A Comparison of Different Classifiers Architectures for Electrocardiogram Artefacts Recognition

Carlos R. Vázquez-Seisdedos[1], Alexander A. Suárez-León[1], and Joao Evangelista-Neto[2]

[1] Center for Neurosciences Studies, Image and Signal Processing, Biomedical Engineering Departament, Universidad de Oriente, Santiago de Cuba
{cvazquez,aasl}@fie.uo.edu.cu
[2] Amazon State University and UniNorte/Laurente Manaus, Brazil
joao_evangelista_neto@yahoo.com

Abstract. Applying heart rate variability (HRV) analysis on ambulatory ECG monitoring is a very useful decision support tool for cardiovascular diagnosis. The presence of non-valid beats (artefacts) on the RR interval time-series affects the diagnosis accuracy using this technique. Despite the importance of artefacts recognition prior to exclusion, no paper was found characterizing quantitatively the performance of, on the one hand, the extracted features and, on the other hand, the clustering methods on artefacts recognition for HRV analysis. In this paper we evaluate the performance of several combinations of three feature extraction methods and four clustering methods (based on machine learning techniques) for the artefacts beats recognition on the ECG signal. The trade-off between performance indexes suggests the use of a non-linear principal component analysis as feature extraction method and a multilayer perceptron (MLP) as clustering method, with sensitivity, specificity and positive-predictive-value (PPV) equal to respectively 95 %, 95.9 % and 98 %.

Keywords: ECG, artefact detection, artificial neural networks, feature extraction, classifier.

1 Introduction

The ambulatory monitoring of electrocardiogram (ECG) during daily activities plays an important role in the diagnosis but presents the challenge of information loss due to the occurrence of technical and physiological artefacts that distort the ECG signal. Typically, more than 80 000 heartbeats per channel are recorded during 24 hours; so many computer-based methods for automatic ECG analysis have been studied for a long time. ECG recognition is a difficult problem even with the aid of a computer, because ECG waveforms may differ significantly even for the same beat type taken from the same patient. The architecture for morphological recognition of beats in ECG includes several stages as showed in figure 1. The core is the classifier composed by features extraction (FE) and clustering stages, both based on computational intelligence techniques.

J. Ruiz-Shulcloper and G. Sanniti di Baja (Eds.): CIARP 2013, Part II, LNCS 8259, pp. 254–261, 2013.

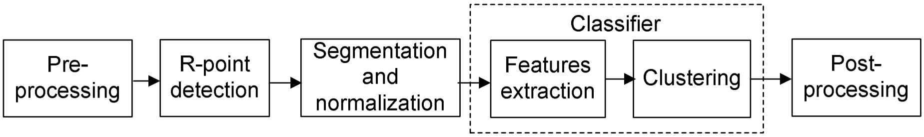

Fig. 1. Stages of an ECG-signal classifier system

Some FE methods on ECG are based on:

1. Morphologic features extracted from signal [1], [2]: amplitudes, interval durations or areas of waves or specific segments.
2. Statistical parameters in time domain (mean, standard deviation, maximum, minimum, self-correlation-coefficients, histogram, etc) as well as in frequency domain (QRS-complex-energy, power spectral density).
3. The use of mathematical models to represent ECG wave and segments [3], like autoregressive models, linear prediction coefficients and curve fitting.
4. The use of transforms: (a) Principal Component Analysis (PCA) [4], (b) Discrete Cosine Transform (DCT) [5], (c) Wavelet Transform [3], (d) Time-frequency distributions and (d) Hermite functions, among others.

Several artificial neural networks (ANN) based clustering methods that automatically classify heart beats have been proposed in the last years. Multilayer Perceptron (MLP) is one of the most referred [1]. Other clustering methods (in descending order) are support vector machine (SVM), learning vector quantization (LVQ) and radial basis functions (RBF).

In order to validate the HRV analysis [6], it should be verified that each detected R-point corresponds to a complete beat resulting from sinus node depolarization without any type of atrioventricular blockade. Otherwise, the beat will be considered as an artefact located in the corresponding positions of the RR time series, and it should be excluded of the analysis. The heart beat artefacts can have either a physiological (e.g. arrhythmias) or a technical (e.g. spikes and noise) origin.

Although there is an extensive diversity of publications about arrhythmia recognition [7-9], no publication was found characterizing quantitatively the performance of the FE and clustering methods for recognition of heart beat artefacts. There is a recent work [10] that compares several FE methods according to simplicity, accuracy and positive predictive value, but only on the qualitative point of view and not including artefacts beats. This paper does not analyze the execution time or other indexes, neither others FE methods as popular as DCT and linear PCA.

In a previous work [11], three FE methods were characterized, using an MLP network as a gold standard for clustering. The higher performance corresponded to non-linear PCA, also named kernel PCA (KPCA). The previous research left the following question: will another cluster method exist with a better performance?

The aim of this work is to validate the performance of three FE methods combined with four clustering approaches to detect non-valid beats (artefacts beats) for HRV analysis.

2 Methodology

2.1 Data

The MIT-BIH arrhythmia database is used for training and validation. This database consists of 48 30-min two-lead recordings (series 100 and 200) sampled at 360 Hz, for a total of 24 hours [12]. The development platform was MATLAB 7.7.

The beat classes' global distribution from this database has 110288 beats: 75056 normal beats and 35232 artefact beats. Thus, around the 70% of the beats were classified as normal beats (resulting from sinus node). There are 17 classes of beats (ectopic, left and right bundle blocks, and others) that are grouped in a class: artefact (ARTF). Every normal beat belongs to the normal (NORM) class.

Initially, a partial evaluation using 4000 beats (2000 for training and 2000 for validation) was made in order to find the best clustering method for this sample. Then, a global evaluation was made for the entire database using the clustering method found. Of the 4000 beats belonging to different database records, were chosen 2000 for each class according to the following criteria: from each record with more than 50 beats of NORM class (40 records), 50 beats were randomly chosen. For the beats of the ARTF class the criterion is the following: from each record with more than 375 beats (25 records) from ARTF class, 80 beats were randomly chosen.

2.2 Stages of the Classification System Beats

Preprocessing: To eliminate baseline drift and high frequency noise, a bandpass filter is used, consisting of a high-pass filter (Butterworth, zero-phase, 6th order, cutoff frequency equal to 0.6 Hz) in cascade with a low-pass filter (Butterworth, zero-phase, 12th order, cutoff frequency equal to 45 Hz). Subsequently, the average value was eliminated so that the signal is converted into a signal of unit variance. This standardization is performed to achieve invariance with respect to the amplitude, for any beat.

R Peak Detection: Because the R peak detection has been broadly described, no further discussion on this subject is pursued in this paper. In [13] there is an extensive review of recent approaches for R peak detection. Any R peak detector with demonstrated robustness can be used. In this work, R peak annotations were used for each beat, which is equivalent to employing an infallible algorithm to estimate the R peaks on the ECG signal. Thus, the results depend only on the clustering methods and not on the R peak detection approach.

Segmentation: An asymmetric window of fixed size around the R peak was used. The length of the window was equal to 235 samples (i.e., the maximum value) including the R peak point. For the selection of the number of samples to the right and to the left around the R peak, the mean of the PR/QT is used for minimal and maxima values, yielding to 39 % @ 360 Hz, resulting in approximately 39% of the samples located on the left (92 samples) and 61% on the right (142 samples).

Classifiers: The classifier consist on the combination of three FE algorithms: DCT, PCA and KPCA with four types of machine learning techniques (MLP, LVQ, RBF, and SVM). From FE stage, it is possible to obtain the following number of components:

DCT	PCA	KPCA
$1 \leq K \leq V_L$	$1 \leq K \leq V_L$	$1 \leq K \leq M$

Where, V_L is the vector length (in samples), M is the vector number and K is the number of components for FE stage. In this case, V_L is equal to 235, M is equal to 2000, and K is equal to 10, 15 and 20 components (generating 12 classifiers in total).

To train the MLP classifiers, a network was created with hidden layer architecture: n - $2n$ - 1, i.e., n input neurons, $2n$ neurons in the hidden layer and one output neuron (architectures: 10 - 20 - 1, 15 - 30 - 1 and 20 - 40 – 1). The activation functions are hyperbolic tangent and linear in the hidden and output layers, respectively.

For the LVQ classifier, n neurons in the input layer and two neurons in the output layer (2 classes) are employed while for the RBF classifier uses a simple algorithm to search the optimal dispersion parameter in the range (0.1 - 10). Each classifier has n radial basis neurons and one linear output neuron.

The SVM classifier is the only one for which the number of beats (N) is reduced to 1000 due to high computational cost. Thus, one might expect a lower performance, because only a quarter of the available set is employed.

The performance evaluation for each classifier was carried out by computing three indexes: Specificity (Sp), Sensitivity (Se) and Positive Predictive Value (PPV), from the confusion table defined for the NORM and ARTF classes (Figure 3):

Estimated Classes		
NORM	TP	FP
ARTF	FN	TN
	NORM	**ARTF**
	Real classes	

$$Sp(\%) = \frac{TN}{TN + FP} \times 100 \qquad Se(\%) = \frac{TP}{TP + FN} \times 100, \qquad PPV(\%) = \frac{TP}{TP + FP} \times 100$$

Fig. 2. Confusion matrix for each classifier. TP: true positive, TN: true negative, FP: False positive, FN: false negative, Se: Sensitivity, Sp: Specificity, PPV: positive predictive value.

2.3 Validation and Comparation of Classification Methods

To test whether the differences between classifiers are statistically significant, we used the McNemar's Test, based on the calculation of McNemar statistic, defined as

$$z = \frac{|n_{01} - n_{10}| - 1}{\sqrt{n_{01} + n_{10}}} \tag{1}$$

Where:
n_{01}: number of miss-classified samples by A but not by B.
n_{10}: number of miss-classified by B but not by A.

The statistic $|z|$ was calculated for all possible combinations of each pair of classifiers. The null hypothesis H0 (the classifiers have the same error) can be rejected with an error probability of 0.05 if $|z| > 1.96$. The alternative hypothesis H1 is that the classifiers have different errors, i.e., the differences in performance indexes are statistically significant.

3 Results and Discussion

3.1 Partial Evaluation

Tables 1 to 6 show the performance indexes of the four classifiers with 10, 15 and 20 components (features) both for training and validation.

Table 1. All Classifiers using DCT (Training)

Training									
Classifier	***Sp*** **(%)**			***Se*** **(%)**			***PPV*** **(%)**		
	10	**15**	**20**	**10**	**15**	**20**	**10**	**15**	**20**
DCT+ MLP	97.8	97.1	97.4	97.7	97.8	97.6	97.8	97.1	93.2
DCT+LVQ	75.8	83.6	79.4	91.6	79.5	95.6	79.5	84.8	76.2
DCT+RBF	87.2	87.6	90.8	86.6	87.3	93.6	87.3	88.0	84.0
DCT+SVM	72.8	80.9	82.0	80.1	76.0	81.7	76.0	82.3	73.8

Table 2. All Classifiers using DCT (Validation)

Validation									
Classifier	***Sp*** **(%)**			***Se*** **(%)**			***PPV*** **(%)**		
	10	**15**	**20**	**10**	**15**	**20**	**10**	**15**	**20**
DCT+ MLP	93.3	93.2	94.6	94.2	96.3	96.6	93.2	93.3	94.6
DCT+LVQ	71.8	80.8	75.2	91.7	88.9	96.2	76.2	82.0	79.2
DCT+RBF	83.9	82.9	89.9	85.9	89.7	92.5	84.0	83.8	90.0
DCT+SVM	72.5	77.5	79.8	79.6	81.2	81.0	73.8	77.9	79.6

Table 3. All Classifiers using PCA (Training)

Training									
Classifier	***Sp*** **(%)**			***Se*** **(%)**			***PPV*** **(%)**		
	10	**15**	**20**	**10**	**15**	**20**	**10**	**15**	**20**
PCA+ MLP	96.9	98.6	98.5	97.8	97.7	97.4	97.0	98.6	98.5
PCA+LVQ	80.2	79.0	84.7	93.8	93.1	95.4	82.8	81.3	86.4
PCA+RBF	90.1	90.4	93.5	89.4	90.5	93.0	90.2	90.6	93.6
PCA+SVM	82.8	84.4	84.4	79.2	84.6	85.9	83.2	85.4	85.6

Table 4. All Classifiers using PCA (Validation)

Validation									
Classifier	***Sp*** **(%)**			***Se*** **(%)**			***PPV*** **(%)**		
	10	**15**	**20**	**10**	**15**	**20**	**10**	**15**	**20**
PCA+ MLP	94.2	95.9	95.9	94.5	96.2	94.9	94.2	95.9	95.8
PCA+LVQ	78.1	81.9	82.2	91.3	94.1	93.8	80.4	84.1	83.9
PCA+RBF	89.2	88.9	92.3	88.5	88.5	91.5	89.0	88.7	92.1
PCA+SVM	81.4	82.5	82.1	81.3	84.7	84.2	81.0	82,6	82.2

Table 5. All Classifiers using KPCA (Training)

Training									
Classifier	***Sp*** **(%)**			***Se*** **(%)**			***PPV*** **(%)**		
	10	**15**	**20**	**10**	**15**	**20**	**10**	**15**	**20**
KPCA+MLP	96.8	98.4	98.3	98.0	98.6	98.8	96.9	98.4	98.3
KPCA+LVQ	84.4	84.2	93.9	89.1	93.1	90.0	85.3	85.7	90.5
KPCA+RBF	87.6	91.8	90.1	88.8	88.6	94.9	87.9	91.7	90.7
KPCA+SVM	79.3	89.8	91.1	84.4	86.7	87.5	81.4	90.2	91.3

Table 6. All Classifiers using KPCA (Validation)

Validation									
Classifier	***Sp*** **(%)**			***Se*** **(%)**			***PPV*** **(%)**		
	10	**15**	**20**	**10**	**15**	**20**	**10**	**15**	**20**
KPCA+MLP	94.0	96.8	96.0	94.9	96.3	95.9	94.0	96.8	96.0
KPCA+LVQ	83.0	82.9	89.5	88.0	90.4	91.9	83.6	83.9	89.6
KPCA+RBF	87.0	91.2	88.6	88.3	88.5	92.7	87.0	90.8	88.9
KPCA+SVM	79.8	88.6	89.2	84.5	87.4	89.0	80.4	88.2	88.9

From the above results, it is evident that the MLP classifier gives the best results for all features extraction variants.

3.2 Global Evaluation

Table 7 shows the results of the evaluation for the entire database (110192 beats). Only, 96 beats were excluded: the first and the last of each record. It is evident that for 10 components in the feature vector, the PCA + MLP method outperforms in specificity and positive predictive value to the other two methods, although it is less sensitive than the DCT + MLP and KPCA + MLP, in that order. For 15 and 20 components, the KPCA + MLP method is better than DCT + MLP and PCA + MLP, showing indexes greater than (or equal to) the best shown by the other two methods.

Table 7. All Classifiers using MLP

Validation									
Classifier	**DCT**			**PCA**			**KPCA**		
	10	**15**	**20**	**10**	**15**	**20**	**10**	**15**	**20**
Sp **(%)**	93.9	95.4	96.2	95.4	95.9	96.7	94.6	96.4	96.7
Se **(%)**	94.6	95.2	94.7	93.9	94.8	93.9	94.4	95.2	95.4
PPV **(%)**	97.1	97.8	98.1	97.8	98.0	98.4	97.4	98.2	98.4

The experiments show that KPCA has a higher performance than PCA and DCT. It can be explained by the capability of nonlinear PCA algorithms to capture nonlinear correlations between the data. It leads to an excellent trade-off in to preserve the biggest information with a minimum number of features.

The values of statistic $|z|$ are shown in Table 8 for all possible combinations of each pair of classifiers. The value for each pair of classifiers is obtained by intercepting the row and column of the table. In all cases, the null hypothesis has to be rejected meaning that differences in the performance indexes for each classifier are statistically significant among methods, validating the results.

Table 8. Results of McNemar's Test for all classifiers and beats the database. The grouping method in MLP is all cases.

	PCA10	**PCA15**	**PCA20**	**DCT10**	**DCT15**	**DCT20**	**KPCA10**	**KPCA15**
PCA10								
PCA15	12.5							
PCA20	7.1	5.6						
DCT10	100.8	106.3	104.2					
DCT15	154.7	159.9	158.8	74.3				
DCT20	170.9	176.2	173.7	88.1	28.1			
KPCA10	208.5	210.5	209.6	140.3	80.2	60.9		
KPCA15	218.8	222.3	221.5	151.0	98.5	83.7	25.3	
KPCA20	193.3	197.1	195.5	119.3	53.6	30.6	32.7	58.1

It was not possible to compare the results with other studies about artefacts recognition, because there are no other publications to our knowledge about this particular topic.

4 Conclusions

The trade-off between performance indexes suggests the use of the non-linear principal component analysis as feature extraction method and a multilayer perceptron as clustering method. In spite of its high runtime, it can be implemented with reasonable resources taking into account the current computer technologies. The future improvement and optimization of KPCA algorithm could ensure its practical application with a greater efficiency and speed, for example, using programmable devices to accelerate the calculation of principal components.

References

1. Hosseini, H.G., Luo, D., Reynolds, K.J.: The comparison of different feed forward neural network architectures for ECG signal diagnosis. Medical Engineering & Physics 28, 372–378 (2006)
2. Mitra, S., Mitra, M., Chaudhuri, B.B.: Pattern defined heuristic rules and directional histogram based online ECG parameter extraction. Measurement 42, 150–156 (2009)
3. Engin, M.: ECG Beat Classification Using Neuro-fuzzy Network. Pattern Recognition Letters 25, 1715–1722 (2004)
4. Joy, M.R., Chakraborty, C., Ajoy, K.R.: A two-stage mechanism for registration and classification of ECG using Gaussian mixture model. Pattern Recognition 42, 2979–2988 (2009)
5. Acir, N.: A support vector machine classifier algorithm based on a perturbation method and its application to ECG beat recognition systems. Expert Systems with Applications 31, 150–158 (2006)
6. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology: "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use". European Heart Journal 17, 354–381 (1996)
7. Ceylan, R., Özbay, Y.: Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network. Expert Systems with Applications 33, 286–295 (2007)
8. Patra, D., Das, M.K., Pradhan, S.: Integration of FCM, PCA and Neural Networks for classification of ECG arrhythmias. IAENG International Journal of Computer Science 36(3) (2009)
9. Özbay, Y., Ceylan, R., Karlik, B.: A Fuzzy Clustering Neural Network Architecture for Classification of ECG Arrhythmias. Computers in Biology and Medicine 36, 376–388 (2006)
10. Karpagachelvi, S., Arthanari, M., Sivakumar, M.: ECG Feature Extraction Techniques-A Survey Approach. International Journal of Computer Science and Information Security 8(1), 76–80 (2010)
11. Neto, J.E., Suárez-León, A.A., Vázquez Seisdedos, C.R., López-Mora, N.A., Leite, J.C., Oliveira, R.C.L.: Métodos de extracción de características en el ECG: análisis comparativo. In: V Latin American Congress on Biomedical Engineering, CLAIB 2011, Cuba. IFMBE Proceedings, vol. 33, pp. 858–861 (2011) (in Spanish), `http://link.springer.com/content/pdf/10.1007%2F978-3-642-21198-0_218`
12. Moody, G.B., Mark, R.G.: The MIT-BIH Arrhythmia Database on CD-ROM and software for use with it. Computers in Cardiology 17, 185–188 (1990)
13. Köhler, B.U., Hennig, C., Orglmeister, R.: The Principles of Software QRS Detection. Reviewing and Comparing Algorithms for Detecting this Important ECG Waveform. IEEE Engineering in Medicine and Biology, 42–57 (2002)