

DISEASE DIAGNOSIS FROM ECG SIGNALS BASED ON OPTIMIZING INDEPENDENT COMPONENT ANALYSIS USING GENETIC ALGORITHM

M.RAMKUMAR

Sri Krishna College Of Engineering And Technology, Coimbatore-641008,, Tamil Nadu, India

Email: ramkumar.mkrishna@gmail.com

Dr.C.GANESH BABU

Bannari Amman Institute Of Technology, Sathyamangalam-638401, Tamil Nadu, India

Email: bits_babu@yahoo.co.in

Abstract: The examination of the ECG can benefit in diagnosing the greater part of the heart illness. The electrocardiogram (ECG) gives all data about electrical action of the heart. Changes in the typical beat of a human heart may bring about various cardiovascular arrhythmias, which might be quickly deadly or make hopeless harm to heart managed over drawn out stretches of time. The capacity to automatically recognize arrhythmias, for example, diabetics and blood pressure from ECG chronicles is critical for clinical analysis and treatment. The fundamental goal is to think of a straightforward strategy having less computational time without bargaining with the proficiency. This paper proposes an improved strategy for the arrhythmia classification and extraction of parameters from the ECG signal which is utilized for information gathering and classification framework. Principal component analysis (PCA) is utilized to diminish dimensionality of electrocardiogram (ECG) information proceeding for performing Independent component analysis (ICA). A recently proposed PCA change estimator by the author has been connected for distinguishing true, actual and false peaks of ECG information files. In this paper, it is felt that the capacity of ICA is also checked for parameterization of ECG signals, which is essential on occasion. Independent components (ICs) of appropriately parameterized ECG signals are more promptly interpretable than the estimations themselves, or their ICs. The original ECG recordings and the samples are organized by statistical measures to evaluate the noise statistics of ECG signals and discover the recreation errors. The capacity of ICA is clarified by finding the true, false and actual peaks of private hospital database ECG files. The preprocessed dataset is then classified utilizing machine learning algorithm named Genetic algorithm. The GA is qualified in improving the weights of the ICA, The genetic algorithm is used as a co-training algorithm for enhancing the connection weights values and minimizing the error value to least possible value. This method has demonstrated a decent outcome and a decent execution. Utilizing AI in examining the ECG signal has sparing time, quicker, and straightforward in diagnosing the illness.

Keywords: Electrocardiogram, PCA variance estimator, Principal component analysis, Independent component analysis, genetic algorithm

I. Introduction

There are different sorts of arrhythmias and each sort is related with an example, and in that capacity, it is conceivable to distinguish and arrange its sort. The arrhythmias can be characterized into two noteworthy classifications. The main class comprises of arrhythmias framed by a single irregular pulse, in this called morphological arrhythmia. The other classification comprises of arrhythmias shaped by an arrangement of unpredictable heartbeats, in this called rhythmic arrhythmias. The order of typical heartbeats and the ones making the previous gathering are on the concentration of this study. These heartbeats create modifications in the morphology or wave frequency, and these adjustments can be distinguished by the ECG exam. The way toward distinguishing and classifying arrhythmias can be exceptionally troublesome for an individual on the grounds that occasionally it is important to break down every pulse of the ECG records, obtained by a holter screen for example, during hours, or even days. Also, there is the likelihood of human blunder amid the ECG records investigation, because of exhaustion. An option is to utilize computational procedures for automatic classification. A full programmed framework for arrhythmia characterization from signals procured by an ECG gadget can be partitioned in four stages (see Fig. 1), as takes after: (1) ECG signal preprocessing; (2) heartbeat segmentation; (3) feature extraction; and (4) learning/classification. In each of the four stages, a move is made and the last target is the discrimination/identification of the type of heartbeat. The initial two stages of a such order framework (ECG signal preprocessing and heartbeat segmentation) have

been broadly investigated in the literature [2– 6]. The strategies utilized amid the preprocessing step straightforwardly impact the last outcomes, and in this manner, ought to be deliberately picked. The outcomes identified with the heartbeat segmentation venture, on account of QRS recognition, are near ideal. Be that as it may, there is still space for investigation and enhancements in the means identified with classification (feature extraction and learning algorithms). Despite the fact that the issue of ECG outline is as yet open, it isn't so valuable for the strategies in the literature overviewed here.

A review of existing investigations found in literature in regards to the ECG-based arrhythmia classification strategies and examines the primary procedures utilized for the development of these programmed frameworks and additionally two fundamental standards utilized for assessment: between inter-patient and intra-patient[7,8]. Furthermore, the most prevalent databases and the issues identified with the assessment of current strategies found in literature are additionally examined. From this discourse, a work process is proposed to direct the assessment procedure of future works. Note that this work process for assessment process constitutes an essential commitment of this study work. In the literature, we discover an overview of information based ECG translation [9] assessing techniques proposed in the twentieth century. Clifford et al.[1] played out a broad review on the strategies utilized for ECG signal investigation. Their investigation concentrated on the physiology of the signal, and it's preparing methods, chiefly on the feature extraction and classification. In this paper, we have proposed a Genetic based Independent Component Analysis for feature extraction and disease finding from ECG signals. The significant impediment in component analysis is we can't anticipate that the picked components have same number of vectors from each class; all things considered the general execution might be diminished. Henceforth, it is smarter to pick same number of components from each class. Genetic Algorithm is utilized for grouping, since it is a strong and solid classifier framework what's more, can perform quick classification. The calculation time of GA is lower than the other classifier calculations, for example, decision tree and an artificial immune system. As of recently, ECG accounts that are utilized for the finding of disease are influenced by noise, which crumbles fundamentally the analytic exactness. In this paper, the noises are expelled by the PCA strategy. Better handling of the noisy ECGs can enhance the precision

of the diagnostic techniques and increment their applications in consistently practice. The execution of the framework was analyzed with respect to the on precision to show our outcomes. Our proposed framework acquired 94% grouping exactness. This execution surpasses that of different examinations connected to the ECG dataset characterization issue up until this point.

The remaining of this paper is organized as follows. Section 2 introduces the fundamental aspects of ECG signals; Section 3 deals with the preprocessing technique PCA using in ECG signals, while Section 4 presents the concept of feature extraction using ICA from the ECG signals and its commonly employed techniques. Section 5 deals with the genetic algorithm for classification, Section 6 discuss the experimental results and its comparison. Section 7 presents the simulation results in Matlab for utilized databases, indicated by the standard, to evaluate the classification arrhythmia methods. Section 8 discusses the conclusion and point out future scope.

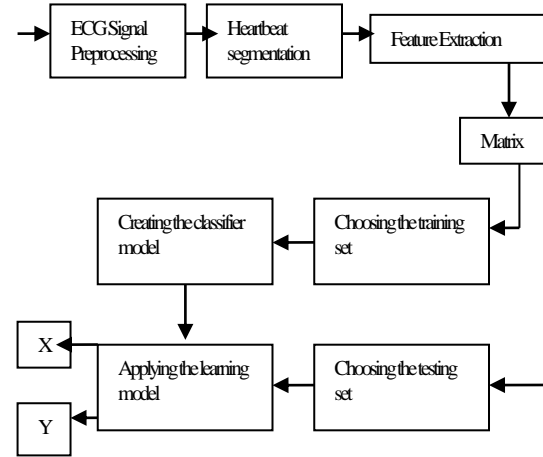


Fig. 1 Arrhythmia classification system.

II. ECG signal

The heart is a muscle that agreements in a rhythmical way, directing blood all through the body. This compression has its start at the atrial sine node that goes about as a characteristic pace-producer, and engenders through rest of the muscle. This electrical signal spread takes after an example [11]. Because of this movement, electrical currents are produced on the surface of the body, provoking variations in the electrical capability of the skin surface. These signals can be caught or measured with the guide of electrodes and proper device. The difference of electrical potential between the points marked by the electrodes on the skin usually is enhanced with the aid of an

instrumentation (operational) amplifier with optic isolation. Then, the signal is submitted to a high-pass filter; and as a second stage, submitted to an initializing low-pass filter. At last, it shows up in an analogical to digital converter. The graphical enrollment of this obtaining procedure is called electrocardiogram (ECG) (see Fig. 2). Since Augustus Desiré Waller showed the principal human ECG in 1887, the electrical action of the heart has been recorded [12]. Even in this way, the capacity to perceive the typical cardiovascular beat as well as arrhythmias did not end up plainly routine in therapeutic registration until the point that 1960.

Nowadays, there are many ways to deal with estimation/record ECG. da Silva et al. [13] gave a scientific categorization of best in class ECG estimation techniques: in-the-individual, on-the-individual and off-the-individual. Inside the in-the-individual classification, there are types of device intended to be utilized inside human body, for example, surgically embedded ones, subdermal applications or even ingested as pills. These gadgets are utilized when less invasive approach are not applicable. Appearing differently in relation to the in-the-individual classification, there is off-the-individual classification.

Gadgets on this classification are intended to gauge ECG without skin contact or with insignificant skin contact. As indicated by [13], this classification is lined up with future patterns of medical application where unavoidable PC frameworks are a reality. Cases of such types of equipment are the ones in view of capacitive device which measure the electric field changes induced by the body permitting ECG estimation at separation of 1 cm or more even with clothing between the body and the sensor [13–15]. The majority of devices utilized for ECG estimations are in the on-the-individual classification. Devices on this class regularly require the utilization of a few electrodes joined to the skin surface. Cases of such types of equipments are bedside monitors and holters. These days, the standard gadgets utilized for heart beat examination originate from this class.

Another generally utilized setup uses 10 electrodes [16], where 5 additional electrodes (other than V or V1 on the chest and LA, RD, LL and RA on legs and arms) are situated on the chest (V2 to V6) permitting an arrangement of 12 leads. The 10 electrodes (12 leads) arrangement can be found in Fig. 3. From these setups, a few unique leads can be developed to visualize the ECG signal.

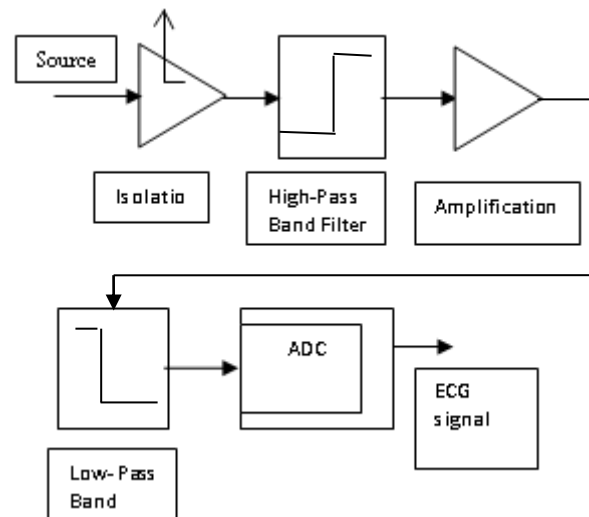


Fig. II ECG signals capture from the hardware.

On equipments having a place into the on-the-individual class, at least three electrodes are utilized to acquire the signal, in which one of them reference in as a source of perspective for the others. Normally, the reference electrode is set close to the right leg. All things considered, there can be distinctive visions of the ECG signal, depending upon the pair of electrodes built the signal. These separated visions are given the name of leads. A broadly utilized arrangement of electrodes is one made out of 5 electrodes [16]: one of the electrodes is situated on the left arm (LA), one on the right arm (RA), one on the left leg (LL), one on the right leg (RL) and one on the chest, to the right of the outer (V or V1).

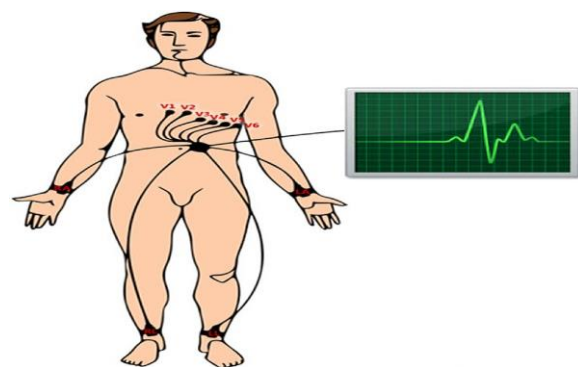


Fig. III Typical 10 electrodes configuration.

For instance, Fig. 4 delineates 3 specific leads: (I) shaped by the electrical potential distinction between the LA and RA electrodes; (II) framed by the electrical potential contrast between the LL and RA electrodes; and (III) framed by the electrical potential distinction between the LL and LA electrodes. The previously portrayed lead II is a standout amongst the most used for diagnosing heart illnesses.



Fig. IV Morphology of the curve for leads I, II and III

It features different portions inside the heartbeat, other than showing three of the most imperative waves: P, QRS and T (see Fig. 5). These waves relate to the field prompted by the electrical wonders happening on the heart surface, named atrial depolarization (P wave), ventral depolarization (QRS complex wave) and repolarization (T wave). The examples incited by arrhythmias can profoundly change these waves. In the interim, lead V and its associate leads (V1, V2) support the grouping of ventricular related arrhythmias, since there are electrodes situated on the chest, enhancing the registry of activity possibilities on ventricular muscle.

In this manner, the leads most used for the automatic heart-beat and arrhythmia classification are leads II and V and the techniques that utilization a mix of these two leads (and different blends) are the ones that present the best outcomes to date [17].

In this sense, the recent work by Tomasic and Trobec [18] audits techniques working with decreased quantities of leads and methodologies for the synthesis of leads, presuming that the customary 12 lead framework can be incorporated from fewer estimations [19]. Conversely, another research published by de Chazal [20] exhibited that comparative viability for ECG arrhythmia grouping can be acquired at a lesser computational cost when utilizing just a single lead, contrasted and strategies utilizing numerous leads [7]. Although on-the-individual is the standard on gadgets pointing heart sicknesses analyze, [13] have demonstrated that information caught with off-the-individual based gadgets can be exceedingly associated to information caught with conventional on-the-individual based equipments.

The authors guarantee that off-the-individual based equipments can broaden preventive solution by permitting ECG checking without impedance on day by day schedule. In that sense, we urge analysts to fabricate ECG databases in light of off-the-individual gadgets to assess and approve heartbeat classification techniques for that category.

A.PCA preprocessing:

Signal processing in general has massively

changed during the most recent 20 years and it is relied upon to change significantly more in the years to come.

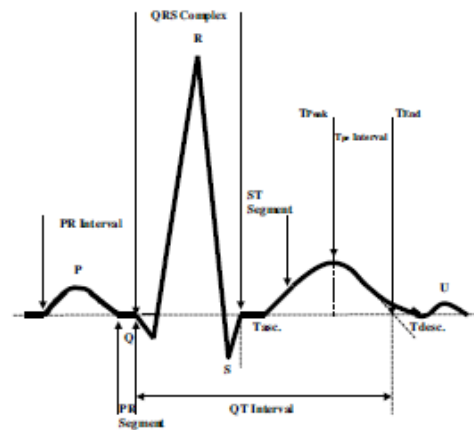


Fig. IV Fiducially points and various usual intervals (waves) of a heartbeat.

What was before pictured as digital signal processing now shapes just a little piece of the new idea of signal processing which may be all the more enough clarified as the techniques for examining, manipulating and passing on natural data. Feature extraction is fundamentally lessening of the accessible data keeping up ECG morphology. Features are representatives of distinguishing proof to a specific subject or example. Examination and feature extraction from electrocardiograms is troublesome until and unless artifacts and noise from the ECG are removed; there are numerous systems accessible in the literature [21 and 24]. Figure 2(a) demonstrates PCA based scheme for ECG data compression while figure 2(b) delineates the PCA plot for ECG processing.

The ECG features of interest are the different intervals and segments of ECG. A preprocessing step based on PCA is suggested, since after pre-whitening, the genuine ECG sources and the whitened vectors are quite recently related through an orthogonal change [21, 22, and 23]. For whatever length of time that the forced conditions are satisfied by the genuine ECG data, this strategy gives preferred execution over traditional ICA procedures in fact that the new data is incorporated into the proposed calculation. The proposed method is tested on real ECG database signals. At times in the beginning of the ECG processing, PCA is used for base line wander removal, which could be better choice. In the present work, PCA offers significant advantages in removing base-line wander (BLW) from ECG as compared to the aforementioned methods [22, 24, 25, and 26]. Figure

2(d) depicts the PCA transformation scheme for ECG classification and figure 2(e) illustrates a PCA based scheme for ECG data dimension reduction.

ECG signals analysis:

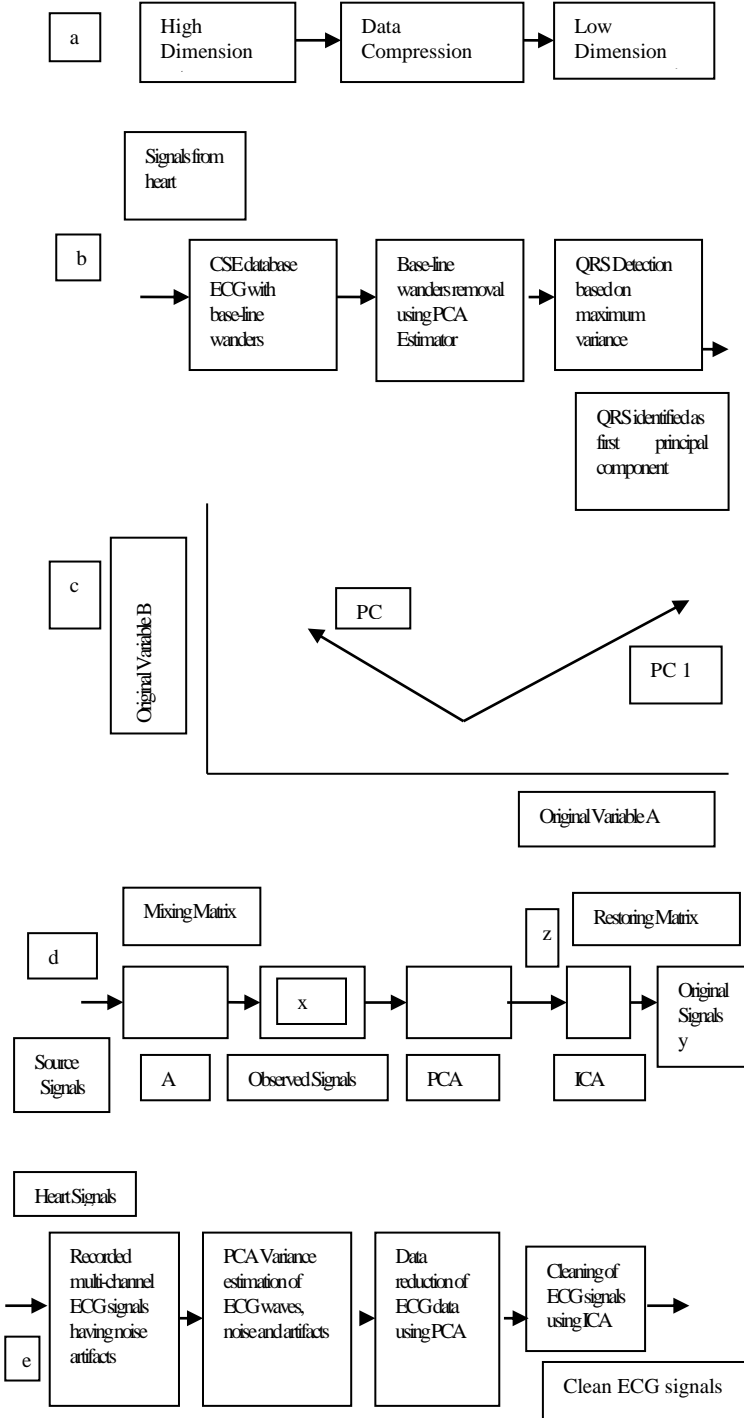


Fig 6: (a) PCA scheme for ECG data compression. (b) PCA scheme for ECG processing. (c) Basis for PCA vectors. (d) Combined PCA-ICA transformation scheme. (e) PCA based ICA scheme for dimension reduction and segment classification

In this paper, for classification purposes, a feature vector from ECG time tests is developed from every ECG lead of a data set. This vector is of distinct length and the R-wave peak is utilized as the reference purpose of this vector. The R-wave peak is the point where contrast between following slope and the previous slope in QRS region is a greatest. Some QRS identification strategies decreased the measure of data that medicinal specialists expected to process in time however needed to endure critical situations, losing diagnostic features. In the proposed strategy, first the peak of the QRS complex is distinguished with its high commanded amplitude in the signal utilizing PCA fluctuation estimator, trailed by recognition of Q and S-waves [21, 22 and 25].

B. Modeling steps of PCA

PCA is the ideal linear procedure which holds the most extreme measure of fluctuation (among every single linear projection) inside the anticipated feature space. The fundamental disadvantage of PCA lies in its worldwide linearity; since the calculation finds just a linear subspace of the original data space, it is imperfect when the hidden structure in the data is inherently nonlinear. PCA utilizes projections onto an 'orthogonal premise's set to isolate the ECG signal from the noise. The initial stage referred to as PCA has three fundamental purposes:

1. To estimate the number of ECG signals.
2. To remove the second order correlations between the temporal ECG waveforms.
3. To normalize the temporal ECG waveform vectors.

To estimate the number of ECG signals, the eigen values or singular values are compared against a detection threshold. The vectors relating to the noise subspace are then ignored. If the number of ECG temporal samples is large, it may be computationally more efficient to implement spatial eigen decomposition than SVD. The orthonormal temporal vectors can be constructed from the spatial eigenvectors.

The following simplified steps are required in modeling, the PCA based ECG analysis:

Step 1: Get the ECG data

Step 2: Subtract the mean

Step 3: Calculate the ECG covariance matrix

Step 4: Calculate eigenvectors and eigenvalues of the covariance matrix

Step 5: Choose principal components and form ECG feature vectors

Feature vector $v = (eig1 \dots eigp)$

Step 6: Deriving the new ECG data set

*Final data=Row feature vector*Row data adjust*

Step 7: Getting the old ECG data back, i.e. reconstruction of ECG data

Step 8: Reconstruct the original dimensionality of the ECG data.

IV. FEATURE EXTRACTION:

A. Extraction of independent components

An ICA strategy is proposed where extra learning about the time and statistical structure of the ECG sources is fused. ICA yields conditions depicting the conduct of the different ECG segments as an element of cardiovascular process duration. ICA can be utilized to synthesis an ECG signal which is a sensible multiplication of the original signal, and furthermore can control parameters, for example, QRS complex amplitude, rise-time, fall-time and the relative amplitudes of the P-and T-waves. The span of every component will consequently track the chosen heart rate in a non-linear form, mirroring its actual conduct. This will give a priceless and practical apparatus for testing, adjusting and keeping up electrocardiographic equipment in doctor's facilities and centers, and for the plan/change of new and existing instrumentation.

PCA isn't an extremely proper procedure for the representation of ECG data and nonlinear dimensionality lessening calculation from the ECG morphological perspective. This is because of the way that it can just reveal linear connections in the ECG data, and is intended to discover the bearings in the ECG data with the most maximum change, which may not generally be the most instructive headings. The subsequent low-dimensional PCA descriptors can be utilized for exploratory ECG data investigation, perception and resulting ECG data demonstrating. It has been accounted for in the literature that the ICA source gauges advance some inverse procedure, following up on the ECG perceptions. In a standard ICA method, this inverse capacity is a summed up linear capacity, i.e. a component of the pseudo-inverse of the mixing matrix, the alleged unmixing matrix and the induced noise demonstrate. ICA is relied upon to expel noise from the ECG with known attributes (Fig 7).

In this paper, propose that the coordination of PCA and

ICA strategies can productively remove the noise and artifacts from the ECG signals. The consequences of PCA and ICA when utilized together demonstrated great viability and informative ECG classification.

A.ICA steps

First step involves the determination of independent components by removing the mean values of the variables also known as 'centering the ECG data'. The second step is to 'whiten the ECG data' also known as 'sphering the ECG data'. In the third step, independent components are obtained by applying a linear transformation to the whitened ECG data.

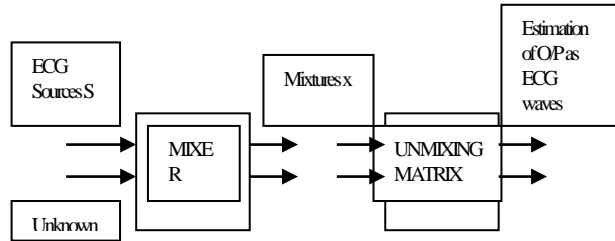


Fig.VII Basic understanding of ICA application for ECG analysis.

To estimate one of the independent components, a linear combination of the x_i is considered. Let us denote this by

$$y = w^T x = \sum_{i=1}^n w_i x_i$$

Where the column vector w is to be determined. The independent components are determined by applying a linear transformation to the whitened data. A given component can be obtained using the linear transformation

$$i_c = b_i^T x,$$

Where i_c gives an independent component which is an estimate of the original signal, b is an appropriate vector to reconstruct the independent components. In order to employ the ECG signal for facilitating interpretation and medical diagnosis, ICA is used to clean the ECG signal by removing some or all the sources of noise. By using ICA, the basic idea is to 'project out' the noise and artifacts from ECG signals and to represent noise and artifacts as independent components.

The ICA algorithmic principle is

$$y = Est(s) = w_{ICA} x,$$

$$y = wx$$

The idea of ICA is to recover the original signals by assuming that they are statistically independent. y is independent and it is desired to find

how w maximizes the independence of y . After estimating A , computation of $w = A^{-1}$ is done, which gives

$$s = wx = A^{-1}x$$

Using ICA, it is required to derive a 'clean ECG signal' from the source ECG signal to find the noise reduction factor.

V. GENETICAL ALGORITHM (LEARNING/ CLASSIFICATION)

In the subsequent stage, Genetic Algorithm is utilized to pick the more significant features. The input data is changed to higher dimension utilizing a non-linear transfer function (polynomial function) and GA is utilized to choose the ideal subset of the component clusters with the fitness work taken as the acknowledgment execution. The constraints in PCA and ICA are, (i) We can't decide the fluctuations (energies) of the components, (ii) We can't decide the order of the components, and (iii) we can't expect the chosen components has same level with number of vectors from each class. Here, the GA is utilized to choose the ideal subset of components from each cluster; thus the above said restrictions can be overcome.

In our proposed method, we are going to choose only F number of components, for each cluster, the reduced feature set will contain $S = NC * F$ number of features, where NC is the number of clusters. In our case we have two clusters: abnormal and normal. Initially an n number of components are selected from each cluster at random. The index of each component is used to construct one chromosome. Similarly N number of chromosomes is generated. ($N=10$). For example, consider the chromosome:

980 726 657 807 240 825.....
2 4 2 5 2 2 2 0 9 2 5 3 4 1 5 5

Each integer represents one component. The first 980 stands for the 980th components from the first cluster, the 24 in the second row represents the 24th component from the second cluster. The total length of the chromosome is equal to the total number of independent components required. Here, we kept the size as 600. For each chromosome, the Euclidian distance within the class (W) and between the classes (B) has been calculated. The fitness value is calculated as:

$$f(x) = B / W \quad (14)$$

The chromosome which has the minimum fitness value (G_{min}) is stored as the best independent component set. Then the genetic operators are applied

to search for the optimum set.

Reproduction (selection) – The selection procedure chooses chromosomes from the mating pool coordinated by the survival of the fittest idea of normal genetic systems. In the corresponding choice system embraced in this article, a chromosome is allocated various duplicates, which is relative to its fitness in the populace that goes into the mating pool for advance genetic operations. Roulette wheel choice is one regular system that executes the corresponding determination procedure.

Crossover – is a probabilistic procedure that trades data between two parent chromosomes for creating two child chromosomes. In this paper, single point crossover with a settled crossover likelihood of $p_c=0.6$ is utilized. For chromosomes of length l , an arbitrary whole number, called the crossover point, is created in the range $[1, l-1]$. The segments of the chromosomes deceiving the privilege of the crossover point are traded to create two offspring.

Mutation – Each chromosome undergoes mutation with a fixed probability $p_m=0.003$. For binary representation of chromosomes, a bit position (or gene) is mutated by simply flipping its value. Since we are considering real numbers in this paper, a random position is chosen in the chromosome and replace by a random number between 0-9.

The new population is generated after the genetic operators are applied. The current best IC set is (L_{min}) selected from the new population and compared with the global one. If the global set contains minimum fitness value then the local, the next iteration is continued with the old population. Otherwise, the current population is considered for the next iteration. This process is repeated for k number of iterations. The algorithm is given as:

1. Construct the initial population (p_1) with random ICs.
2. Calculate the fitness value $f(x) = B / W$.
3. Find out the Global minimum (G_{min})
4. For $i = 1$ to k do
 - a. Perform reproduction
 - b. Apply the crossover operator between each parent.
 - c. Perform mutation and get the new population. (p_2)
 - d. Calculate the local minimum (L_{min}).
 - e. If $G_{min} < L_{min}$ then
 - i. $G_{min} = L_{min}$;
 - ii. $p_1 = p_2$;

VI. EXPERIMENTS AND RESULTS

The Database is used for evaluation was taken from private hospital in our proposed algorithm. This database consists of ECG recordings from 600 subjects. The subjects were 300 men aged 30 to 84, and 300 women aged 55 to 71. The database includes ST segment change and of T-wave change. Each record is two hours in duration and contains two signals, each sampled at 250 samples per second with 12-bit resolution over a nominal 20 millivolt input range. The sample values were rescaled after digitization with reference to calibration signals in the original analog recordings, in order to obtain a uniform scale of 200 ADC units per millivolt for all signals. (The calibration signals are not included in the signal files.) The header files include information about the leads used, the patient's age, sex, and medications, the clinical findings, and the recording equipment. Each of the signal files is 5,400,000 bytes long. Two cardiologists worked independently to annotate each record beat-by-beat and for changes in ST segment and T-wave morphology, rhythm, and signal quality. ST segment and Twave changes were identified in both leads (using predefined criteria which were applied uniformly in all cases), and their onsets, extrema, and ends were annotated. Annotations made by the two cardiologists were compared, disagreements were resolved by the coordinating group, and the reference annotation files were prepared

In this paper, we have taken the full length ECG signals from patients with normal, diabetics and blood pressure conditions. And the signals are sliced with 60 seconds time interval; each signal will be translated into 120 samples and totally 2040 beats. This dimensionality is reduced by GICA as discussed in the earlier section.

The performance of our proposed method is compared with linear PCA, ICA classifier. Also the performance is studied with Genetic based ICA, here the clustering step is excluded and the components have been chosen directly with GA. The performance is studied with the measurements like classification accuracy, sensitivity and specificity analysis. The classification accuracies for the datasets are measured using the equation:

$$Accuracy(T) = \frac{\sum_{i=1}^{|T|} assess(t_i)}{|T|}, t_i \in T_r$$

$$Assess(t) = \begin{cases} 1, & \text{if } classify(t) = t.c, \\ 0, & \text{otherwise} \end{cases}$$

where T is the set of data items to be classified (the test set), $t \in T$, t.c is the class of item t, and classify(t) returns the classification of t by genetic.

The Sensitivity, specificity, TP rate, FP rate, and accuracy are calculated as:

$$Sensitivity = (TP / (TP+FN)) (\%)$$

$$Specificity = (TN / (FP+TN)) (\%)$$

$$FPrate = (FP / (FP+TN))$$

$$TPrate = (TP / (TP+FN))$$

Accuracy = (TP+TN) / (TP+FN+FP+TN) where TP, TN, FP, and FN denotes true positives, true negatives, false positives, and false negatives, respectively.

TABLE I. PERFORMANCE ANALYSIS OF BEAT CLASSIFICATION

METHODS	PCA	ICA	GICA
Sensitivity (%)	88	92	94
Specificity (%)	85	90	92.5
FP rate (%)	15	10	7.5
TP rate (%)	88	92	94
Accuracy (%)	86.7	91	93.3
Az Value	0.87	0.92	0.94

TABLE II. THE VALUE OF SENSITIVITY AT EACH FOLD

	PCA	ICA	GICA
1	84	91	94
2	86	91	94
3	87	91	93
4	88	93	94
5	88	92	94
6	88	92	94
7	89	92	94
8	90	92	94
9	89	93	94
10	91	93	95

VII. SIMULATION RESULTS

GA is a searching algorithm, when running this algorithm the GA will start searching for the solutions in the solution space (to find the best values for the weights) according to specific criterion, sometimes the initial values of the population (the initialization of the population is random) is too close to the desired value so that it will not take a very long time. The GA is used in this paper to enhance the training of ICA trained in standard learning technique and minimize the error value. After training this ICA, the weight evaluated by the classical algorithm is used as initial population to the GA, and the GA is applied to find the optimum values to the weights.

When GA parameters are configured, the algorithm is ready to run for optimizing the connection weights of the ICA. Before running the program the input signal must be prepared in the Matlab work space. The program starts analyzing the ECG input signal. After few minutes the analysis is completed and the report is displayed. The figures show the analysis report window with three different cases of ECG signals.

The program can diagnose three different cases related with the rhythm of the heart, they are:

- Normal condition
- Diabetics
- Blood pressure

1. Train the signals for diabetics, normal and blood pressure.
2. After training complete input a test ECG signal and proceed analyze.

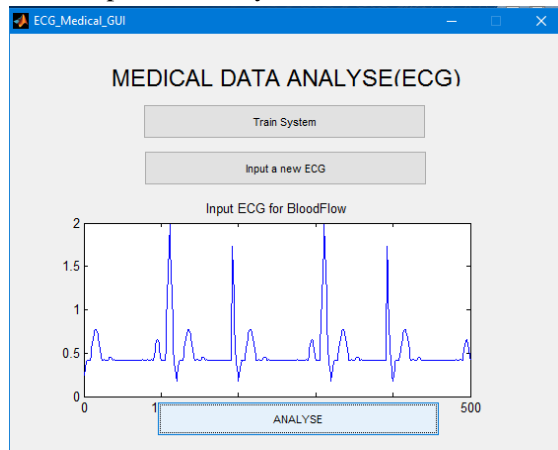


Fig 8: GUI for proposed model

3. In analyze section it shows PCA and ICA graphs for each input

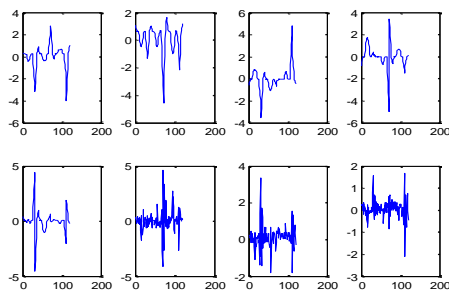


Fig 9: waveform for PCA (diabetics)

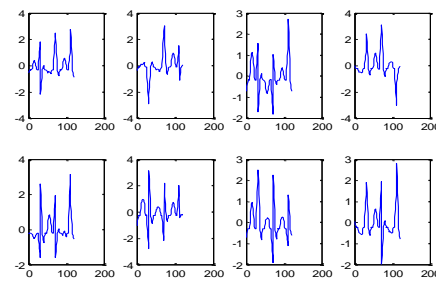


Fig 10: waveform for ICA (diabetics)

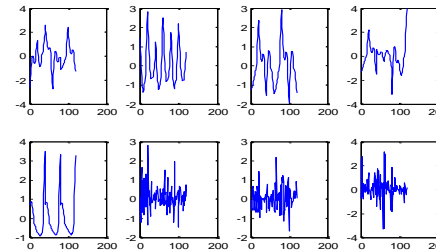


Fig 11: waveform for PCA (low pressure)

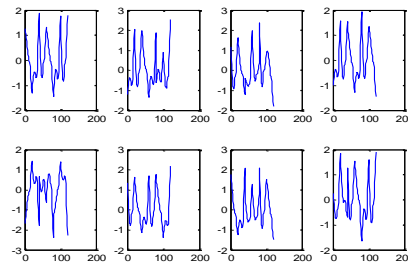


Fig 12: waveform for ICA (low pressure)

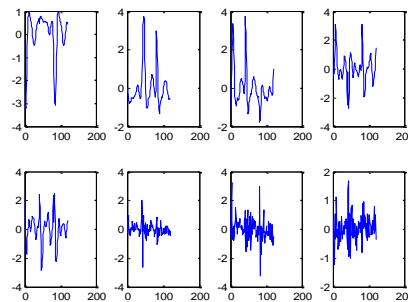


Fig 13: waveform for PCA (normal)

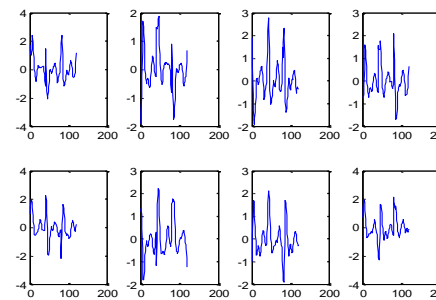


Fig 14: waveform for ICA (normal)

3. In diagnosis section it predicts the disease using GA.

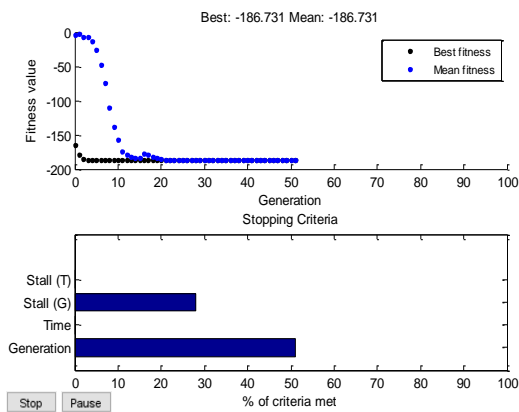


Fig 15: waveform for optimized peak detector weight (diabetics)

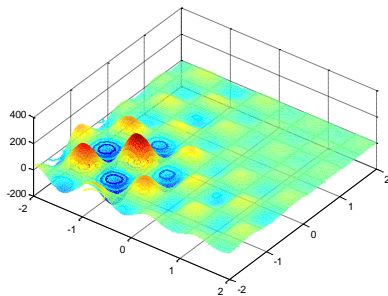


Fig 16: Surface graph (diabetics)

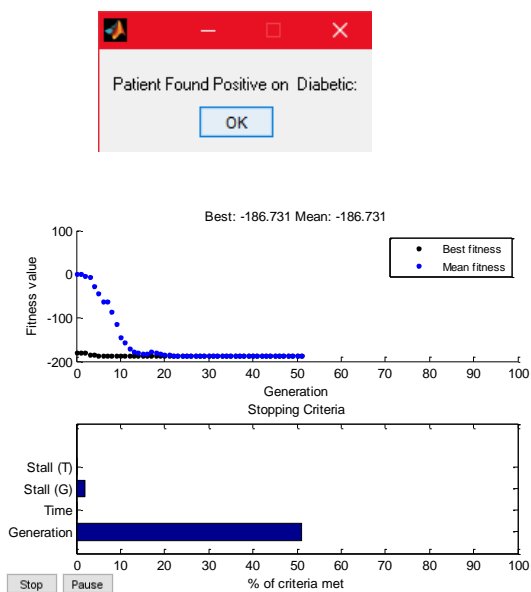


Fig 17: waveform for optimized peak detector weight (low pressure)

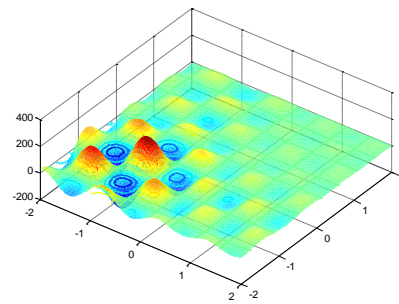


Fig 18: Surface graph (low pressure)

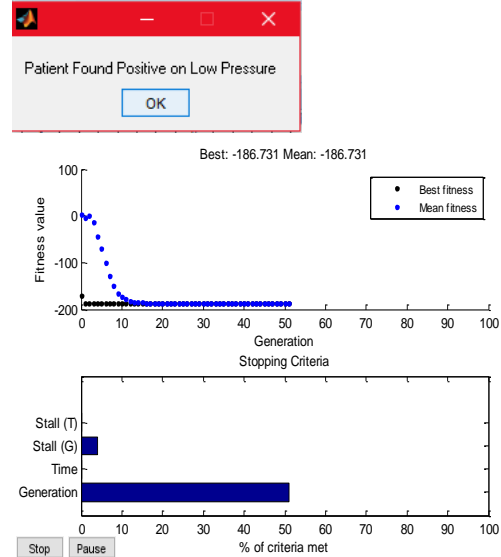


Fig 19: waveform for optimized peak detector weight (normal)

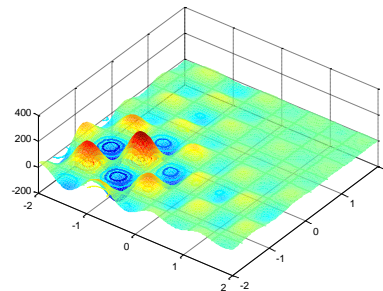
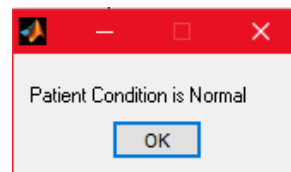


Fig 20: Surface graph (normal)



4. In diagnose section for different input ECG signals it predicts the mean value is 76.3481 and index value is 2 for diabetic case, for low pressure the mean value is 120.0273 and index value is 10 and for normal case the mean value is 180.8475

and index value is 6. The index value is the predefined threshold limit.

VIII. CONCLUSION

In this work, a technique for automatic programmed classification of ECG signals from three distinct clusters normal, diabetics and blood pressure has been proposed. Technique initially utilizes smoothness priors way to deal with pre-process all ECG signals from the database keeping in mind the end goal to diminish the baseline drifts and different patterns in the signals utilizing PCA. ICA based strategy is utilized to distinguish peaks in the preprocessed ECG signals connected in the second phase of the proposed framework before the coefficients of a prepared signal model are separated and used to characterize each area of ECG signal into one of three possible cluster. Choice about the sort of every ECG signal from the test set is then made relying upon the gathering into which the vast majority of the areas from a similar ECG have been assembled. Extracted features, parameter sets are well separated in feature space and accurately classified, indicating that the high classification accuracy can be expected in the practical application using Genetic algorithm on the proposed system. For the standard set of ECG test signals has taken from private hospital managed to achieve 96% accurate classification of three heart conditions. Future work includes repeating the experiment with other advanced machine learning algorithms such as PSO and Neural network.

REFERENCES

- [1] G.D. Clifford, F. Azuaje, P. McSharry, *Advanced Methods And Tools for ECG Data Analysis*, 1st ed., Artech House Publishers, 2006.
- [2] O. Sayadi, M.B. Shamsollahi, *Multiadaptive bionic wavelet transform: application to ECG denoising and baseline wandering reduction*, *EURASIP J. Adv. Signal Process.* 2007(14) (2007) 1–11.
- [3] O. Sayadi, M.B. Shamsollahi, *ECG denoising and compression using a modified extended Kalman filter structure*, *IEEE Trans. Biomed. Eng.* 55 (9) (2008) 2240–2248.
- [4] J.P. Martinez, R. Almeida, S. Olmos, A.P. Rocha, P. Laguna, *A wavelet-based ECG delineator: evaluation on standard databases*, *IEEE Trans. Biomed. Eng.* 51 (4) (2004) 570–581.
- [5] C. Li, C. Zheng, C. Tai, *Detection of ECG characteristic points using wavelet transforms*, *IEEE Trans. Biomed. Eng.* 42 (1) (1995) 21–28.
- [6] M. Bahoura, M. Hassani, M. Hubin, *DSP implementation of wavelet transform for real time ECG wave forms detection and heart rate analysis*, *Comput. Method Programs Biomed.* 52 (1) (2007) 35–44.
- [7] P. de Chazal, M. O'Dwyer, R.B. Reilly, *Automatic classification of heartbeats using ECG morphology and heartbeat interval features*, *IEEE Trans. Biomed. Eng.* 51 (7)(2004) 1196–1206.
- [8] G. de Lannoy, D. Francis, J. Delbeke, M. Verleysen, *Weighted conditional random fields for supervised inter patient heartbeat classification*, *IEEE Trans. Biomed. Eng.* 59 (1) (2012) 241–247.
- [9] M. Kundu, M. Nasipuri, D.K. Basu, *Knowledge-based ECG interpretation: a critical review*, *Pattern Recogn.* 33 (3)(2000) 351–373.
- [10] ANSI/AAMI, *Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms*, American National Standards Institute, Inc. (ANSI), Association for the Advancement of Medical Instrumentation (AAMI), ANSI/AAMI/ISO EC57,1998-(R)2008, 2008.
- [11] M.S. Spach, J.M. Kootsey, *The nature of electrical propagation in cardiac muscle*, *Am. J. Physiol. Heart Circ. Physiol.* 244 (H) (1983) 3–22.
- [12] E. Besterman, R. Creese, *Waller-pioneer of electrocardiography*, *Br. Heart J.* 42 (1) (1979) 61–64.
- [13] H.P. da Silva, C. Carreiras, A. Lourenc,o, A. Fred, R.C. dasNeves, R. Ferreira, *Off-the-person electrocardiography: performance assessment and clinical correlation*, *HealthTechnol.* 4 (4) (2015) 309–318.
- [14] Y.M. Chi, T.-P. Jung, G. Cauwenberghs, *Dry-contact and noncontact biopotential electrodes: methodological review*, *IEEE Rev. Biomed. Eng.* 3 (2010) 106–119.
- [15] R.C. Martins, D. Primor, T. Paiva, *High-performance groundless EEG/ECG capacitive electrodes*, in: 2011 IEEE International Workshop on Medical Measurements and Applications Proceedings (MeMeA), 2011, pp. 503–506.
- [16] T. Barill, *The Six Second ECG: A Practical Guidebook to Basic ECG Interpretation*, nursecom, 2003.
- [17] M. Llamedo, J.P. Martinez, *An automatic patient-adapted ECG heartbeat classifier allowing expert assistance*, *IEEE Trans. Biomed. Eng.* 59 (8) (2012) 2312–2320.
- [18] I. Toma'si'c, R. Trobec, *Electrocardiographic systems with reduced numbers of leads – synthesis of the 12-lead ECG*, *IEEE Rev. Biomed. Eng.* 7 (2014) 126–142.
- [19] J.C. Principe, *Editorial*, *IEEE Rev. Biomed. Eng.* 7 (2014) 1–2.
- [20] P. de Chazal, *Detection of supraventricular and ventricular ectopic beats using a single lead ECG*, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, pp. 45–48.
- [21] Chawla, M.P.S., Verma, H.K. and Kumar, V., 2007, *A new statistical PCA-ICA algorithm for location of R-peaks in ECG*, *International Journal of Cardiology*, (in press, available online 25th July 2007).
- [22] Chawla, M.P.S., Verma, H.K. and Kumar, V., 2006, *ECG modeling and QRS detection using principal component analysis*. Paper no. 04, in IET Proceedings, International Other, MEDSIP-06, Glasgow, UK, July.
- [23] Chawla, M.P.S., Verma, H.K. and Kumar, V., 2006,

Data reduction and removal of base-line wander using principal component analysis, Communicated in DSP, Elsevier in May 2006 (Under review).

[24] Gao, P., Chang, E.C. and Wyse, L., 2003, Blind separation of fetal ECG from single mixture using SVD and ICA, Proceedings of ICICS-PCM 2003, 15–18 December 2003, Singapore.

[25] Chawla, M.P.S., Verma, H.K. and Kumar, V., 2006, PCA-ICA Method for Detection of QRS Complexes and Location of R-Peaks in Electrocardiograms, Communicated in DSP, Elsevier in May, 2006 (Under review).

[26] Chawla, M.P.S., 2007, A comparative analysis of principal component and independent component techniques for electrocardiograms, International Journal of Neural Computing and Applications (NCA), (in press).