IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 5, Issue 9, April 2017)

ECG Signal Classification using PCA, DWT and Neural Network Classifier

Ankish Dangi
M. E. Scholar
Computer Science & Engineering
Shiv Kumar Singh Institute of Technology, Indore (India)
ankishdangi@gmail.com

Dr. G. D. Gidwani
Ex. Director
Shiv Kumar Singh Institute of Technology, Indore (India)
ed@sksits.com

Abstract –**This paper offers ECG signal classification system using Principal Component Analysis (PCA) technique to reduce the dimensionality of test signal. Discrete Wavelet Transform (DWT) is used for feature extraction. Power Spectral Density (PSD) is another feature for the spectrum of ECG. This process helps in enhancing the classification accuracy. Classification is done using Neural Network classifier. In this paper, the signal processing and neural network toolbox are used in MATLAB environment. The processed signal source came from the Massachusetts Institute of Technology Beth Israel Hospital (MIT-BIH) arrhythmia database which was developed for research in cardiac electrophysiology.**

Keywords – **DWT, ECG, MIT-BIH, PCA, PSD.**

I. INTRODUCTION

Electrocardiogram (ECG) is the record of the electrical potentials produced by the heart. The electrical wave is generated by depolarization and repolarization of certain cells due to movement of Na+ and k+ ions in the blood. The ECG signal is typically in the range of 2 mV and requires a recording bandwidth of 0.1 to 120 Hz [1]. The ECG is acquired by a non-invasive technique, i.e. placing electrodes at standardized locations on the skin of the patient [2]. The ECG signal and heart rate reflects the cardiac health of human heart. Any disorder in heart rate or rhythm or change in the morphological pattern of ECG signal is an indication of cardiac arrhythmia. It is detected and diagnosed by analysis of the recorded ECG waveform. The amplitude and duration of the P-QRS-T-U wave contains useful information about the nature of disease related to heart.

An Electrocardiogram or ECG is an electrical recording of the heart and is used in the investigation of heart disease. This ECG can be classified as normal and abnormal signals. One of the most important problems in ECG analysis is automatic beat delineation. This is needed in many cases ranging from simple heart rate computations to serving as the first stage of complex automatic diagnosis. Beat delineation techniques have to start by identifying features in the ECG signal that can constantly be detected in each heartbeat. Simply by looking at an ECG plot, it can be noticed that the QRS complex is the predominant feature in every beat. The other features of the ECG signal, like the P wave and T wave, are sometimes too small to be detected [3]. This makes the QRS complex the feature that can yield the best detection accuracy.

The generalization performance of the traditional classifiers are not sufficient for the correct classification of ECG signals. To overcome this problem Neural Network (NN) classifier is used which work by searching for the best value of the parameters that tune its discriminant function, and upstream by looking for the best subset of features that feed the classifier. The main problem in automatic ECG signal recognition and classification is that related features are very susceptible to variations of ECG morphology and temporal characteristics of ECG. This problem is resolved by applying dimensionality reduction using PCA before Discrete Wavelet Transform (DWT) to reduce the training overhead and thus we get higher accuracy.

The main objective of this paper is to propose a computer based automated system for classification of the ECG signal. Normal, Premature Ventricular Contraction (PVC) and Fusion signals of the MIT-BIH Database are used as reference. Principal Component Analysis (PCA) technique is used to reduce the dimensionality of test signal. Discrete Wavelet Transform is used to calculate spectral energy of ECG signal. Spectral flatness is considered as an additional feature and combining DWT and spectral flatness gives feature vector to

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 5, Issue 9, April 2017)

increase the accuracy. Classification is done using Neural Network classifier.

II. Proposed Method

Figure 4.1 shows the flow of proposed research work. The first stage is pre-processing stage including five levels of data processing which are Principal Component Analysis, Empirical Mode Decomposition (EMD), and Extraction of Attribute Vectors by the SVD method, Discrete Wavelet Transform (DWT), Power Spectral Density and trained by using Neural Network classifier to classify LBBB, NSR, AFIB, RBBB, SBR, AFL, PVC and BII features of ECG signal.

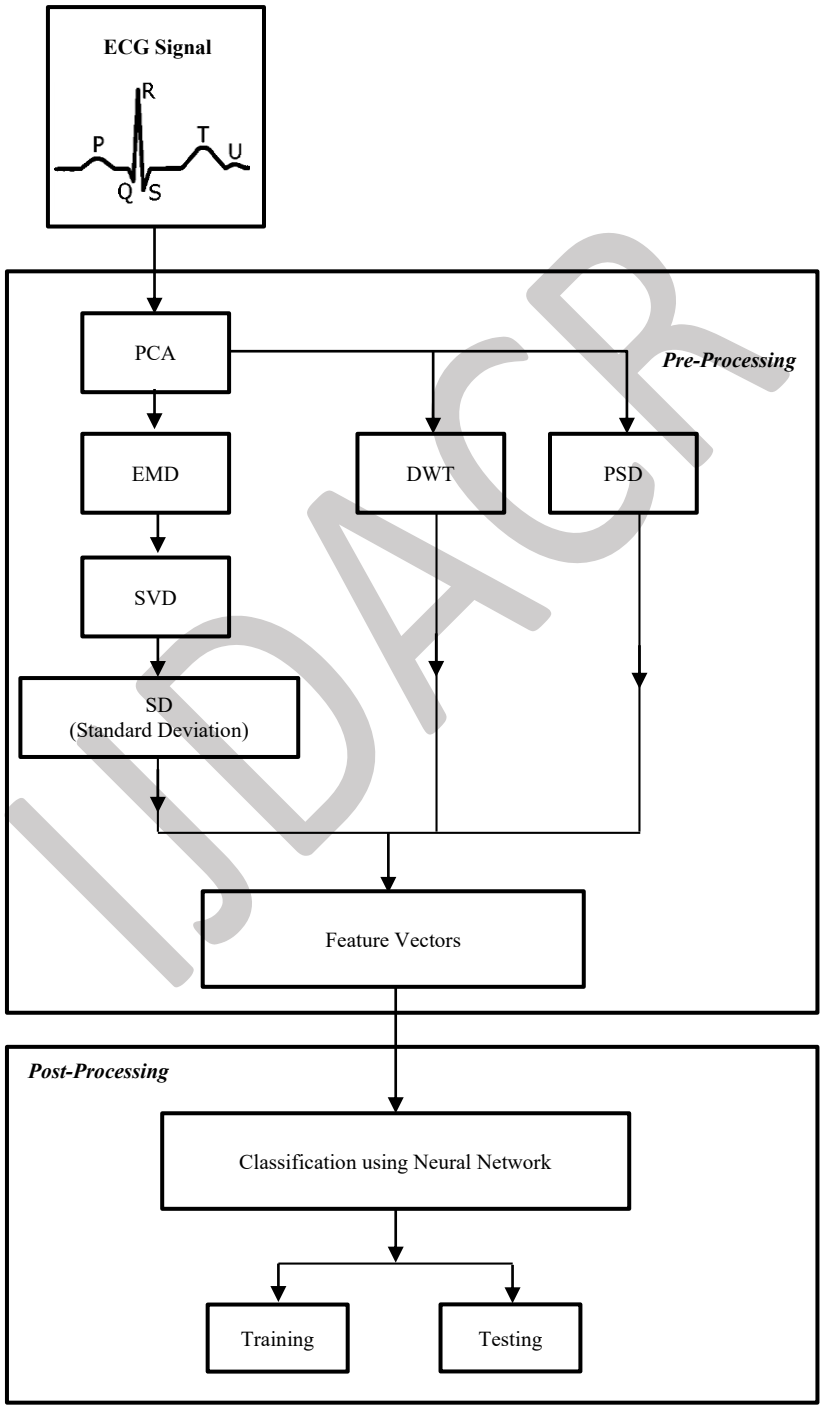

Figure 1: Flow diagram of proposed research work

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 5, Issue 9, April 2017)

Dimension Reduction using Principal Component Analysis

A gigantic measure of information is not effective to perform a pattern recognition process. In the third phase of preprocessing, PCA is applied to reduce the dimensionality.

As other transformation like Fourier transformation, PCA transform the data into another representation where a new set of basis vectors are used. In PCA, however, these basic vectors are not constant as in the case in many other transformation. Instead they are calculated based on the data to be transformed.

A. Statistics

The entire subject of statistics is based around the idea that we have this big set of data, and we want to analyze that set in terms of the relationships between the individual points in that data set.

B. Standard Deviation

To understand standard deviation, we need a data set. Statisticians are usually concerned with taking a sample of a population. To use election polls as an example, the population is all the people in the country, whereas a sample is a subset of the population that the statisticians measure. Here's an example set:

$$X = [1\ 2\ 4\ 6\ 12\ 15\ 25\ 45\ 68\ 67\ 65\ 98\ ]$$

X refers to this entire set of numbers. The mean of the sample is given by the formula:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad (1)$$

The Standard Deviation (SD) of a data set is a measure of how spread out the data is. SD is given by the formula

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{(n-1)}} \quad (2)$$

C. Variance

Variance is another measure of the spread of data in a data set. In fact it is almost identical to the standard deviation. The formula is simply the SD squared.

$$s^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{(n-1)} \quad (3)$$

D. Covariance

Covariance is always measured between two dimensions. If you calculate the covariance between one dimension and itself, you get the variance. So, if you had a three-dimensional data set (x, y, z), then you could measure the covariance between x and y dimensions, the x and z dimensions and the y and z dimensions. The formula for covariance is very similar to the formula for variance. The formula for variance could also be written like this,

$$var(x) = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(X_i-\bar{X})}{(n-1)} \quad (4)$$

The formula for covariance is given by,

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{(n-1)} \quad (5)$$

The exact value of covariance is not as important as its sign (i.e. Positive or negative). If the value is positive, then it indicates that both dimensions increase together. If the value is negative, then as one dimension increases, the other decreases. If the covariance is zero, it indicates that the two dimensions are independent of each other.

E. Covariance Matrix

Covariance (cov) is always measured between two dimensions. If we have a data set with more than two dimensions, there is more than one cov measurement that can be calculated. For example, from a three dimensional data set ($dimensions\ x, y, z$) you could calculate $cov\ (x, y)$, cov (y, z), cov (y, z). In fact, for an n dimensional data set, you can calculate different covariance values.

$$\frac{n!}{(n-2)!*2} \quad (6)$$

A useful way to get all the possible cov values between all the different dimensions is to calculate them all and put them in a matrix. So, the definition for the cov matrix for a set of data with n dimensions is:

$$C^{m\times n} = \left(c_{i,j}, c_{i,j} = cov(Dim_i, Dim_i)\right) \quad (7)$$

Where, $C^{m\times n}$ is a matrix with n rows and n columns and Dim_x is the x^{th} dimension. The entry on row 2, column 3 is the cov value calculated between the 2nd dimension and the 3rd dimension.

F. Eigenvalues and Eigenvectors of Matrix

If the matrix is small, we can compute them symbolically using the characteristic polynomial. However, this is often impossible for larger matrices, in which case we must use a numerical method.

An important tool for describing eigenvalues of square matrices is the characteristic polynomial: saying that λ is an eigenvalue of A is equivalent to stating that the system of linear equations $(A - \lambda I)\ v = 0$ (where I is the identity matrix) has a non-zero solution v (an eigenvector), and so it is equivalent to the determinant:

$$\det(A - \lambda I) = 0 \quad (8)$$

The function $p(\lambda) = det(A - \lambda I)$ is a polynomial in λ since determinants are defined as sums of products. This is the characteristic polynomial of A: the eigenvalues of a matrix are the zeros of its characteristic polynomial.

All the eigenvalues of a matrix A can be computed by solving the equation $pA(\lambda) = 0$. If A is a $n \times n$ matrix, then pA have degree n and A can therefore have at most n eigenvalues. Conversely, the

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 5, Issue 9, April 2017)

fundamental theorem of algebra says that this equation has exactly *n* roots (zeroes), counted with multiplicity. All real polynomials of odd degree have a real number as a root, so for odd n, every real matrix has at least one real eigenvalue. In the case of a real matrix, for even and odd *n*, the non-real eigenvalues come in conjugate pairs.

Once the eigenvalues λ are known, the eigenvectors can then be found by solving:

$$(A - \lambda I)v = 0 \quad (9)$$

An example of a matrix with no real eigenvalues is the 90-degree clockwise rotation (equation 9) whose characteristic polynomial is $\lambda^2 + 1$ and so its eigenvalues are the pair of complex conjugates $i, -i$. The associated eigenvectors are also not real.

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (10)$$

Empirical Mode Decomposition (EMD)

The

The total number of IMF components is roughly limited to log2*N*, where N is the total number of data points. It satisfies all the requirements for a meaningful instantaneous frequency through Hilbert transform. An EMD decomposes the original speech signal x (t) in to set of IMF through an iterative procedure called Sifting algorithm is given below.

Stepwise EMD algorithm can be described as follows:

Step 1: Initialize $-r_0(t) = x(t), i = 1, r_i(t) = r_0(t)$

Step 2: Procedure to extract the i^{th} IMF

- Initialize: $h_0(t) = r_i(t), J = 1$.
- Extract all the local minima and maxima of $h_{J-1}(t)$.
- Interpolate the local maxima and the local minima by a cubic spline to construct upper and lower envelopes of $h_{J-1}(t)$.
- Calculate the mean $m_{J-1}(t)$ of the upper and lower envelopes.
- $h_J(t) = h_{J-1}(t) - m_{J-1}(t)$
- if stopping criterion is satisfied then set $imfi(t) = h_J(t)$ else go to (b) with $J = J + 1$

Step 3: ri+1(t) = ri−1(t) − imfi(t)

Step 4: if ri+1(t)still has at least 2 extrema then go to 2 with i = i + 1 else the decomposition procedure ends. And ri(t)is the residue.

Finally, when EMD procedure is completed after n iterations the original signal can be reconstructed as:

$$x(t) = \sum_{j=1}^{n} imf_j + r(t) \quad (11)$$

Extraction of Attribute Vectors by the SVD Method

This approach extract relevant information using the SVD method. In other words, it extract the singular values of the matrix. Another concern is the extraction of the characteristics of the non-stationary signals that consist in approximating the Fourier transform representation in a finite number of rectangles using the singular vectors (SV) derived from the SVD method, so that the Fourier transform regions which have uniform energy will be represented by the rectangle approach. In the context of extracting the characteristics for the ECG signals, use the probability distribution function calculated from the two left and right straight. Let $S_{(m \times n)}$ be the time-frequency matrix of the transform in S. A decomposition in singular values of the matrix S is a factorization of the form:

$$S = U \Sigma V^T \quad (12)$$

With U is an orthogonal matrix of size $m \times m$ and V is an orthogonal matrix of size $n \times n$ and Σ is a diagonal matrix of dimension $m \times n$, with the singular values $\sigma_{ij} = 0$ if $i \neq j$ and $\sigma_{ij} \geq 0$ otherwise.

Discrete Wavelet Transform (DWT) is a mathemarical tool for dynamic disintegration of a signal. It is useful for changing of non-stationary signals. The transform is centered on wavelets of variable frequency and obliged time duration. Wavelet transform gives both frequency and spatial delineation for a signal. Not at all like conventional Fourier transform, temporary information is held in this system. Wavelets are made by understandings and extensions of a modified capability called mother wavelet. This segment details about the correctness of DWT and gives preferences of utilizing DWT as against different transformation methods.

The DWT is only an arrangement of filters. There are two filters incorporated;

Wavelet Filter

Scaling filter.

The scaling filter is a low pass filter while the wavelet filter is a high pass filter. DWT incorporates numerous sorts of transforms, for example, Daubechies wavelet, Haar wavelet and others. We have utilized Haar wavelet as a part of present research work. Figure 4.3 demonstrates the capacity of DWT for measurement reduction.

IJDACR
ISSN: 2319-4863

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 5, Issue 9, April 2017)

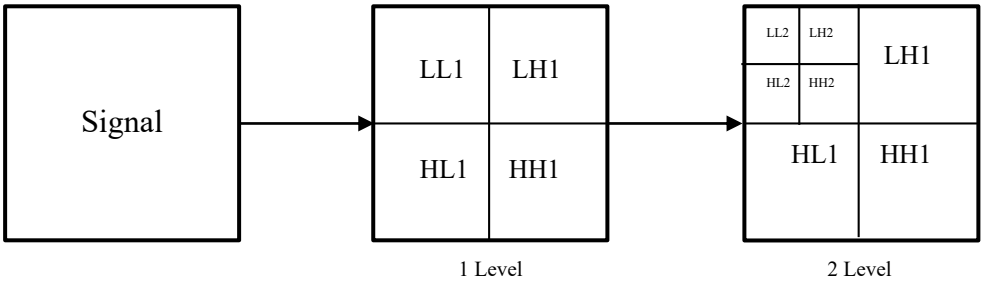

Figure 2: Function of DWT for dimension reduction

Power Spectral Density (PSD)

In parametric spectral estimation, one assumes that the signal is modeled by a stationary process which has a spectral density function (SDF) $S(f; a_1, \dots, a_p)$ that is a function of the frequency f and p parameters $a_1, \dots, a_p$. The estimation problem then becomes one of estimating these parameters.

The most common form of parametric SDF estimate uses as a model an autoregressive model $AR(p)$ of order p. A signal sequence $\{Y_t\}$ obeying a zero mean $AR(p)$ process satisfies the equation:

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + \epsilon_t \quad (13)$$

Where $\emptyset_1, \dots . \emptyset_p$ are fixed coefficients and ϵ_t is a white noise process with zero mean and innovation variance σ_p^2. The SDF for this process is

$$S(f; \emptyset_1, \dots, \emptyset_p, \sigma_p^2) = \frac{\sigma_p^2 \Delta t}{\left|1 - \sum_{k=1}^{P} \emptyset_k e^{-2i\pi f k \Delta t}\right|^2} \quad (14)$$

$$|f| < f_N$$

With Δt the sampling time interval and f_N the Nyquist frequency. There are a number of approaches to estimating the parameters $\emptyset_1, \dots, \emptyset_p, \sigma_p^2$ of the $AR(p)$ process and this the spectral density. The Yule-Walker estimators are found by recursively solving the Yule-Walker equations for a $AR(p)$ process.

Neural Network Classifier

Back Propagation Neural Network (BPNN) generates complex decision boundaries in feature space. BPNN in specific circumstances resembles Bayesian Posterior Probabilities at its output. These conditions are essential to achieve low error performance for given set of features along with selection of parameters such as training samples, hidden layer nodes and learning rate. In else case, the performance of BPNN could not be evaluated. For W number of weights and N number of nodes, numbers of samples (m) are depicted to correctly classify future samples in following manner:

$$m \geq O\left(\frac{W}{\epsilon} \log \frac{N}{\epsilon}\right) \quad (15)$$

The theoretical computation of number of hidden nodes is not a specific process for hidden layers. Testing method is commonly entertained for selection of these followed in the constrained environment of performance [4].

Back-Propagation Algorithm

The back-propagation algorithm for a 3-layer network (only one hidden layer) is as follows:

initialize the weights in the network (often small random values)
do
  for *each data i in the training set of database O = neural-network-output(network, i)*
  T = desired output for i
  calculate error (T - O) at the output units;
  calculate δ_h for all weights from hidden layer to output layer;
  calculate δ_i for all weights from input layer to hidden layer;
  update the weights to minimize error in the network; ***until*** *some stopping criterion satisfied*
return *the network*

III. Simulation And Results

The performance of proposed algorithms has been studied by means of MATLAB simulation.

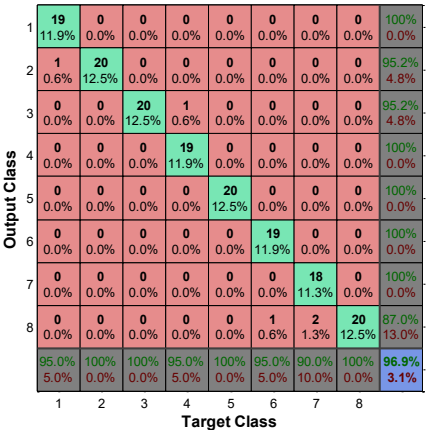

Figure 3: Confusion Matrix plot for proposed scheme with dimension reduction using PCA

IJDACR
International Journal Of Digital Application & Contemporary Research

International Journal of Digital Application & Contemporary Research
Website: www.ijdacr.com (Volume 5, Issue 9, April 2017)

The row and column are the class of ECG samples, taken from the database. There are 8 set of classes and each class having 20 ECG samples of with different classes of beats i.e. LBBB, NSR, AFIB, RBBB, SBR, AFL, PVC and BII. The associated accuracies achieved for different classes are 95%, 100%, 100%, 950%, 100%, 95%, 90%, and 100% respectively.

The confusion matrix plot indicates the overall accuracy i.e. 96.9% for proposed scheme with dimension reduction using PCA.

IV. Conclusion

Automatic detection of heart arrhythmias could be very important in clinical usage and lead to early detection of a fairly common malady and could help contribute to reduced mortality. In this research work, the use of Neural Networks for classification of the ECG beats is presented. Several stages of pre-processing have been used in order to prepare the most appropriate input vector for the neural classifier.

It was found that the accuracy of proposed algorithm is nearly around 96.9.5 % for proposed scheme with dimension reduction using PCA

Reference

[1] R. Acharya, J. S. Suri, J. A. E. Spaan and S .M. Krishnan, “Advances in Cardiac Signal Processing”, Springer Berlin Heidelberg, New York , 2007.

[2] J. Moss and S. Stern., “Non-invasive electrocardiology: Clinical aspects of holter monitoring”, 1st Edition, W.B. Saunders, Philadelphia, 1996.

[3] Yang Xiao, Hui Chen, Frank, “Handbook on Sensor Networks”, Published by World Scientific Publishing Co. Pvt. Ltd., 2010.

[4] Christos Stergiou and Dimitrios Siganos, “Neural Networks”, Report available at: http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html

[5] 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.

[6] X. Tang, L. Shu, “Classification of Electrocardiogram Signals with RS and Quantum Neural Networks”, International Journal of Multimedia and Ubiquitous Engineering, Vol. 9, No. 2, PP. 363-372, March 2014.

[7] Taiseer Mohammed Siddig, Mohmmed Ahmed Mohmmed, “A Study of ECG Signal Classification using Fuzzy Logic Control”, International Journal of Science and Research (IJSR), ISSN: 2319-7064, Vol. 3 Issue 2, February 2014.

[8] R. Kavitha, T. Christopher, “A Study on ECG Signal Classification Techniques”, International Journal of Computer Applications, ISSN: 0975-8887, Vol. 86, No. 14, January 2014.