored classes for our final selected entry using 3-fold cross validation on the public training set: Area Under the ROC Curve, Sensitivity, Specificity and G metric; mean of the standard deviation was ±0.007.

## 5. Discussion

Results obtained in this work showed low 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. Classification using only one or two leads outperformed the ones using three and four leads, possibly due to leads V2 and III added some bias in

their extracted features respect the ones in leads I and II. In this sense, further studies should get the classification performance for each single lead in order to know their individual performance.

On the other hand, since the signals from leads I and II share many characteristics with those offered by wearables devices, the resulting classification models may eventually be suitable for clinical use in wearable or automated control systems. This approach would benefit from low

computational costs consumption during classification.

Nevertheless, poor results in the Challenge score metric could highlight that this approach should be used cautiously when detecting cardiac conditions with low performance in our results. Future optimization of these classifiers should improve 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 validation of the first public training fold.

## 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. Interestingly, the classification results using only one or two leads outperformed the ones using three and four leads, and almost matched the ones with twelve leads, showing lower computational costs and been more suitable for wearable monitoring devices. Improving the identification of some cardiac rhythms by incorporating more specific features for those cases where the performance was low, should be 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](mailto:sanjiser@upv.es)