International Journal of Engineering Research and General Science Volume 2, Issue 3, April-May 2014
ISSN 2091-2730

Morphological & Dynamic Feature Based Heartbeat Classification

N P. Joshi[\1],Shweta A.Tambe[1], P.S. Topannavar[1]

[1]Scholar, ICOER, Pune, India

ABSTRACT - In this paper, a new approach for heartbeat classification is proposed. The system uses the combination of morphological and dynamic features of ECG signal. Morphological features extracted using Wavelet transform and independent component analysis (ICA). Each heartbeat undergoes both the techniques separately. The dynamic features extracted are RR interval features. Support vector machine is used as a classifier, after concatenating the results of both the feature extraction techniques, to classify the heartbeat signals into 16 classes.
Whole process is applied to both the lead signals and then the classifier results are fused to make final decision about the classification. The overall accuracy in classifying the signals from MIT-BIH arrhythmia database should be 99% in “class-oriented” evaluation and an accuracy more than 86% in the “subject-oriented” evaluation.

Keywords—heartbeat classification, support vector machine, independent component analysis, wavelet transform, RR features, ECG signal, evaluation schemes.

I. INTRODUCTION

Electrocardiogram (ECG) analysis is basically used to control cardiac disorders. Cardiac disorders are the conditions such as abnormal behaviour of heart. if the medical emergencies are not provided properly then the subject can cause impulsive death.
There is another class of arrhythmias which is not critical for life but still should be given attention and treated. Classification of heartbeats depending on classes of heartbeats is an important step towards treatment. Classes are based on consecutive heartbeat signal [1]. To satisfy the requirements of real-time diagnosis, online monitoring of cardiac activity is preferred on human monitoring & interpretation. Also automatic ECG analysis is preferred for online monitoring & detection of abnormal activity observed in heart. Hence, automatic heartbeat classification using parameters or characteristic features of ECG signals is discussed in this paper.

II. DATASET

A .Classes of ECG signal

The MIT-BIH arrhythmia database [2] is the standard material used for training & testing of the algorithm developed for detection & classification of arrhythmia ECG signals. By using this database we cancompare the proposed method with the approaches in published results. MIT-BIH arrhythmia database is exploited for testing the system.
There are total 48 records. All signals are two lead signals denoted as lead A & lead B signal. These signals are filtered using BPF at 0.1 Hz - 100Hz. Sampling of thesesignals is performed at 360 Hz.

TABLE I
CLASSES OF ECG SIGNAL

Heartbeat type	Annotation
Normal Beat	N
Left Bundle Branch Block	L
Right Bundle Branch Block	R
Atrial Premature Contraction	A
Premature Ventricular Contraction	V
Paced Beat	P
Aberrated Atrial Premature Beat	A
Ventricular Flutter Wave	!
Fusion of Ventricular and Normal Beat	F
Blocked Atrial Premature Beat	X
Nodal (Junctional) Escape Beat	J
Fusion of Paced and Normal Beat	F
Ventricular Escape Beat	E
Nodal (Junctional) Premature Beat	J
Atrial Escape Beat	E
Unclassifiable Beat	Q
TOTAL:	**16**

All 48 records belong to one of the classes of ECG signal shown in TABLE I. According to the clinical terms, V1 to V6 leads represent the area of heart. In 45 records, Lead A signal is a modified limb Lead II while Lead B is a modified Lead V1. In remaining 3 records, Lead A is from position V5 and Lead B signal is V2.

B. Evaluation schemes

Previous literature [3]-[9] is divided into two categories according to the evaluation scheme followed. Following evaluation schemes are used:

1) Class-oriented evaluation.
2) Subject-oriented evaluation.

All 48 records contained in MIT-BIH arrhythmia database are not used. 4 ECG signals are excluded as they are paced beats. Each ECG signal contains its own annotation file in database. Those annotations of QRS complex are used for segmentation of ECG signals from which heartbeat segments can be obtained.44 ECG signals are divided into 2 datasets. One of them is used as a training dataset and another one is used as testing dataset. This division is done for the experiment purpose.Above datasets are prepared by selecting a random amount of fraction from each of the 16 classes. Training dataset constitutes following fractions of beats. Normal class contributes 13% of the beats, 40% contribution is from each of the five bigger classes i.e. 'L','A','R','V' & 'P' while 50% contribution is of all the small ten classes. Mapping of these 16 classes is done in 5 classes as shown in TABLE II.

TABLE II
MAPPING OF MIT-BIH CLASSES TO AAMI CLASSES

AAMI Classes	MIT-BIH Classes
N	NOR, LBBB, RBBB, AE, NE
S	APC, AP, BAP, NP
V	PVC, VE, VF
F	VFN
Q	FPN,UN

III. PROPOSED METHODOLOGY

Section I contains a brief introduction of the proposed automatic heartbeat classification system. In this section, we have discussed all the theoretical details and the techniques used in the process. Fig. 2 shows the flow of the proposed system. The process has following blocks Pre-processing, Heartbeat segmentation, Feature extraction, Classification, Two-lead fusion and Decision. Lead I & Lead II signals are nothing but raw ECG signals. Artefacts contained in these raw ECG signals are removed by using the first block of the process i.e. Pre-processing. After pre-processing, these ECG signals are divided to obtain heartbeat segments. For this purpose, we use provided R peak locations.

We apply Wavelet transform (WT) and independent component analysis (ICA) separately to each heartbeat and concatenate corresponding coefficients. Now we use Principal component analysis (PCA) and represent these coefficients in a lower dimensional space. Now the resulting principal components that represent most of the variance are selected and a morphological descriptor of the heartbeat is obtained by utilizing these components. RR interval features are derived, which give descriptive information about dynamic features of the heartbeat.

After performing the feature extraction, the main classification algorithm is applied. Heartbeats are then classified into 16 above classes by using a classifier based on Support vector machine (SVM) is used. According to the data given in [2] all the ECG signals are two-lead signals, all the above process is separately applied to the signals from leads A & B. The two sovereign decisions for each heartbeat are obtained, which then are fused to build the final composed decision of heartbeat classification. By integrating both leads signals, confidence about classification can be improved for the final decision.

A. Pre-processing

It is necessary to perform the pre-processing of raw ECG signals as they can contain various types of noise. These noises must be reduced so that signal-to-noise ratio (SNR) is improved. Improved SNR helps in detection of the subsequent fiducial point. Types of noise like power-line interference, baseline wander, artifacts due to muscle contraction, and electrode movement affect the quality of ECG signals. In this study, the pre-processing of ECG signals consists of baseline wander correction. The baseline wander is removed by subtracting mean of the signal from signal itself. The pre-processed signals were used in subsequent processing.

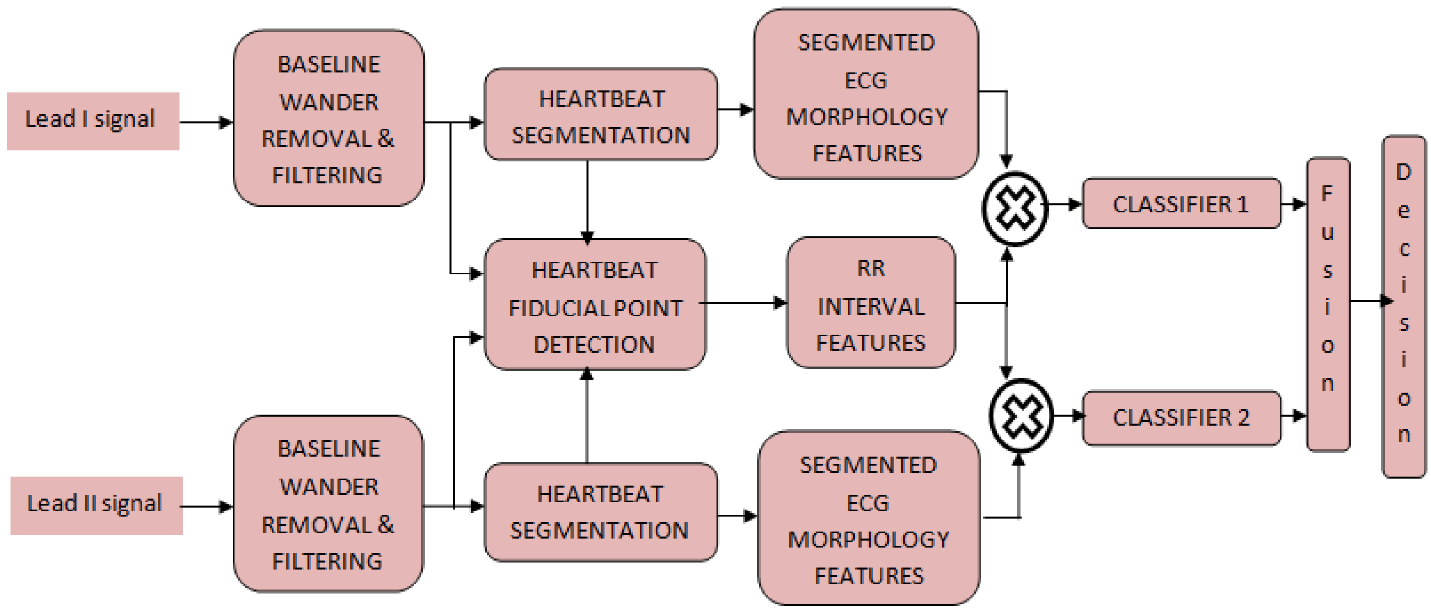

Fig. 1 Flow of proposed system

B. Heartbeat segmentation

There are three waveform components of one heartbeat of ECG signal known as P wave, QRS complex and T wave. To have the full segmentation of ECG signal, the boundaries and peak locations i.e. fiducial points should be properly detected. To obtain heartbeat segments or R-peaks the annotations provided for R-peak locations are utilized. In real applications, an automatic R-peak detector may be used so that the classification method can actually be fully automatic. But there are two disadvantages of this automatic R-peak detector. 1. If some leading heartbeats are missed then error may get added in those heartbeat signals and hence they cannot be classified correctly.

A number of heartbeat detection schemes do exist [7], [12], [13], which have capacity to detect heartbeat signals present in MIT-BIH arrhythmia database with an error rate less than 0.5%. 2. The quality of RR interval features will be degraded to some extent because of addition of the error by the automatic R peak detector. The sampling rate is given as 360 Hz. Hence in each heartbeat segment there are 100 samples before the R peak location as the pre-R segment & 200 samples after the R peak as the pro-R segment, i.e., a total of 300 samples. The segment size is selected such that it includes most of the information of one heart cycle. The segment size of heartbeat is kept fixed. The ratio of lengths of the pre-R segment and the pro-R segment is kept so that it matches with lengths of PR interval & the QT interval. There is an advantage to keep the fixed segment size it avoids the detection of the P wave and T wave.

C. Wavelet transform- Morphological Feature extraction

ECG signals i.e. biomedical signals in real exhibit non-stationary nature. Non-stationary nature actually means the presence of some statistical characteristics. These signals change over position or time. Due to this nature, they cannot be responsive and hence cannot be analysed using classical Fourier transform (FT). Therefore, it becomes must to use wavelet transform (WT). Wavelet transform is capable of performing analysis in both the domains i.e. time & frequency domains. It is possible to analyse ECG signal by using WT.

There are various purposes of using WT in ECG signal processing. It includes de-noising, heartbeat detection and feature extraction. We use WT as a feature extraction method in this study. As can be seen, Daubechies wavelets of order 8 have most similar characteristics as that of QRS complex, hence are selected. Since the sampling frequency is given to be 360 Hz, the maximum frequency is 180 Hz.

C. Independent component analysis- Morphological Feature extraction

In this study, ICA is used for feature extraction[15].Five sample beats are randomly selected from every class for preparation of training set. These training sets are used to compute Independent components. If the total number of beats in any of the recording is less than five, then all beats are taken. This makes a training set of total 626 beats taken from all 16 classes. These beats are used for calculating ICs. The ICs obtained are used as source signals for ICA and hence applied to both the datasets viz. training and testing datasets. To obtain actual number of ICs, tenfold cross validation is evaluated. Number of independent components are varied between 10 & 30, the ICA coefficients obtained after that are actually considered as features and given as input to SVM classifier. This process is performed in 5 iterations. And average is taken. When average performance is observed, the accuracy increases at number of ICs between 10 & 14 and afterwards it decreases. So number of ICs is selected to be 14.

D. Principal component analysis-Morphological feature extraction

The two features obtained i.e. ICA features and wavelet features are combined together and PCA is applied to obtain the reduction in feature dimension. Then 10-fold cross validation is performed and final morphological features are obtained.

E. RR Interval Features

RR interval features are extracted to obtain dynamic information of the heartbeat signal input. These are known as “dynamic” features. There are four RR interval features namely, previous RR, post RR, local RR, and average RR interval features.
The previous RR feature is nothing but the interval between a present R peak and the previous R peak. Post RR feature is calculated as the interval between current R peak and next R peak. The local RR interval is calculated by taking average of all the RR intervals within past 10-s period of the given heartbeat. Likewise, the average RR interval is calculated as the average of RR intervals within past 5-min period of the heartbeat.

In previous literature, the local RR & average RR feature extraction shows poor performance when applied in real-time application. The local RR feature is calculated as average of consecutive 10 heartbeats whose centre will be at given beat. Whereas average RR feature is calculated as average of all beats from same recording. In the proposed method, these features are calculated such that they ensure to work at real-time.

F. Support vector machine

Support vector machines are nothing but binary classifiers. This classifier is given by Vapnik. It builds a prime hyperplane which separates two classes from each other due to increase in margin between them. As this approach has an excellent ability to build the classification model on general basis it is enough powerful to be used in many applications.A number of multiclass classification strategies have been developed to extend SVM to address multiclass classification problem [14], such as heartbeat classification problem.In this paper, the technique used for classification of the heartbeats is an SVM classifier which classifies the heartbeat under consideration into one of the 16 classes.

The training set contains N examples. It is used in two-class classification problem. N examples are given as $\{(\mathbf{x}_i, y_i), i = 1, . . . ,N\}$, where $\mathbf{x}_i$ is nothing but d-dimensional feature vector of the ith example and $x_i \in d$ whereas y_i is the class label of ith example and $y_i \in \{\pm1\}$. Now a decision function is to be constructed on the basis of the training set. This function is used to predict output class labels of test examples. These are based on input feature vector. The resultant decisionfunction is given as

$$f(x) = \text{sign}\left(\sum_{i \in SVs} \alpha_i y_i K(x_i, x) + b \right)$$

where, $K(., .)$ is kernel function. $\alpha_i$is Lagrangemultiplier for each training datasample. Few Lagrange multipliers are nonzero. The examples of training set which are nonzero are known as Support vectors. These support vectors actually determine f(x). Two seperate classifiers are applied to signals from lead A & lead B.

G. Two-lead fusion

As two different classifiers are applied, each classifier gives its seperate answer. Now the two answers are fused together to get a final answer which actually gives the class of the heartbeat it belong to. Two seperate answers can be fused together by using rejection approach.

IV. RESULTS& ANALYSIS

As seen in Fig.2 & Fig.5, the original lead 1 signal is shifted from its axis and offset is added in it. This happens because of patient’s movement or in-line interference. The pre-processing of signal results into reduction of these noises. The shift of axis is called as baseline wander. The baseline wander is removed after pre-processing. Pre-processing also helps in R-peak detection. The amplitude of the signal is compared with the threshold and hence the R peaks are found out from the pre-processed signal. These R peaks give post-R and pre-R features. While average of these R-peaks are found to get average-R feature. Segmentation gives the actual separation of the QRS complex from whole recording. Sampling frequency is kept to be 180 Hz. We take 100 samples before and 200 samples after to get proper segment, which actually contains total QRS complex, P wave and T wave. This helps in finding out the exact class of the ECG signal after applying feature extraction techniques and classifier. Hence segmentation size is kept fixed.

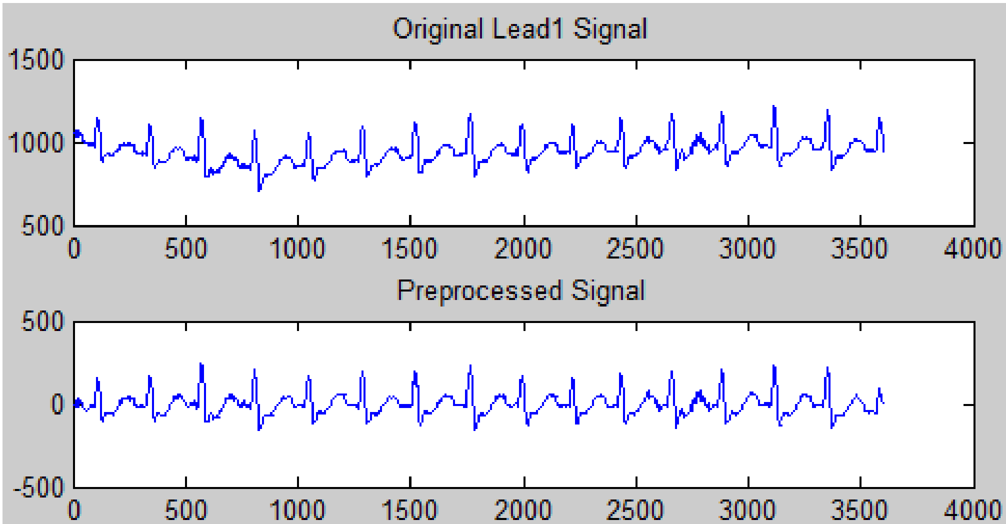

Fig. 2 Results showing pre-processing for lead 1 signal of 109 ECG recording.

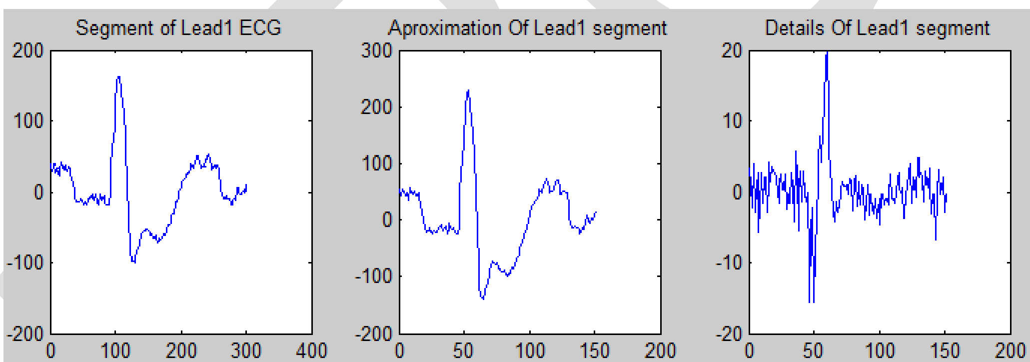

Fig. 3 Results showing segmentation of lead 1 signal of 109 ECG recording

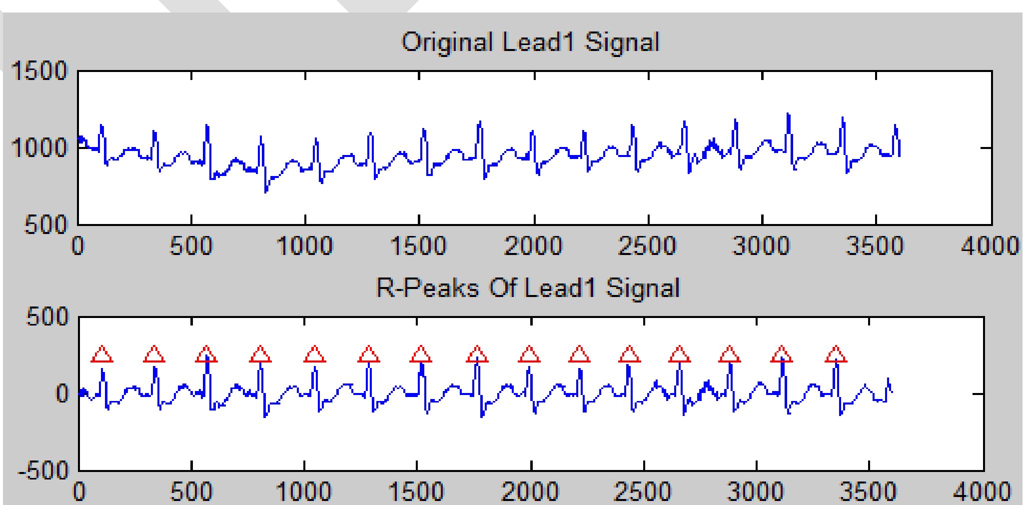

Fig.4 Results showing R-peak detection of lead 1 signal of 109 ECG recording

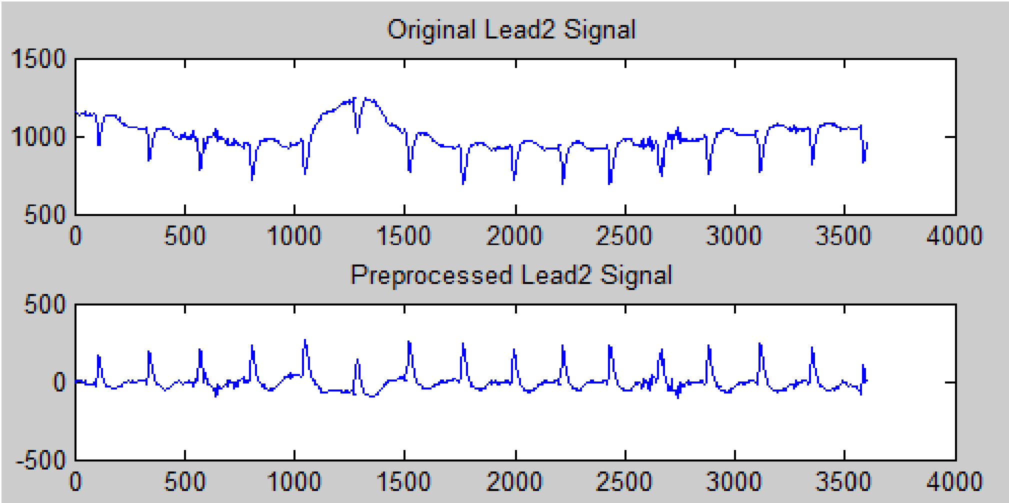

Fig.5 Results showing pre-processing of lead 2 signal of 109 ECG recording

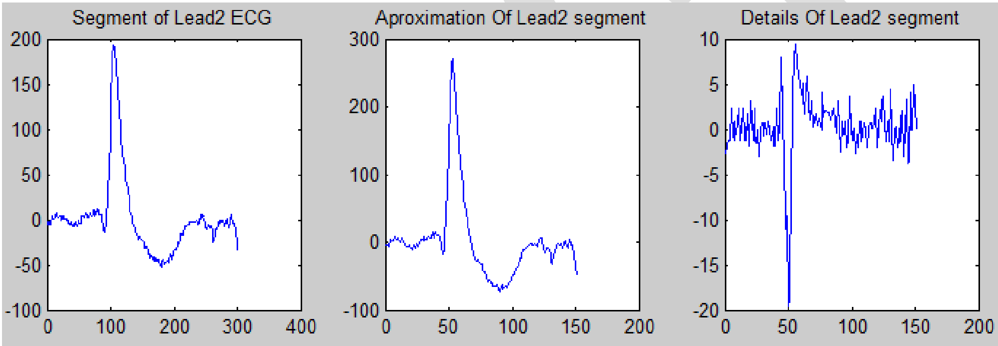

Fig. 6 Results showing segmentation of lead 2 signal of 109 ECG recording

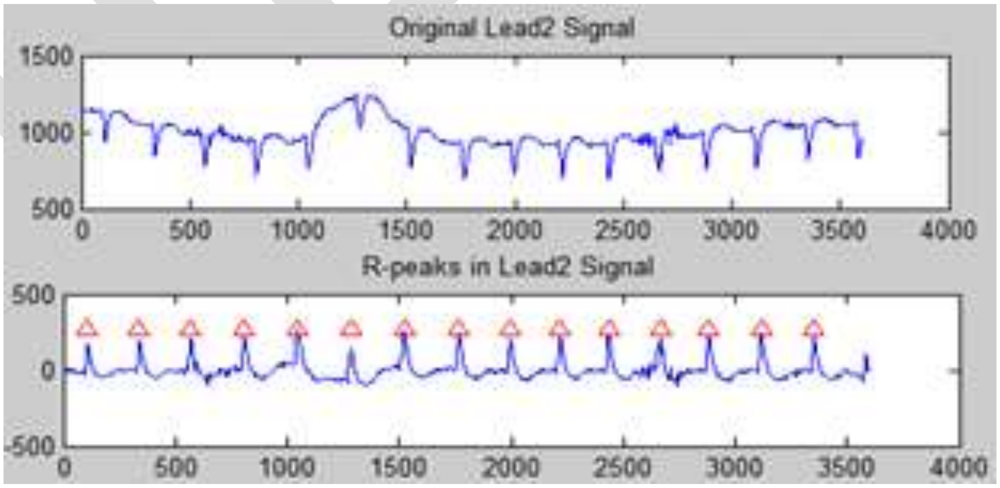

Fig.7 Results showing R-peak detection of lead 2 signal of 109 ECG recording

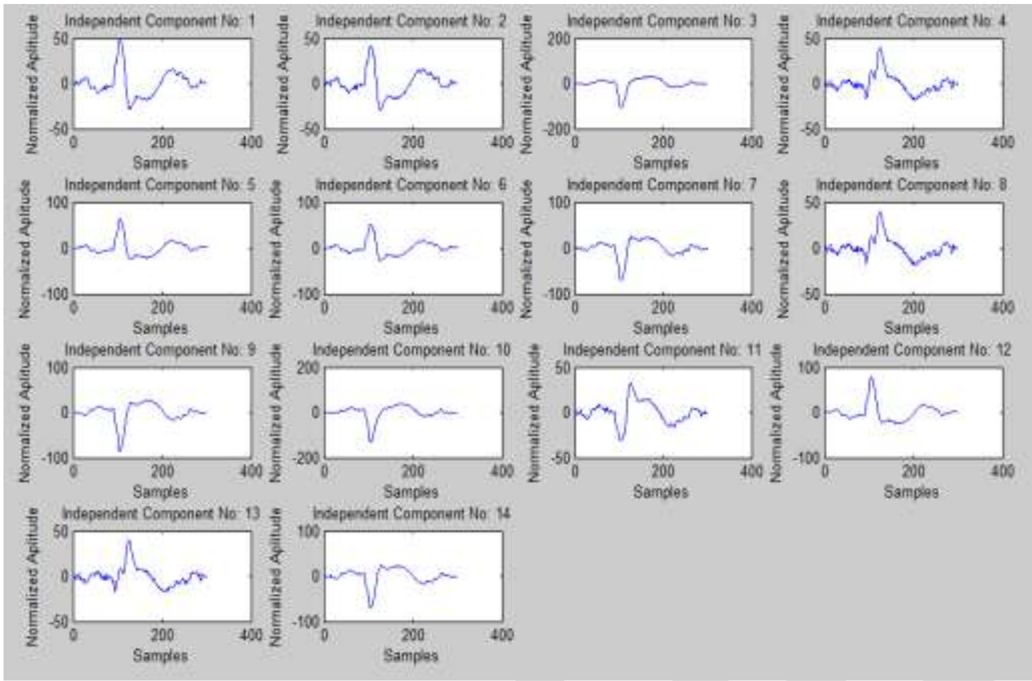

Fig.8 Results showing 14 Independent components of 109 ECG recording

REFERENCES:

[1] Can Ye, B. V. K. Vijaya Kumar, Miguel Tavares Coimbra, "Heartbeat Classification Using Morphological and Dynamic Featuresof ECG Signals," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 10, pp. 2930-2941, Oct. 2012.

[2] MIT-BIH Arrhythmias Database. Available online at: http://www.physionet.org/physiobank/database/mitdb/.

[3] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo,"Clustering ECG complexes using Hermite functions and self-organizing maps," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 7, pp. 838–848, Jul. 2000.(6)

[4] P. de Chazal, M. O. Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features,"*IEEETrans. Biomed. Eng.*, vol. 51, no. 7, pp. 1196–1206, Jul. 2004. (8)

[5] S. Osowski, L. T. Hoa, and T. Markiewic, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Trans. Biomed.Eng.*, vol. 51, no. 4, pp. 582–589, Apr. 2004. (9)

[6] J. Rodriguez, A. Goni, and A. Illarramendi, "Real-time classification of ECGs on a PDA," *IEEE Trans. Info. Tech. Biomed.*, vol. 9, no. 1, pp. 23–34, Mar. 2005. (10)

[7] P. Laguna, R. Jane, and P. Caminal, "Automatic detection of wave boundaries in multilead ECG signals: Validation with the CSE database," *Comput.Biomed. Res.*,vol. 27, no. 1, 1994. (12)

[8]W. Jiang and G. S. Kong, "Block-based neural networks for personalized ecg signal classification," *IEEE Trans. Neural Networks*, vol. 18, no. 6,pp. 1750–1761, Nov. 2007. (13)

[9] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ECG signals," *IEEE Trans.Biomed. Eng.*, vol. 56, no. 5, pp. 1415–1426, May 2009. (14)

[10] M. Llamedo and J. P. Martinez, "Heartbeat classification using feature selection driven by database generalization criteria," *IEEE Trans. Biomed.Eng.*, vol. 58, no. 3, pp. 616–625, Mar. 2011. (15)

[11] G. de Lannoy, D. Francois, J. Delbeke, andM. Verleysen, "Weighted conditional random fields for supervised interpatient heartbeat classification," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 1, pp. 241–247, Jan. 2012. (17)

[12] V. X. Afonso, W. J. Tompkins, T. Q. Nguyen, and L. Shen, "ECG beat detection using filter banks," *IEEE Trans. Biomed. Eng.*, vol. 46, no. 2, pp. 192–202, Feb. 1999. (25)

[13] S. Kadambe, R. Murray, and G. F. Boudreaux, "Wavelet transform-based QRScomplex detector," *IEEE Trans.Biomed. Eng.*, vol. 46, no. 7, pp. 838– 848, Jul. 1999. (27)

[14] C. Cortes and V. N. Vapnik, "Support vector networks," *J. Mach. Learn.*,vol. 20, pp. 1–25, 1995. (34)

[15] Jarno M.A. Tanskanen, Jari J. Viik, *Independent Component Analysis in ECG Signal Processing*, Finland, Tampere University of Technology and Institute of Biosciences and Medical Technology.