ECG LEAD SELECTION AND MORPHOLOGICAL FEATURE EXTRACTION USING PCA

Manish Yadav, Vikas Sharma, Ashish Kulshrestha ,Nupur Yadav

Abstract— The processing of electrocardiogram signals (ECG) using principal component analysis (PCA) is presented in this paper. The proposed framework enables complex ECG signal processing with all the necessary steps resolved by PCA approach. Firstly, PCA is used for lead selection from 12 Lead ECG signals. Secondly, morphological features are extracted from optimal ECG lead like QRS complex, T waves and P wave. The methods are applied to the MIT/BIH 12 Leads ECG database. The results are particularly efficient promising that this new approach can be useful for lead selection, data compression, and morphological features extraction, of ECG signals. The main aim of this paper is its catholic presentation of Principal Component analysis (PCA) theory and its applications in the field of biological signal processing.

Index Terms— Electrocardiogram (ECG); Principal Component Analysis (PCA), QRS Complex, Base line Wander (BLW)

1 INTRODUCTION

Principal component analysis is a mathematical method that transforms a large set of correlated variables into a small set of uncorrelated variables called principal components. Every principal component provides new source of information regarding the data set and arranged so that the earliest few principal components contain most of the variability. In signal processing applications, PCA is applied on a set of time samples rather than on a data set of variables. If the signal is periodic in nature, like the ECG signal, the investigation is often based on samples extracted from the same segment location of different periods of the signal.

PCA is one of the most established techniques in multivariate statistical analysis and has been applied to ECG data compression, de-noising and de-correlation of noisy and useful ECG components or signals. The idea to use PCA is due to the fact that, in most of the ECG data there is a large amount of noise and artifacts as well as redundant information which is unnecessary for diagnostic applications.

Signal processing is today found in virtually any system for ECG analysis, and has clearly demonstrated its importance for achieving improved diagnosis of a wide variety of cardiac pathologies. Signal processing is employed to deal with diverse issues in ECG analysis such as data compression, beat detection and classification, noise reduction, signal separation, and feature extraction.

- *Mr. Manish Yadav is Asst. Prof. Dept. of electronics & communication engg. in JECRC Jaipur India,. E-mail: manishyadav.ece@jecrc.ac.in*
- *Mr. Vikas Sharma is Asst. Prof. Dept. of electronics & communication engg. in JECRC Jaipur India,. E-mail: vikassharma.ece@jecrc.ac.in*
- *Mr. Ashish Kulshrestha is Asst. Prof. Dept. of electronics & communication engg. in JECRC Jaipur India,. E-mail: ashishkulshrestha.ece@jecrc.ac.in*
- *Mrs Nupur Yadav is Asst. Prof. Dept of electrical engg. in JECRC Jaipur India,. E-mail: nupuryadav.ee@jecrc.ac.in*

Principal component analysis has become an important tool for successfully addressing many of these issues, and was first considered for the purpose of efficient storage retrieval of ECGs. Over the years, this issue has remained central as a research topic, although the driving force has gradually changed from having been tiny hard disks to become slow transmission links. Noise reduction may be closely related to data compression as reconstruction of the original signal usually involves a set of eigenvectors whose noise level is low, and thus the reconstructed signal becomes low noise; such reduction is, however, mostly effective for noise with muscular origin. Classification of waveform morphologies in arrhythmia monitoring is another early application of PCA, in which a subset of the principal components serves as features which are used to distinguish between normal sinus beats and abnormal waveforms such as premature ventricular beats.

This paper presents the overview of PCA in ECG signal processing. Section 2 contains a brief description of ECG fundamentals and an explanation of the ECG waveform. The remaining sections of the paper are focus on the use of PCA in ECG applications. The present overview is confined to those particular applications where the output of PCA, is considered, whereas applications involving ECG lead selection out of 12 leads ECG Signals. The latter types of applications include Principal Component analysis-(PCA)-based techniques for morphological features extraction of the optimal ECG leads.

1.1 ECG SIGNALS

ECG signals are biomedical signals originating from the actions of the human heart. The human heart operates cyclically pumping blood into the arterial system, which causes bioelectrical variations detectable from human skin. The action of the heart causing the ECG signal is depicted in Figure .1, where we have one real cardiac cycle of an ECG signal. The whole ECG recording consists of several consecutive cardiac cycles. In the cardiac cycle there are

fairly periodic waves and peaks corresponding to the consecutive phase of the heart's action.

A cardiac cycle of an ECG signal includes a QRS complex as its primary and most dominant sub pattern. Before the QRS complex there is a P wave in the ECG, and after the QRS complex a T wave, which is larger than the P wave. Flat segments between the above-mentioned components are PQ, ST and TP segments. An RR interval, which is the delay between two consecutive QRS complexes, gives us useful information about the action of the heart. The most important parameters for the cardiologists are the durations and the amplitudes of the above mentioned sub patterns.

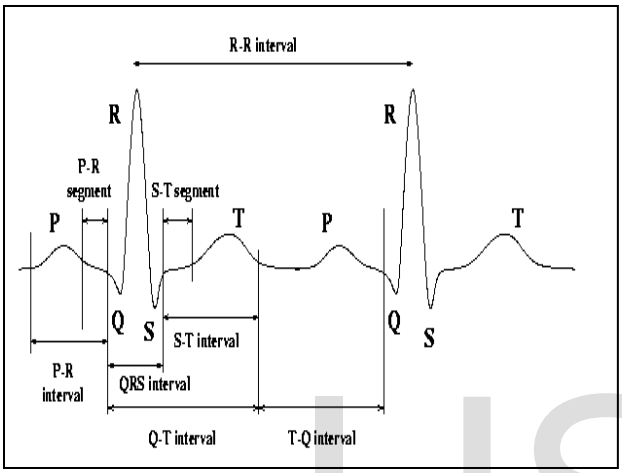

Figure 1: ECG Wave

1.2 ECG DATABASE

In this analysis number of ECG samples taken in a ECG file is 3863 and the sampling frequency of 450Hz. In this analysis 12 leads ECG signals of MIT/BIH database is used.

2 PRINCIPAL COMPONENT ANALYSES

The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first *few* retain most of the variation present in *all* of the original variables.

2.1 GOALS OF PCA

The goals of PCA are to

1) Extract the most important information from the data table.
2) Compress the size of the data set by keeping only this important information,
3) Simplify the description of the data set
4) Analyze the structure of the observations and the variables.

In order to achieve these goals, PCA computes new variables called principal components which are obtained as linear combinations of the original variables. The first principal component is required to have the largest possible variance. The second component is computed under the constraint of being orthogonal to the first component and to have the largest possible inertia. The other components are computed likewise. The values of these new variables for the observations are called factor scores, these factors scores can be interpreted geometrically as the projections of the observations onto the principal components.

2.2 ECG LEAD SELECTION BASED ON PCA

In the first stage of analysis in this paper, optimal leads are selected from the 12 leads ECG signals using PCA. In the 12 lead ECG signals the correlation exist between different ECG leads, that is same information is present in between. If the correlation coefficient between the ECG leads reduces then noise is better separated or reduced which results in more diagnostic information. On the other hand if correlation coefficient increases then noise is more predominant which results in lesser diagnostic information.
Values of correlation coefficients between various lead combinations give different values of correlation coefficients and also indicate the noise content sustain in the data as well the artifacts removed before and after application of PCA.

The following simplified steps are applied to the PCA based ECG analysis:

1) Get the 12 Leads ECG data
2) Subtract the mean from each ECG leads
3) Calculate the covariance matrix of ECG leads
4) Calculate eigenvectors and Eigen values of the covariance matrix.
5) Sort the Eigen values and Eigen vectors, on the basis of highest Eigen values
6) Select principal components and create ECG Feature Eigen vectors

$$\text{Feature Eigen Vector} = [eig_1; eig_2; \cdots eig_n] \quad (1)$$

7) Derive the new optimal leads ECG signals

$$\text{Optimal Leads} = [\text{Feature vector}]^{\text{T}} * [\text{ 12 leads ECG signal}] \quad (2)$$

The 12 lead ECG signals are pile up to a vector Y_i , contain all L leads of the i[th] beat.

$$Y_i = [x_{i,1}x_{i,2}x_{i,3}x_{i,4} \dots \dots \dots L] \quad (3)$$

Therefore by selecting all L leads for a *n* time, represented by the vector

$$x_i(n) = \left[x_{i,1}(n)x_{i,2}(n)x_{i,3}(n) \dots \dots \dots x_{i,L}(n)\right]^T \quad (4)$$

The PCA is used to concentrate the information into fewer leads

$$O_i(n) = PC^T x_i(n) \quad (5)$$

Where PC is the principal component

The original 12 Lead ECG signal and optimal lead ECG signal is shown in figure 2-3.

It is apparent that the energy of the original leads is redistributed so that only 3 out of the 12 transformed leads have important energy; the remaining leads generally account for noise even though small residues of ventricular activity can be identify.

Table 1 show that 12 Leads ECG signal Eigen value and % of total variance of 12 Leads ECG signals. The result show that leads V3 has the maximum Eigen value, afterwards lead L2 and AR. This means that in first principal components of Eigen vectors has the maximum information is present in the V3 lead and then second and third principal component of L2 and AR leads, remaining principal component whether noise or error is present in small amplitude. Figure 4 show the bar representation of principal component for the 12 leads ECG signal and % of total variance.

Table 1 Eigen Values and % Variances of Principal Components of 12 Leads ECG Signals

S. No	Eigen Value Number	Magnitude of Eigen Value	12 Leads ECG	Total Variance of 12 Leads ECG Signals (%)
1	Eigen value 1	0.2505	V3	71.8697
2	Eigen value 2	0.0649	L2	18.6075
3	Eigen value 3	0.0295	AR	8.4664
4	Eigen value 4	0.0018	AF	0.5294
5	Eigen value 5	0.0015	AL	0.4185
6	Eigen value 6	0.0003	L1	0.0751
7	Eigen value 7	0.0001	L3	0.0210
8	Eigen value 8	0.0000	V1	0.0124
9	Eigen value 9	0.0000	V2	0.0000
10	Eigen value 10	0.0000	V4	0.0000
11	Eigen value 11	0.0000	V5	0.0000
12	Eigen value 12	0.0000	V6	0.0000

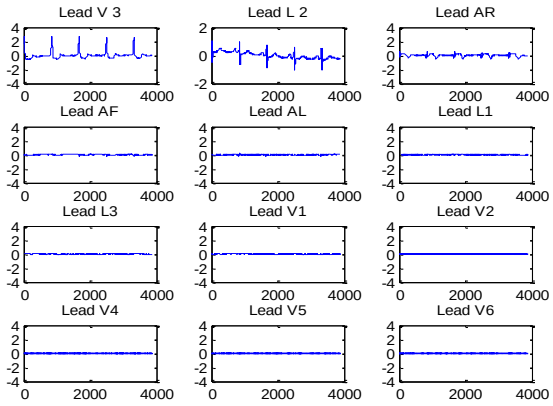

Figure 2: The standard 12-lead ECG Signals

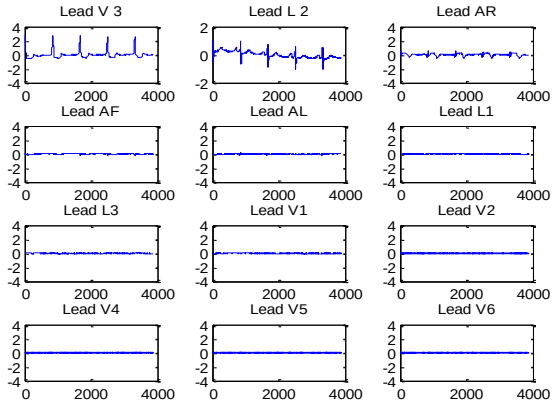

Figure 3: Optimal lead ECG Signals

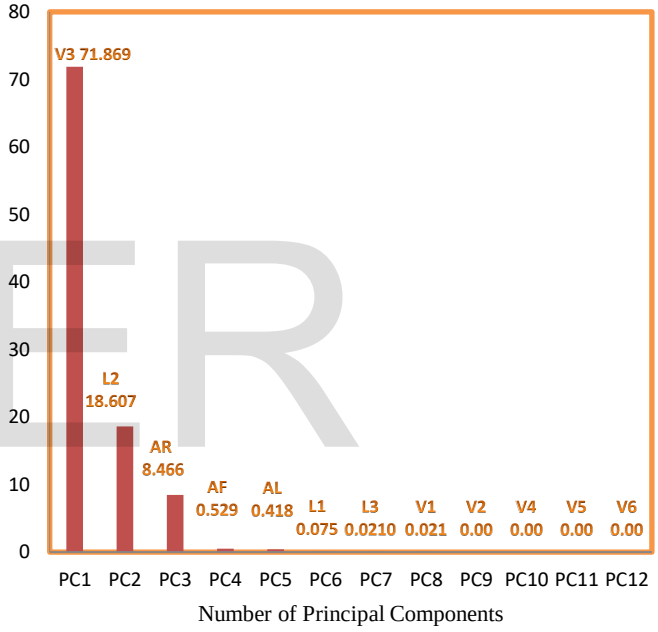

Figure 4: Bar Representation of Principal Component of 12 Leads ECG Signal

3 MORPHOLOGICAL FEATURE EXTRACTIONS

In this section of paper we apply the PCA for morphological feature extraction of optimal ECG lead which are obtained with the help of PCA in previous section of this paper. In the ECG signal morphological features are QRS complex, P wave T wave and RR interval etc. In the ECG signal some sort of noise and base line wander are also come into the picture at time of ECG recording, with the help of PCA we remove these noise or base line wander from ECG signal. PCA is applied only optimal leads like V3 and L2 leads these are obtained with the help of PCA.

In the PCA based features extraction process, first ECG signal is segmented into a series of consecutive slab where each slab is subjected to one beat (i.e. one heart beat or one complete cycle of ECG signal). Every slab is situated around QRS complex, opening at a permanent distance before the QRS including the P wave

IJSER

and extending ahead of the end of the T wave to the starting of the next beat.

It have to be prominent that partition of the ECG is bound to fail if certain chaotic arrhythmias are encountered such as ventricular fibrillation throughout which no QRS complexes are identified.

The following step is performed for morphological feature extractions from optimal leads are as

(1) Collect the n dimensional ECG beat (i.e. one heart beat contain one data set)

(2) Calculate the mean and subtract it from each heat beat.

(3) Calculate the covariance matrix of the ECG heart beat.

$$\sum = XX^T \quad (6)$$

(4) Apply the PCA to decompose the covariance matrix.

$$\sum = \sum_{i=1}^{N} \lambda_i PC_i PC_i^T \quad (7)$$

Where λ_i is the eigen values and PC_i is the principal component of beats

(5) Calculate the Eigen values and Eigen vector of the covariance matrix.

(6) Arrange the Eigen value and the corresponding eigenvector, on the basis of decreasing value of Eigen values/Eigen vector.

(7) Select principal components and create ECG Feature Eigen vectors

$$\text{Feature Eigen Vector} = [eig_1; eig_2; \cdots eig_n] \quad (8)$$

(8) Principal components are rotated using varimax rotation.

(9) Varimax rotation is an orthogonal transform that rotates the principal components such that the variance of the factor is maximized.

(10) Analyze each principal component.

The first principal component looks somewhat like part of the central QRS complex in a normal ECG likewise. Second and third principal component resembles element of T wave and P wave. Fourth and fifth principal component look are noise and base line wander.

Table 2 and 3 show that eigen values magnitude, ECG segment and % of total variance in ECG segments for optimal leads V3 and L2 respectively.

Figure 5-6 show the bar representation of principal components for leads V3 and L2. Figure 7 shows the detected QRS complex for lead V3 as first principal component.

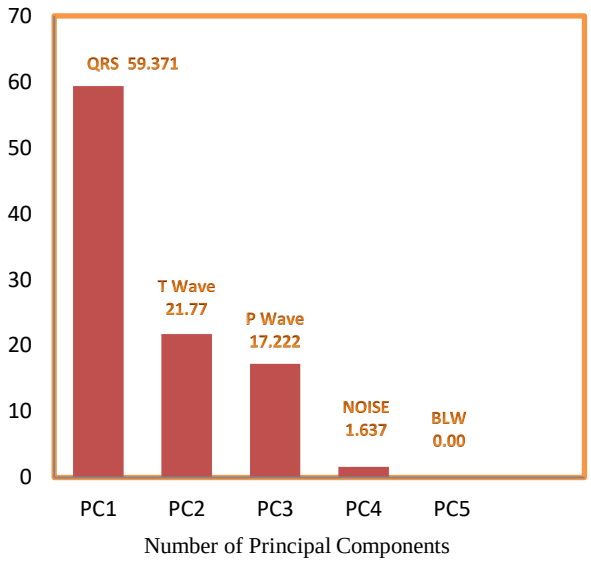

Figure 5: Bar Representation of Principal Component of V3 Lead

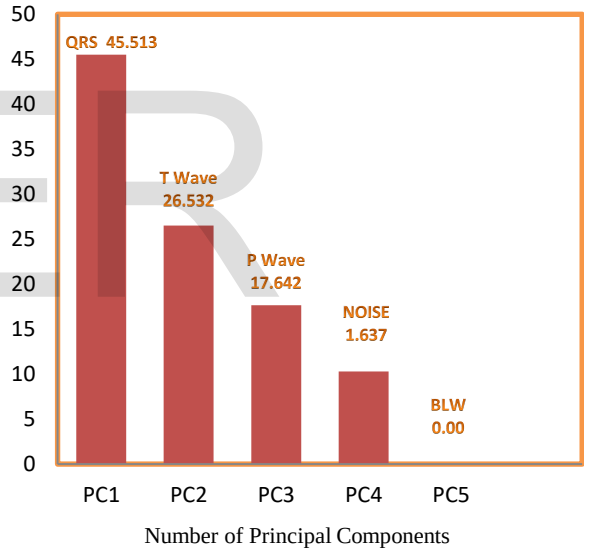

Figure 6: Bar Representation of Principal Component of L2 Lead

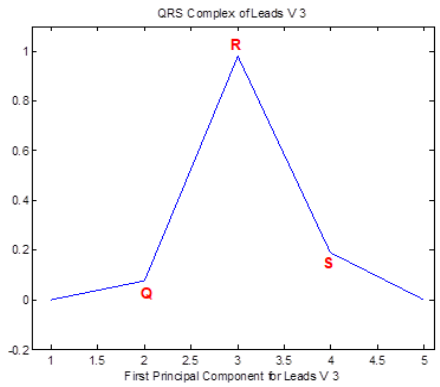

Figure 7: Detected QRS Complex of Lead V3

IJSER

Table 2 Eigen Values and % Variances of Principal Components of Lead V3

S. No.	Eigen Value Number	Magnitude of Eigen Value	ECG Segments	Total Variance of ECG Segments (%)
1	Eigen value 1	0.6841	QRS Complex	59.371
2	Eigen value 2	0.2508	T wave	21.770
3	Eigen value 3	0.1984	P Wave	17.222
4	Eigen value 4	0.0189	Noise	1.637
5	Eigen value 5	0.0000	BLW	0.0000

Table 3 Eigen Values and % Variances of Principal Components of Lead L2

S. No.	Eigen Value Number	Magnitude of Eigen Value	ECG Segments	Total variance of ECG Segments (%)
1	Eigen value 1	0.0662	QRS Complex	45.513
2	Eigen value 2	0.0386	T wave	26.532
3	Eigen value 3	0.0256	P Wave	17.642
4	Eigen value 4	0.0150	BLW	10.313
5	Eigen value 5	0.0000	Noise	0.0000

4 CONCLUSION

In this paper we have presented the PCA approach to analysis ECG signal. A linear PCA is introduced to select the optimal leads from the 12 Leads ECG signals and then extract the morphological features from the optimal leads.
The features are QRS Complex, T wave and P Wave. The test results show that PCA is better and faster than other techniques.

IJSER

REFERENCES

[1] T. Joliffe, *Principal Component Analysis*, Springer, New York, NY, USA, 2002.

[2] P. M. Okin, R. B. Devereux, R. R. Fabsitz, E. T. Lee, J. M. Galloway,and B. V. Howard, "Principal component analysis of the T wave and prediction of cardiovascular mortality in American Indians: the strong heart study," *Circulation*, vol. 105,no. 6, pp. 714–719, 2002.

[3] L. De Lathauwer, B. De Moor, and J. Vandewalle, "*SVD-based methodologies for fetal electrocardiogram extraction*," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '00)*, vol. 6, pp. 3771–3774, Istanbul, Turkey, June 2000.

[4] S. G. Priori, D. W. Mortara, C. Napolitano, et al., "*Evaluation of the spatial aspects of T-wave complexity* in *the long-QT syndrome*," Circulation, vol. 96, no. 9, pp. 3006–3012, 1997.

[5] L. S¨ornmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, Elsevier/Academic Press, Amsterdam, The Netherlands, 2005.

[6] Draper, B., Baek, K., Bartlett, M.S. and Beveridge, J.R., 2003, *Recognizing faces with PCA and ICA*, Computer Vision and Image Understanding, 91(1–2), 115–137.

[7] F. Kornreich, P. M. Rautaharju, J. W. Warren, B. M. Horacek, and M. Dramaix, "*Effective extraction of diagnostic ECG waveform information using orthonormal basis functions derived from body surface potential maps*," Journal of Electrocardiology, vol. 18, no. 4, pp. 341–350, 1985.

[8] M. Zabel, M. Malik, K. Hnatkova, et al., "Analysis of T-wave morphology from the 12-lead electrocardiogram for prediction of long-term prognosis in male US veterans," *Circulation*, vol. 105, no. 9, pp. 1066–1070, 2002.

[9] J. Garc´ıa, M. A° stro¨m, J. Mendive, P. Laguna, and L. So¨rnmo,"*ECG-based detection of body position changes in ischemia monitoring*," IEEE Transactions on Biomedical Engineering, vol. 50, no. 6, pp. 677–685, 2003.