Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 11, November -2017

# Analysis of Automatic Classification of Electrocardiogram (ECG) Beats Using Wavelet Transform and SVM and PCA-SVM

*Shantanu Choudhary[1], S S Mehta[2]*

[1]*Lecturer Electrical Engineering, Govt. Polytechnic College, Government of Rajasthan, Jodhpur, India- 342001*
[2] *Professor Electrical Engineering, MBM Engineering College, Jai Narain Vyas University, Jodhpur, India- 342001*

**Abstract —** *A method for the automatic classification of cardio beats from an electrocardiogram (ECG) is presented in the paper. This beats classification is based on an analysis of QRS and DWT based feature extraction. The principal component analysis (PCA) is used for parameter analysis and recognition of cardiac beats. These parameters are calculated for beats with 4 types of classes (L, A, P and R) from ECG records retrieved from the MIT-BIH arrhythmia database. Further SVM is applied as classifier for automatic detection of heart beats. Analysis of the different groups shows the overall recognition performance was 96.43% with SVM and 97.75% with PCA-SVM.]*

***Keywords***- *ECG, PCA, MIT-BIH, SVM, QRS and DWT.*

## I. INTRODUCTION

The Wavelet Transform in recent years has become a technique that has been the object of study by researchers in the analysis of signals from a wide variety of areas of Science, Engineering and Medicine.

The Wavelet Transform analysis is being applied to a wide variety of biomedical signals including electromyographic (EMG) signals, electroencephalographic (EEG) signals, clinical sounds, respiratory patterns, blood pressure trends, and deoxyribonucleic acid sequences (DNA), along with the signals object of this project the electrocardiographic (ECG).

At present there are a large number of applications for the processing of signals of physiological origin, mainly due to the complexity in the extraction of rules and specific characteristics for the implementation of algorithms that unequivocally reflect the medical knowledge derived from the interpretation of the biological / biomedical signals treated.

A bibliographical review of the methods and techniques used in the pre-processing, feature extraction, segmentation, as well as the sources related to the reduction of characteristics related to the stages of the ECG signal analysis process is discussed.

In summary, it is a question of elaborating the framework in which the present project fits, reviewing the investigations carried out in the study and analysis of the biomedical signals, the electrocardiographic signals in particular, situating, in this way, the context in the will locate the present study.

To this end, and as a summary of the bibliographical references studied regarding the work on ECG signals, it is observed that a scheme commonly accepted in this process is shown in Figure 1, structured in several stages.

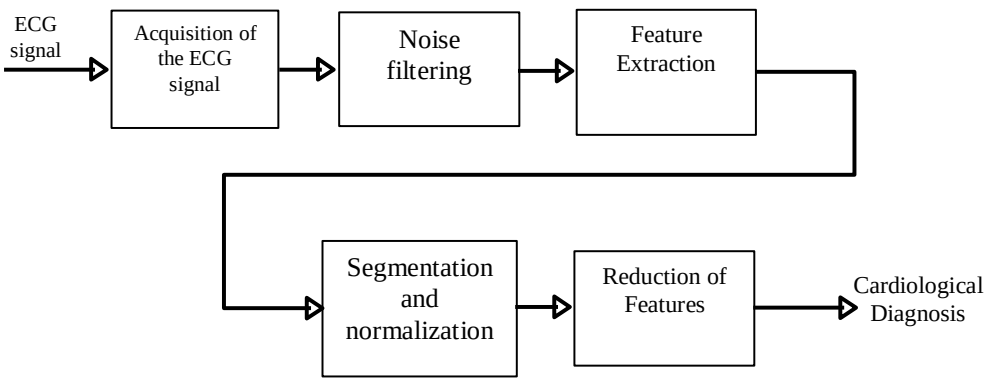


***Figure 1. General process developed in the analysis of the beats of a Holter ECG signal***

 

The analysis and classification of the heart beats can be made from temporal and morphological studies of the electrocardiographic signal.

In general, to achieve the characterization of the shape of a QRS complex by means of its destinguibles points, it is inevitable the accumulation of large amounts of data that are not always essential to obtain results accurate.

The Principal Component Analysis or PCA (Principal Component Analysis), is a statistical synthesis technique of information or reduction of the dimension (number of variables). In data banks with many variables, the PCA technique allows to reduce the number of such variables, without losing substantial information. The new factors or components will be a linear combination of the original variables, and independent of each other.

A key aspect in the Analysis of Principal Components is the interpretation of the factors, which is not given to priori, but it is deduced after observing the relation of the results with the initial variables. This task is not always trivial.

The objective of the present work is to extract the points of the electrocardiographic signal (ECG) that describe each one of the QRS complexes for their morphological analysis and then apply the principal component analysis techniques (PCA) to minimize data redundancy and obtain an appropriate model.

## I. MATERIALS AND METHODS

### A. *Electrocardiographic Signal*

A 5-minute extract of signal No. 100 from the MIT-BIH arrhythmia database was used. This signal belongs to the group of electrocardiographic signals considered normal, that is without arrhythmias.

### B. *Algorithm for the determination of QRS points*

The algorithm uses both the electrocardiogram signal and the annotation file also extracted from the base of Arrhythmia data from MIT-BIH corresponding to that signal. This log file contains, in the first column, the time in which the Q wave occurs in each ventricular beat (which marks the beginning of the QRS complex); in another column the order of the sample corresponding to that data is indicated; the code of classification of the beat ("N" if it is a normal beat). The first two rows of the archive.

0: 00.194 70 N
0: 01.005 362 N (1)

The algorithm, designed with MATLAB® (The MathWorks, Natick, Massachusetts, USA), combines the two files, and from the agreement between the data file and the signal annotation, a matrix with the data is generated that correspond to the QRS complex, one row for each beat. To determine the end of the event, the S wave is identified which is the first point of inflection after the peak R (Figure 2). The turning point is recognized by a change in the sign of the slope, or by having a zero slope or by a significant change thereof. It was determined for it the derivative of the difference of two points shown in equation 1. When this minimum point is detected, the charge of the matrix for that heartbeat.

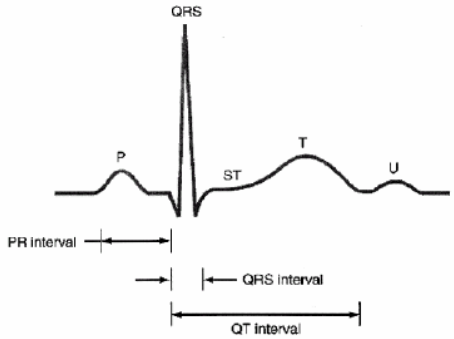


***Figure 2. QRS complex. Determination of QRS complex through peaks Q and S***

Irregular beats form two groups: the first comprises fibrillation and ventricular tachycardia which are dangerous and the ECG signal represents the electrical activity of the need for immediate therapy with a heart, is probably the diagnostic technique of the defibrillator; the second group concerns the most common cardiac pathologies. ECG is non-

 

hazardous arrhythmias but which require a non-stationary randomized signal, structured preventive therapy of other problem. These disorders (Figure 2) by the succession of wave and cardiac forms are also classified according to the frequency and the interval of the (P, Q, R, S, and T). Any rhythm modification described by the number of cycles or morphological or temporal contractions of its events (or ventricular (depolarization of the ventricles) by chronic) constitutes a cardiac pathology. The minutes, marked on the ECG by the number of changes concerning the complex rhythm or frequency QRS are cardiac arrhythmias (fatal diseases). A number of works [1], [2], [3], [4], [5], [6], and [7] have been successfully performed for clinical and technological problem highlighting arrhythmias, but they defy by (means and approaches). They are varied and the choice of types and number of parameters to approach them are also varied, significant that characterize the arrhythmia in question. It is possible to define the premature ventricles) in cardiac arrhythmias by the irregularity of the beats [6] [8] with cardiac arrhythmias or as a group of different parameters, approaches and results.

On the other hand, several classification methods have been proposed in the literature to classify the arrhythmic beats in normal or abnormal (arrhythmic) beats for some and others according to different rhythms (TV and BV) [4] , [9] and [10] compares three different procedures to characterize the rhythms to be classified by an LVQ network. But in most cases the essential parameters are measured around the QRS complex (width, shape and amplitude) and more particularly the detection of the R wave which is a dominant parameter [4], [6], [9]. Especially for atrial arrhythmias (atrial fibrillation) the P wave and the PR interval are privileged.

Inspired by several works cited, we propose an approach of automatic recognition to detect several arrhythmias. It involves methods of non-linear analysis by neural networks and tools of non-stationary analysis by the wavelet transform (WT) which have very interesting adaptation properties.

Currently the wavelet method has become a technique, par excellence, very effective to extract the parameters that characterize the arrhythmia in its various forms. Many works using wavelets have been successfully carried out, in particular the articles [11], [12] and [13] have amply demonstrated the contribution of the ECG signal. It makes it possible to produce the start and end times of waves and intervals and to filter the signal.

### *C. Feature Extraction*

The matrix containing the data describing each QRS complex has a total of 30 columns. In most of the cases the beats do not exceed an average of 20 samples, in the case of normal beats. That is to say that there is data redundancy. The PCA method aims to eliminate this redundancy.

$$y(nT) = \frac{1}{2}[x(nT) + x(nT - T)] \tag{2}$$

Let $y(t)$ be the signal (ECG) to be analyzed, the mother wavelet $\psi(t)$ is chosen to serve as a prototype for all the windows. All the windows that are used (daughter wavelets) are dilated (or compressed) and offset versions of the mother wavelet.

The DWT is a function of two parameters τ and s which is the coefficient of the (WT) of the signal $x(t)$ using the parent ocean of analysis $\psi(.)$:

$$DWT_X^{\psi}(\tau, s) = \psi_X^{\psi}(\tau, s) = \int_{-\infty}^{\infty} Y(t)\, \psi_{\tau,s}^{*} \tag{3}$$

Once the mother wavelet is chosen, the calculation begins with $s = 1$. This first value of s corresponds to the most compressed wavelet, the DWT is calculated for all the values of s <1 and s>1. Then the Two parameters τ and s are increased by a sufficiently small step. This corresponds to the sampling of the time scale plan.

### *D. Dimension Reduction*

*PCA Technique*

In many applications, a set of n objects are represented through a collection of m descriptors, indexes or parameters. In some cases, m is a very large number, which hinders the analysis of the data set in all its dimensionality; that is to say that the n objects can be considered as n points located in a space of m dimensions. The objective is to classify those objects and represent them in a space with a smaller dimension p (p <m), in such a way that the projection in that space is optimal.

In the PCA methodology, the descriptors are ordered in a matrix A of dimension n×m (223×30 in our case). The mathematical criterion used to achieve the dimensionality reduction is such that, for a predetermined value of p, the

maximum total statistical variance of the original data is retained in that subspace. The most common is to obtain vector columns centred and normalized dimensionless. So to each column $a_i$ of matrix A,

$$A = (a_1\ a_2\ \dots\dots\ a_m) \tag{4}$$

The calculated average is

$$\bar{a}_i = \frac{1}{n}\ \sum_{i=1}^{n} a_{ij} \tag{5}$$

And the standard deviations multiplied by n,

$$s_j = \sqrt{\sum_{i=1}^{n}\left(a_{ij} - \bar{a}_j\right)^2} \tag{6}$$

Obtaining the following dimensionless matrix of variables:

$$A \to Z = (z_1\ z_2\ \dots\dots z_m) \tag{7}$$

Where each $z_j$ column vector is defined from the transformation

$$a_i \to Z_i = \frac{a_i - \bar{a}_i}{s_j} \tag{8}$$

The matrix of dimensionless homogenized variables allows to calculate the matrix of the correlation coefficients between each pair of data columns:

$$R = Z^T Z \tag{9}$$

This matrix is of dimension $m \times m$.

The principal components are given by the eigenvectors of the matrix R. All eigenvalues are not negative, since the matrix Z is obtained in such a way that it is defined non-negative. These eigenvalues are the parameters that indicate what fraction of the original total variance retains each new Main Component:

$$f_i = 100\frac{\lambda_i}{\sum_{i=1}^{m}\lambda_j}\% \tag{10}$$

Therefore, the order, from highest to lowest, of the eigenvalues induces an order of preference of the principal components. We will assume that:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \tag{11}$$

Now,

$$R \to X = (x_1\ x_2\ \dots\dots\dots x_m) \tag{12}$$

Where $x_1$ is the eigenvector associated with $\lambda_1$ , $x_2$ and $\lambda_2$, so on up to m.

The first Principal Component, x1 represents the largest amount of variance of the original data, $x_2$ retains the second greater variance, and so on up to m. The set of the Main Components generates a new coordinate matrix. To the coefficients of each own vector $x_j$ are called weights (loadings) and indicate which linear combinations of the original variables should be build to define the new dimensionless coordinates. The most usual, in order to reduce the dimensionality of the problem, is to choose p = 2 or p = 3 first Main components. In our case p = 5.

### *E. Classification using Support Vector Machine (SVM)*

Our objective is to evaluate the performance of a binary classifier with rejection, we first retained the annotations of the American Heart Association (AHA) [14] and the practical recommendations of the AAMI standard to form two classes of beats (normal beats and ectopic beats)

- The positive class (P) represents ectopic beats, premature beats and some unknown beats.
- Negative class (N) represents the normal beats (about 70% of the study base) and some abnormal beats LBBB (Left Bundle Branch Block), Right Bundle Branch Block (RBBB).

According to the AAMI (American Association for Medical Instrumentation) guidelines, records (102, 104, 107, 217) containing beats from pacemakers are excluded from this study. Recordings not containing PVC beats (11 records) are also excluded [15]. We then have 33 recordings of interest. Note that no selection of signals is based on their quality.

***Learning Base***

To constitute a good learning base that would allow us to generalize our classifier and thus obtain a global classifier, we took in a random way 10 records from which we used the beats present in the first 5 minutes, i.e. 1500 beats tagged by cardiologists.

The cardiac beats being segmented and quantified by discriminating parameters, are each represented by a characteristic vector.

 

***Test Basis***

From each recording, we use the last 25 minutes for the test phase. Thus, the learning base is completely dissociated from the test database. This allows us to evaluate the generalization capacity of our classification algorithm.

## II. RESULTS

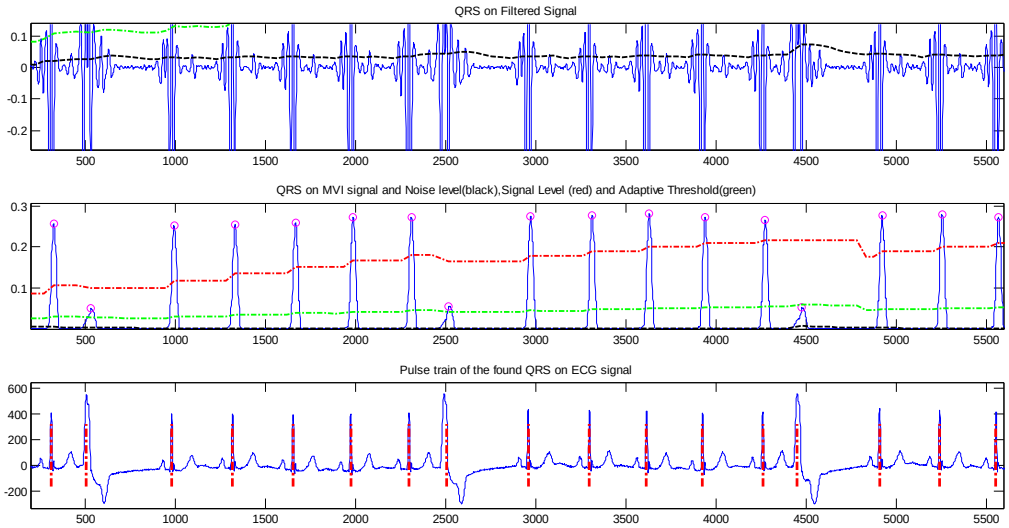


***Figure 3: QRS on filtered signal***

***Table 1. Accuracy with training overhead reduction***

| Methods | Accuracy | Training overhead |
|---|---|---|
| SVM | 95.7% | 1000x150 |
| PCA-SVM | 97.1% | 1000x118 |

After applying the tools provided by MATLAB for the determination of the Principal Components, obtained the correlation matrix that has a dimension of 30×30, which was arbitrarily transformed into the matrix X in ascending order according to the weights of the eigenvalues of the matrix.

This comparison helps to identify and focus the few vital factors differentiating them from the many useful factors. The application of the same allows to visually display in order of importance, the contribution of each element in the total effect.

## III. CONCLUSION

The Principal Component Analysis (PCA) method is effective, and allowed to meet the objectives mentioned above. It was possible to reduce a data matrix of 12 % of original data. With this tool, data redundancy was eliminated to speed up the computational times, which is a primary objective in information processing. Further SVM is applied as classifier for automatic detection of heart beats. Analysis of the different groups shows the overall recognition performance was 96.43 % with SVM and 97.75 % with PCA-SVM**.**

## REFERENCES

[1] Babloyantz, A. and lvanov, V.V., 1998. Neural networks in cardiac arrhythmias. In *Industrial Applications of Neural Networks* (pp. 403-417).

[2] Silipo, R. and Marchesi, C., 1998. Artificial neural networks for automatic ECG analysis. *IEEE transactions on signal processing*, *46*(5), pp.1417-1425.

[3] Gao, G.Q., 2003. *Computerised detection and classification of five cardiac conditions* (Doctoral dissertation, Auckland University of Technology).

 

[4] Exarchos, T.P., Tsipouras, M.G., Exarchos, C.P., Papaloukas, C., Fotiadis, D.I. and Michalis, L.K., 2007. A methodology for the automated creation of fuzzy expert systems for ischaemic and arrhythmic beat classification based on a set of rules obtained by a decision tree. *Artificial Intelligence in medicine*, *40*(3), pp.187-200.

[5] Tsipouras, M.G., Fotiadis, D.I. and Sideris, D., 2005. An arrhythmia classification system based on the RR-interval signal. *Artificial intelligence in medicine*, *33*(3), pp.237-250.

[6] Shyu, L.Y., Wu, Y.H. and Hu, W., 2004. Using wavelet transform and fuzzy neural network for VPC detection from the Holter ECG. *IEEE Transactions on Biomedical Engineering*, *51*(7), pp.1269-1273.

[7] Monzon, J.E. and Pisarello, M.I., 2005. Cardiac beat classification using a fuzzy inference system. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the* (pp. 5582-5584). IEEE.

[8] Saadi, S., Bettayeb, M., Guessoum, A. and Abdelhafidi, M., 2012. Artificial bee colony optimized neural network model for ECG signals classification. In *Neural Information Processing* (pp. 339-346). Springer Berlin/Heidelberg.

[9] Assaleh, K. and Al-Nashash, H., 2005. A novel technique for the extraction of fetal ECG using polynomial networks. *IEEE Transactions on Biomedical Engineering*, *52*(6), pp.1148-1152.

[10] Kristiansen, D.E., Husøy, J.H., Eftestøl, T., Erik, D. and John, K., 1995. Rhythm detection in ECG signals.

[11] Senhadji, L., Carrault, G., Bellanger, J.J. and Passariello, G., 1995. Comparing wavelet transforms for recognizing cardiac patterns. *IEEE Engineering in Medicine and Biology Magazine*, *14*(2), pp.167-173.

[12] Polat, K., Şahan, S. and Güneş, S., 2006. A new method to medical diagnosis: Artificial immune recognition system (AIRS) with fuzzy weighted pre-processing and application to ECG arrhythmia. *Expert Systems with Applications*, *31*(2), pp.264-269.

[13] Jaylaxmi C. Mannurmath, Raveendra M., "MATLAB Based ECG Signal Classification", International Journal of Science, Engineering and Technology Research (IJSETR), Vol. 3, Issue 7, July 2014.

[14] R. Mark, G. Moody. MIT-BIH Arrhythmia database directory. Massachusette Inst. Technol. (MIT), 1988

[15] Mark, R. and Wallen, R., 1987. AAMI-recommended practice: Testing and reporting performance results of ventricular arrhythmia detection algorithms. *Association for the Advancement of Medical Instrumentation, AAMI ECAR-1987*.

 