

Atrial Fibrillation Detection Using Feedforward Neural Networks and Automatically Extracted Signal Features

Santiago Jiménez-Serrano¹, Jaime Yagüe-Mayans¹, Elena Simarro-Mondéjar¹, Conrado J. Calvo¹, Francisco Castells¹, José Millet¹

¹Instituto ITACA, Universitat Politècnica de València, València, Spain

Abstract

Atrial Fibrillation (AF) is becoming an increasingly significant clinical matter as its prevalence keeps growing. Therefore, developing algorithms to accurately detect AF episodes from single-lead ECG recordings would benefit early and automatic diagnosis from monitoring devices. The aim of this study is to provide a Feedforward Neural Network (FFNN) classification model and assess its performance in order to discriminate short single-lead ECG registers among 4 groups: Normal (N), AF (A), Other rhythms (O) and noisy (~).

We extracted automatically 72 features derived from ventricular activity from each of the 8528 ECG records provided by the 2017 PhysioNet/Computing in Cardiology Challenge. Next, we performed a Feature Selection (FS) and a training/validation process through a grid search over a set of FFNN training parameters. We used a F1 scoring in order to assess the classification performance.

Results shown that filtering 50 features during the FS stage improved the initial classification performance from $F1=0.70$ to $F1=0.73$. The following tuning of FFNN training parameters showed the best results during the 10-Fold Cross-Validation with $F1=0.76$ ($F1n=0.87$, $F1a=0.78$, $F1o=0.65$, $F1p=0.45$) using 200 epochs, $\alpha=0.7$, $\beta=0.0$, and one hidden layer made of 128 units. The final score on the test data was $F1=0.77$, demonstrating the robustness of the presented method.

Our strategy revealed promising classification scores using a robust validation approach. The resulting classification model is computationally low consuming during classification, hence is a good candidate to be implemented in wearable patient management systems.

1. Introduction

The clinical importance of cardiac arrhythmias is increasing along with their incidence and prevalence mostly associated with population aging [1]. Among these diseases, atrial fibrillation (AF) stands out [2] since its growing trend is more significant and since in many

occasions it can be difficult to diagnose, due to paroxistic behaviour and the absence of symptoms in some cases.

On the other hand, nowadays we are in a scenario where wearable devices are gaining great interest as monitoring devices in both research and clinical ambits [3]. However, automatic methods to provide a reliable diagnostic of AF using the ECG signals provided by wearable devices is still challenging, particularly if other normal or abnormal rhythms are also considered.

The aim of this study is to provide a classification model and assess its performance in order to discriminate short single-lead ECG registers among 4 groups: Normal (N), AF (A), Other rhythms (O) and noisy (~); in the context of the 2017 PhysioNet/Computing in Cardiology Challenge [4].

The strategy proposed in this work is based on automatic signal Feature Extraction (FE) derived from ventricular activity, plus a posterior Feature Selection (FS). Finally, those selected features were used in order to train and validate the classification performance of a set of different Feedforward Neural Networks (FFNN) through a tuning of training parameters.

Since the signals used share many characteristics with those offered by wearables devices, the resulting classification model could be a good candidate to be implemented in wearable patient management systems being as this approach is computationally low consuming during classification.

2. Materials

As database for this study we used the 8528 single-lead ECG registers provided by the competition. Deeper explanation of the database can be found in [4].

3. Methods

This section describes the signal feature extraction and selection processes, plus the model validation methodology used during this work. All these stages were carried on through MATLAB (2017a, The MathWorks).

3.1. Signal preprocessing

Provided signals were already band-pass filtered between 0.5Hz and 40Hz, so we performed an artefacts filtering as unique signal preprocessing stage. 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 not considered for analysis purposes.

3.2. Scoring

2017 PhysioNet/Computing in Cardiology Challenge scoring rules are described in [4], where F1a, F1n, F1o and F1p are the individual score for A, N, O and ~ classes respectively. F1 represents the mean among F1a, F1n and F1o values, and consequently the classification performance for a given model in this work. We report F1p although is not necessary in order to calculate F1.

3.3. Feature Extraction

We automatically extracted 72 signal features derived from ventricular activity from each ECG record. Atrial activity related features were avoided since atrial signal can be easily obscured by artifacts and noise, or even not be present in the signal. 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*).

Using the above information, the extracted signal features can be grouped as:

1. Basic statistics (mean, standard deviation, kurtosis, skewness) over the *RR*, *RRd1* and *RRd2* sequences. 9 features.
2. Features based on *RRd1*: *RMSSD*, *pNN25*, *pNN50*, *pNN75*, where *pNNxx* denotes the percentage of intervals between normal beats exceeding *xx* ms. 4 features.
3. 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.
4. Same statistics as in points 1 and 2, 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.
5. 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.
6. Dynamic Time Warping (*DTW*) distance value among

the mean *pNNx* vector [5] for each class (*pNNx(A)*, *pNNx(N)*, *pNNx(O)*, *pNNx(~)*) and the corresponding *pNNx(i)* vector for each *i* sample. 4 features.

The last features group was the unique that needed a previous processing of the dataset in order to obtain the mean *pNNx* vectors for each class. Figures 1 and 2 show the result from varying the threshold, denoted by *x*, to obtain the ratio of consecutive *RR* that differs more than *x* ms. Figure 2 also shows standard deviation of measures for N and A groups.

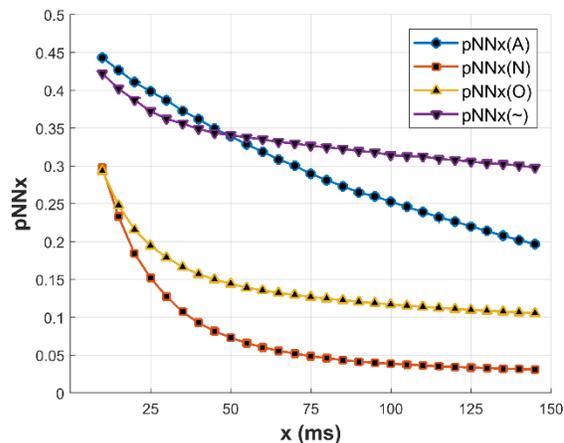


Figure 1: Mean *pNNx* vectors [5] obtained at the provided dataset for each class: *pNNx(A)*, *pNNx(N)*, *pNNx(O)* and *pNNx(~)*. These vectors are necessary in order to get the *DTW* distance of any other *i* sample with its own *pNNx(i)*.

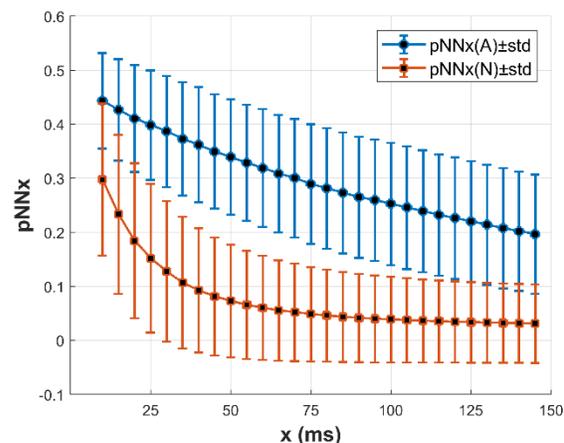


Figure 2: Mean *pNNx* vectors [5] and their standard deviation for class A and N (*pNNx(A)*, *pNNx(N)*).

Using the vectors presented above, we propose 4 new features (group 6) using the *DTW* metric in order to calculate a distance among the *pNNx* vector generated for each *i* sample (*pNNx(i)*), and each mean *pNNx* vectors corresponding to each class: *pNNx(A)*, *pNNx(N)*, *pNNx(O)* and *pNNx(~)*. Hence, these new distance features are

denoted as

$$\begin{aligned} DtwNNA(i) &= DTW(pNNx(A), pNNx(i)) \\ DtwNNn(i) &= DTW(pNNx(N), pNNx(i)) \\ DtwNNO(i) &= DTW(pNNx(O), pNNx(i)) \\ DtwNNp(i) &= DTW(pNNx(\sim), pNNx(i)) \end{aligned}$$

Figure 3 shows the boxplot using $DtwNNn$ values in our dataset. Using all the above detailed features, we created a dataset consisting of 8528 samples and 72 features for each sample.

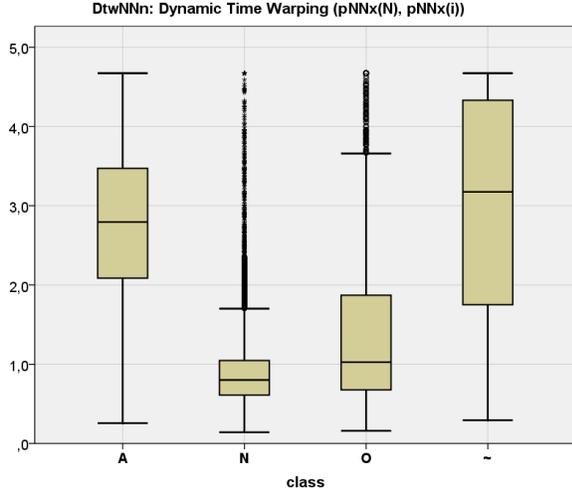


Figure 3: Boxplot of the $DtwNNn$ proposed feature values obtained using the DTW distance among each $pNNx(i)$ value of the database and the mean $pNNx(N)$ vector corresponding to the Normal rhythm class.

3.4. 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, we removed that sample from the dataset. According to this rule, 25 samples were removed from the dataset, obtaining a final dataset with 8503 samples. Since 64% of removed samples corresponded to Noisy class, during the test stage we classify as (\sim) those containing some NaN value. Finally, we performed a z-score using the training set to rescale the whole dataset.

3.5. Feature Selection

A FFNN with 1 hidden layer of 128 units was used in order to determine the $F1$ score through a hold-out strategy (50%-50% train/test split), using 75 epochs to avoid overfitting. To determine the best learning rate (α) and momentum (β) FFNN training parameters for this stage, an initial grid-search was performed. Next, we carried out a

Feature Selection (FS) using a backward-elimination approach. Initially, all the variables are included in the dataset. A relaxation condition was added to the FS process in order to avoid early stopping criteria: if during the current iteration (j) of the algorithm, $F1$ score does not increase respect the ($j-1$) iteration, we select as feature to be removed the one that maximized $F1$ during the iteration (j) if and only if $F1(j) > (F1(j-1) - 0.05)$.

3.6. FFNN training and validation

Using as inputs the selected features, we performed a grid search over the training parameters of a FFNN (α , β , architecture) in two stages: 1) A wide grid-search using a hold-out validation approach (50%-50% train/test split) using 100 epochs; 2) Assessment of the 10 best training parameters combinations through a 10-Fold Cross-Validation (10F-CV) using 100 and 200 epochs respectively. Finally, parameters allowing to achieve higher $F1$ score were used in order to train the final FFNN with the whole dataset, to be submitted for the Challenge competition.

4. Results

Results showed an initial $F1$ score of 0.709 using the 72 original features, achieved with $\alpha=0.6$, $\beta=0.1$ (FFNN 1 hidden layer of 128 units). During the FS stage, we removed 50 features, achieving a new $F1$ score of 0.734. We found that the $DtwNNn$ and $DtwNNp$ proposed features were within the selected set of features.

Table 1. Best $F1$ scores achieved during hold-out validation stage using 100 epochs. Training parameters: number of units in the hidden layer (Architecture), learning rate (α), momentum (β).

Architecture	(α)	(β)	$F1$
[1024]	0.7	0.1	0.7360
[512]	0.6	0.1	0.7356
[512]	0.5	0.1	0.7332
[128]	1.0	0.1	0.7320
[512]	0.5	0.4	0.7318
[128]	0.7	0.0	0.7310
[256]	0.5	0.1	0.7301
[256]	0.5	0.4	0.7301
[512]	0.8	0.1	0.7299
[512]	0.7	0.4	0.7298

Table 1 shows the results during the hold-out validation stage and Table 2 shows the results during the 10F-CV stage. All the FFNN reported consisted in 1 hidden layer. Best $F1$ score value was 0.7644, achieved with a FFNN made of 118 units in the hidden layer, using 200 epochs, $\alpha=0.7$ and $\beta=0.0$. Finally, the FFNN trained with the whole training dataset and the best training parameters achieve a

final score value of $F1=0.77$ on the test data.

Table 2. Best $F1$ ($F1a$, $F1n$, $F1o$, $F1p$) scores achieved during the 10-Fold Cross-Validation stage.

#Epochs	Architecture	(α)	(β)	$F1$	$F1a$	$F1n$	$F1o$	$F1p$
200	[128]	0.7	0.0	0.7644	0.7768	0.8668	0.6496	0.4490
200	[128]	0.6	0.1	0.7629	0.7761	0.8668	0.6458	0.4624
100	[512]	0.5	0.4	0.7627	0.7723	0.8645	0.6512	0.5112
100	[128]	0.7	0.0	0.7604	0.7766	0.8646	0.6402	0.4587
100	[512]	0.6	0.1	0.7603	0.7694	0.8637	0.6477	0.4844
200	[512]	0.5	0.1	0.7601	0.7656	0.8652	0.6494	0.4810
100	[1024]	0.7	0.1	0.7600	0.7649	0.8638	0.6513	0.3974
100	[512]	0.5	0.1	0.7599	0.7754	0.8632	0.6410	0.5097
100	[512]	0.8	0.1	0.7585	0.7681	0.8646	0.6429	0.4962
200	[1024]	0.7	0.1	0.7580	0.7629	0.8653	0.6457	0.4795
100	[512]	0.7	0.4	0.7578	0.7681	0.8593	0.6460	0.4959
200	[512]	0.5	0.4	0.7575	0.7635	0.8618	0.6471	0.4680

5. Discussion

Results obtained in this work showed differences in simple ventricular activity derived features for different rhythms, even when the arrhythmia had a supraventricular origin (AF). This allows for avoiding problems associated with absent or masked atrial activity. Also, FFNN showed good performance taking profit of those differences in order to classify the records in the 4 considered classes. Furthermore, the results of this method shows a high robustness since best $F1$ score in training and test data were 0.76 and 0.77 respectively, showing no overfitting of the model.

However, there were difficulties when discriminating O and ~ groups, since these groups aggregate records with different rhythms and characteristics, which results in great variability in considered features. Given this, we believe that including noise related features and further sub-grouping are strategies that could enhance the obtained results.

6. Conclusion

We presented and evaluated a robust method for short single-lead ECG registers classification that combines ventricular feature extraction and selection, and a FFNN training and validation approach. The result is a low computationally consuming classification model suitable for wearable monitoring devices. Improving the identification of other rhythms other than normal or AF incorporating more specific features for these cases into the proposed method, is an interesting direction to explore in the future.

Acknowledgements

This work has been funded in part by the *Spanish Ministry of Economy and Competitiveness*: Programa Promoción de Empleo Joven e Implantación de la Garantía Juvenil 2014 (PEJ-2014-A-36316).

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] Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra, M, Benjamin EJ, Murray CJL (2014). Worldwide Epidemiology of Atrial Fibrillation: A Global Burden of Disease 2010 Study. *Circulation*, 129(8), 837–847.
- [3] 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.
- [4] Gari Clifford, Chengyu Liu, Benjamin Moody, Ikaro Silva, Qiao Li, Alistair Johnson, Roger Mark, "AF Classification from a Short Single Lead ECG Recording: the PhysioNet Computing in Cardiology Challenge 2017", *Computing in Cardiology (Rennes: IEEE)*, Vol 44, 2017.
- [5] 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