

Biometric Recognition using Area under Curve Analysis of Electrocardiogram

Anita Pal

Department of Computer Science & Engineering,
Institute of Engineering & Technology,
Dr. A. P. J. Abdul Kalam Technical University,
Lucknow, Uttar Pradesh, India

Yogendra Narain Singh

Department of Computer Science & Engineering,
Institute of Engineering & Technology,
Dr. A. P. J. Abdul Kalam Technical University,
Lucknow, Uttar Pradesh, India

Abstract—In this paper, we introduce a human recognition system that utilizes lead I electrocardiogram (ECG). It proposes an efficient method for ECG analysis that corrects the signal and extract all major features of its waveform. FIR equiripple high pass filter is used for denoising ECG signal. R peak is detected using Haar wavelet transform. A novel class of features called as area under curve are computed from dominant fiducials of ECG waveform along with other known class of features such as interval features, amplitude features and angle features. The feasibility of an electrocardiogram as a new biometric is tested on selected features that reports the authentication performance 99.49% on QT database, 98.96% on PTB database and 98.48% on MIT-BIH arrhythmia database. The results obtained from the proposed approach surpasses the other conventional methods of biometric applications.

Keywords—Electrocardiogram; biometric; area under curve features

I. INTRODUCTION

First generation biometrics comprised of fingerprint, face, iris and voice are now common. It is a time to move the next generation which is beyond the PINs and passwords towards more sophisticated security solutions that is ECG signal [2], [3]. ECG is an internal biometric, it has an intrinsic advantage of using a live signal for both accuracy and liveness detection, without the requirement of costly gear. ECG biometric validation offers staggering flexibility and the open door for cost-effective and highly-secure solution, to handle the developing threat of cyber-crime in present world. It can also reduce hacking or spoofing risks for greater security and convenience.

ECG is a biometric signal that is subjective to an individual and for this reason it is harder to mimic these signals. These are highly secure and prevent from any fear of imitation. The first report on ECG appeared in 1875 by Richard Caton [4]. The first human electrocardiogram was published by the British physiologist Augustus D. Waller in 1887 [4]. In 1895, Willem Einthoven improved the Electrometer and defined the main elements of ECG are P wave, QRS-complex and T waves [4]. Now, in most of hospitals the 12-leads method and the 5-leads are used for capturing the 1D ECG signal [5]. Improving the sensitivity of the electrodes and increasing the comfort of the measurement of the ECG are the hot topics of ECG measurement in medical research. For identification based on the ECG, convenience and accuracy are very important factors.

The relaxation as well as contraction of cardiac muscle

results from the repolarisation and depolarization of myocardial cells. The electrodes placed on the chest wall and limbs are recorded due to these electrical changes in the heart and electrocardiogram are produced by transcription onto graph sheet. The electrical potential is contractile heart muscle cells known as cardiomyocytes. The electrocardiogram (ECG) is utilized to examine a few kinds of unusual heart function, including conduction distribution, arrhythmias, and in addition heart morphology [1]. A few wordings used as a part of ECG waveform are as per the following: (1) Isoelectric line is a benchmark or even line when there is no electrical movement on ECG. (2) Segments are the span of the isoelectric line between waves. (3) Intervals are the time between similar portions of contiguous waves. The P-wave is the primary deflection of the ECG signal, due to depolarization of the atria. The QRS complex followed by P wave and represents ventricular depolarization and T wave represents to ventricular repolarization i.e. rebuilding of the resting membrane potential. In around one fourth of population, a U wave can be found after the T wave. This more often than not has an indistinguishable extremity from the former T wave. It has been recommended that the U wave is caused after potentials that are likely created by mechanical electric input. The PQ segment relates to electrical driving forces transmitted through the SA node to bundle of His, bundle of His to its branches and then move to Purkinje fibers. The PQ interval communicates the time slipped from atrial depolarization to the beginning of ventricular depolarization. The ST to T interim corresponds with the moderate and fast repolarization of ventricular muscles. The QT interval represents ventricular activity potential and repolarization. Atria ventricles are in diastole at the point of TP interval.

Current research acknowledged that the heartbeat is a potential biometrics for identifying the individuals. The comparison of different biometric modalities, heartbeat is distinctive, difficult to be counterfeited and more universal. Biel et al. extracted fiducial feature consists of P, QRS and T waveforms and the feasibility of ECG signal is evaluated for human identification [6]. They used 12 lead for recording of ECG signals from 20 individuals of different age groups. They performed multivariate analysis for classification and accomplished identification rate 100%. The issue was that small quantities of test data set were used for validation. Yochum et al. proposed the continuous wavelet transformation to distinguish P, QRS and T wave depiction and histogram discover the veil choice of P, QRS and T wave [7]. This

methodology tested on MIT-BIH Arrhythmia and Computers in Cardiology Challenge 2011 databases. Wubbeler et al. proposed an ECG signal based on the human recognition for extracting biometric features on different leads. Genuine pairs were seen as those which were palatably related, while for another situation data signals were not accepted. The report of false acceptance rates and false rejection rates were 0.2% and 2.5%, respectively while the equal error rate was 2.8% [8]. The recognition rate was 99% for 74 individuals. Israel et al. demonstrated the Wilks Lambda technique and classification on linear discriminant analysis [9]. This system was tested on a galaxy set of 29 individuals and achieved 100% recognition rate. Shen et al. exhibited single lead ECG signal based on 7 fiducial features with identity verification with QRS complex [10]. The outcome of identity verification was 80% for decision based neural network, 95% for template matching and 100% for the combination of the two strategies from a social event of 20 individuals. Singh and Gupta have proposed P and T wave delineators alongside QRS complex to extricate diverse features from predominant fiducials of ECG signals [2] [12] [13] [3]. This framework tested on 50 individuals, accomplished the classification accuracy to 98%.

In this paper, a robust and an efficient method of ECG biometric recognition is proposed. For denoising ECG signal, FIR equiripple high pass filter is used that removes baseline noise. The FIR equiripple low pass filter removes the power interference noise. The accurate detection of the R peak (R_{peak}) with the help of Haar wavelet transforms. All other dominant features of the ECG waveform are detected with respect to the R_{peak} by setting of the windows whose size depend on the length of the corresponding wave duration and location. Extracted features of the ECG signal are successfully detected, these features are interval features, amplitude features, angle features, and area features. The algorithm is applied on 287 ECG signals of PTB database, 38 ECG signals from MIT-BIH arrhythmia database and 27 ECG signal from QT database from physionet bank and detect 36 features from each ECG signals. We use kernel principal component analysis reduction method on the extracted features. Finally the similarities within the components of feature set are calculated on the basis of euclidean distance.

The paper is structured as follows. Section II presents the methodology for identification based on ECG signal. The delineation techniques of P wave, QRS-complex and T wave are demonstrated with detailed description of ECG data. Section III describes the feature selection with the help of kernel principal component analysis. And the experimental results of the recognition system are presented in Section III. Section IV introduces the issues related to the recording of ECG signal. Finally, conclusions are drawn in Section V.

II. METHODOLOGY

The schematic depiction of the ECG individual recognition framework is appeared in Fig. 1. The strategy is implemented in a progression of following steps: (1) Preprocessing: incorporates modification of signal from noise artifacts and classify the ECG waveform (2) Data Representation: includes dominant characteristics of ECG signal (3) Feature Extraction: recognition of dominant features such as P wave, QRS complex and T wave (4) Classification (5) Decision making are based on

the method of template matching. The processing of the ECG biometric framework is shown in Fig. 1. It consists of the preprocessing, data representation and recognition. First, the heartbeat acquired from public database is preprocessed. The filtering process to remove noise and artifacts from the signal. The data representation step contain heartbeat detection and heartbeat segmentation process. The feature extraction includes the interval features, amplitude features, angle features and area under curve features from a group heartbeats. The feature vectors of kernel principal component analysis attributes are stored in the template database.

For authentication process the euclidean distances are calculated by selecting the first window from each subject. The euclidean distances are calculated within the windows of same subject, it is known as genuine score whereas euclidean distances are calculated within the windows of different subjects is known as imposter score. The euclidean distance are calculated between two feature vector F1 and F2 are as follows,

$$\text{Euclidean distance (F1, F2)} = \sqrt{(F1 - F2)^2}$$

The performance of ECG biometric authentication system can be evaluated on basis of equal error rate (EER). The point where the proportion of false acceptance rate (FAR) is the same as false reject rate (FRR) that is (FAR = FRR) is known as equal error rate. The genuine acceptance rate (GAR) can be evaluated as

$$\text{GAR} = 100 - \text{FRR}$$

The receiver operating characteristic (ROC) curve plot is a decision threshold which plots the rate of false acceptance rate against the false rejection rate. The accuracy of the recognition system is determined as

$$\text{Accuracy} = 100 - \text{EER}$$

In ECG biometric identification system compares the unknown subject to all the subjects stored in a database to determine if there is a match (1 : N), it compares each existing subject stored in the database against the newly enrolled subject.

A. Data Preprocessing

An electrocardiogram parades the electrical activity in the heart, that can be represented using P, Q, R, S, and T waveforms. At the point when an ECG signal is recorded, it may be corrupted with different kinds of noise. The exploitation of unadulterated ECG signal from noisy estimations has been one of the essential contemplation of biomedical signal processing. The required techniques are applied to maintain the important information of the recorded ECG signal. Different types of artifact and interference can contaminate the real amplitude and time period of the ECG signal. ECG signals are mostly affected by baseline wander noise and power line interference noise. These artifacts and interference produce the incorrect diagnosis of the ECG signal. It is difficult to eliminate the artifacts and interference present in ECG signal [14]. Digital filters are mostly used to improve the quality of the ECG signal .

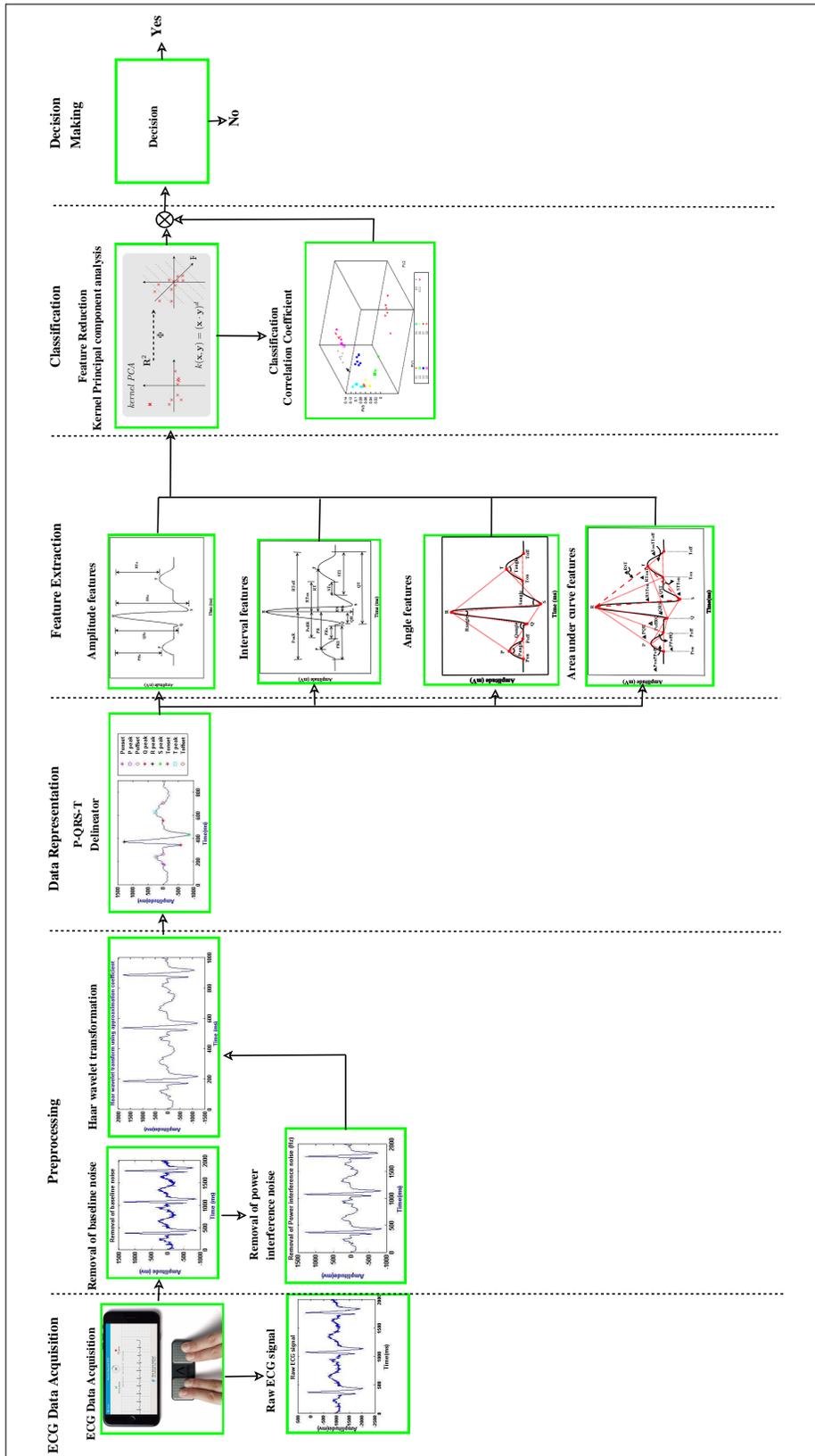


Fig. 1. ECG biometric recognition system.

Using signal processing techniques the quality of ECG signal improves. For example, signal filtering allows certain frequency components to pass through the system without any distortion and attenuated other frequency components, this operation is known as filter [14]. In passband the range of frequencies is permitted to pass through the filter and in stopband the range of frequencies is hindered by the filter. A lowpass filter allows to pass all lower frequency components below the cutoff frequency and blocks all higher frequency components above them [14]. A highpass filter allows all frequency component above cutoff frequency and stop other frequency components.

Equiripple filter has equal ripples in stopband as well as in passband. The signals are distorted at the edge of the passband. The presence of large ripple in equiripple design is avoided. Equiripple design limits the total passband width and has a large transition band. Equiripple filters seek to minimize the maximum error between the desired filter response and the designed approximation [14].

1) *Equiripple linear phase FIR filter*: The linear phase filter acquired by minimizing the weighted error of the peak estimated value ε is known as equiripple FIR filter. At the point when ε minimize, the weighted error function shows an equiripple behavior in the frequency range of intrigue. The principle of Parks-McClellan algorithm, is broadly utilized for outlining the equiripple linear phase FIR filter [14]. The general condition of the frequency response $G(e^{j\omega})$ of a linear phase FIR filter whose length $N+1$ can be defined as follows,

$$G(e^{j\omega}) = e^{j\frac{N\omega}{2}} e^{j\beta} \hat{G}(\omega) \quad (1)$$

where $\hat{G}(\omega)$ is the amplitude response of the signal. The weighted error function with respect to amplitude response can be defined as

$$\varepsilon(\omega) = W(\omega)[\hat{G}(\omega) - D(\omega)] \quad (2)$$

where preferred amplitude response is $D(\omega)$ and the positive weight function is $W(\omega)$. $W(\omega)$ is used to control the span of peak errors in the specific frequency bands. The Parks-McClellan calculation depends on iterative modifying of the amplitude response until the peak absolute value of $\varepsilon(\omega)$ is minimized. If, minimum estimation of the peak absolute value of $\varepsilon(\omega)$ in band $\omega_a \leq \omega \leq \omega_b$ is ε_0 , then the absolute error fulfill the following equality,

$$|\hat{G}(\omega) - D(\omega)| \leq \frac{\varepsilon_0}{|W(\omega)|}, \omega_a \leq \omega \leq \omega_b \quad (3)$$

In some condition the preferred amplifier response is found to be,

$$D(\omega) = \begin{cases} 1, & \text{passband,} \\ 0, & \text{stopband.} \end{cases} \quad (4)$$

The amplifier response $\hat{G}(\omega)$ is necessary to fulfill the above desired response with a ripple positive as well as negative φ_p in the passband and a ripple φ_s in the stopband. The weighted function can be taken as follows,

$$W(\omega) = \begin{cases} 1, & \text{passband,} \\ \frac{\varphi_p}{\varphi_s}, & \text{stopband.} \end{cases} \quad (5)$$

The amplitude response for all 4 types of FIR filters can be written as

$$\hat{G}(\omega) = R(\omega)B(\omega), \quad (6)$$

where the factors of $R(\omega)$ are

$$R(\omega) = \begin{cases} 1, & \text{Type 1,} \\ \cos(\frac{\omega}{2}), & \text{Type 2,} \\ \sin(\omega), & \text{Type 3,} \\ \sin(\frac{\omega}{2}), & \text{Type 3.} \end{cases} \quad (7)$$

The factor $B(\omega)$ is represented as

$$B(\omega) = \sum_{k=0}^P \hat{b}[k] \cos(\omega k), \quad (8)$$

where

$$\hat{b}[k] = \begin{cases} b[k], & \text{Type 1,} \\ \hat{x}[k], & \text{Type 2,} \\ \hat{y}[k], & \text{Type 3,} \\ \hat{z}[k], & \text{Type 3.} \end{cases} \quad (9)$$

$$P = \begin{cases} F, & \text{Type 1,} \\ \frac{2F-1}{2}, & \text{Type 2,} \\ F-1, & \text{Type 3,} \\ \frac{2F-1}{2}, & \text{Type 3.} \end{cases} \quad (10)$$

The modified weighted approximation function is as follows

$$\varepsilon(\omega) = W(\omega)[R(\omega)B(\omega) - D(\omega)] \quad (11)$$

$$= W(\omega)R(\omega) \left[B(\omega) - \frac{D(\omega)}{R(\omega)} \right] \quad (12)$$

The optimization issue now turns into the determination of the coefficients $\hat{b}[k], 0 \leq k \leq P$ which limits the peak absolute value ε of the weighted approximation error $\varepsilon(\omega)$ over the specified frequency band. After the coefficients have been determined the corresponding coefficient of the original amplitude response are computed from which the filter coefficient are then obtained [14]. If we design a filter is of Type 2 then $x[k] = a[k]$ and $F = (\frac{2P+1}{2})$ from Eq.(10), we find the next $x[k]$.

B. Baseline Wander Noise Removal

The baseline wander noise effect the base axis of ECG signal that is viewed on a screen to move up and down rather than being straight. Therefore it causes the entire signal to shift from its normal base. This is due to improper electrodes and movement of the patient or by respiration [15]. Equiripple highpass filter can remove this noise completely without affecting the dominant fiducials of the ECG signal [16]. Equiripple highpass filter allows the dominant fiducials of ECG signal to pass through it [14]. The built-in function (filtfilt) requires the length of data signal to be more than three times greater than the filter order. The Equiripple highpass filter has a stop frequency of 2 Hz, filter order of 2700, cutoff frequency of 1 Hz, and stop attenuation of 80 dB. Fig. 2 shows the removal of baseline noise from the raw ECG signal.

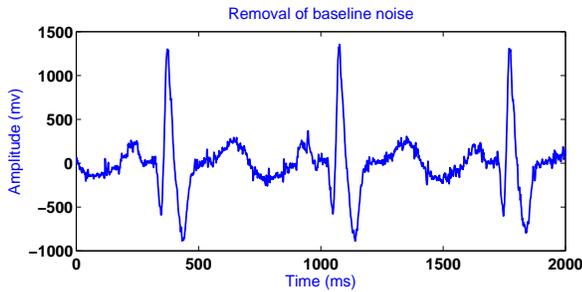


Fig. 2. Baseline noise removal

C. Power Interference Noise Removal

Power interference noise in ECG signal is due electromagnetic fields and addition of 50 or 60 Hz depending on power frequency standard. The power interference noise comes into a sight as a spike in frequency components analysis (FFT) at 50 Hz [14]. This FFT can be removed by using a band pass filter. The low pass filter is adequate for the removal of power interference noise. FIR equiripple lowpass filter components defined as filter order of 506 [17] and the cutoff frequency is 40 Hz. This filter is followed by another filter with zero phase to avoiding time delay by utilizing a Matlab builtin function filtfilt. The removal of power interference noise is shown in Fig. 3

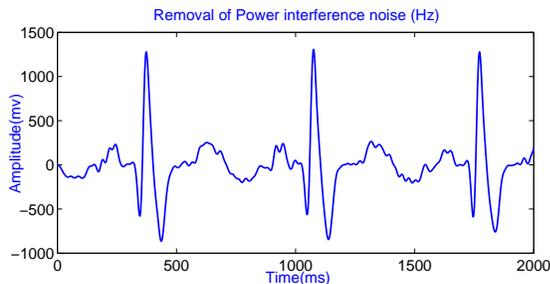


Fig. 3. Power interference noise (Hz) removal

D. Haar Wavelet Transformation

Haar wavelet is applied to ECG signal for feature extraction and it introduces the highest accuracy. Using the Haar wavelet method, R_{peak} is easily obtained. Haar wavelet transformation generates two coefficients called approximation and detail coefficients [18]. In the second order detail coefficient, R_{peak} are dominant feature since the $QRS_{complex}$ has a higher frequency in a shorter time interval [29]. Fig. 4 shows the decomposed ECG signal of second detail coefficient by using Matlab function wavdec and detcoef [20].

E. Peak Detection

1) *QRS-Complex Detection*: For the detection of R_{peak} , we firstly divide the input ECG signal into different subbands with the help of wavelet transformation. Then, reconstruct and calculate the second approximation coefficient by Haar transformation. Therefore, by using an adaptive threshold

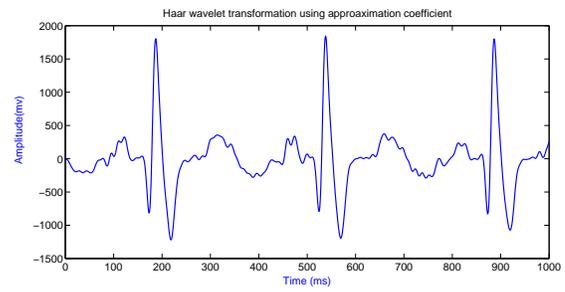


Fig. 4. Reconstruction of ECG signal from approximation coefficient using Haar wavelet transform

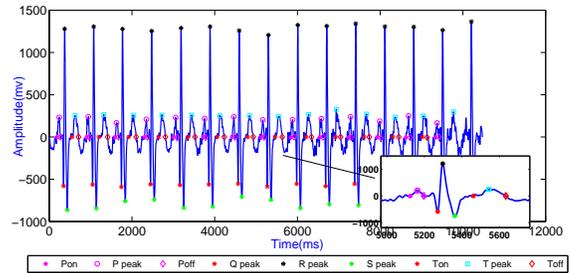


Fig. 5. Detection of ECG waveforms fiducial points such as P peak, Pon, Poff, Q peak, R peak, S peak, T peak, Ton and Toff.

method the maximum amplitude value of the ECG signal is detected that is R_{peak} . With the help of R_{peak} locations, we can find P, Q, S and T waves location. The number of heartbeats per minute (N_{BPM}) is calculated as follows:

$$N_{BPM} = \frac{R_{peak} * Y}{F * 60} \quad (13)$$

where Y is the ECG signal of one heartbeat and F is the frequency of the signal. Q_{peak} is detected by setting the window width of 90 ms. This window starts at 10 ms on the left of R_{peak} and ends at 100 ms away from R_{peak} . Within the boundary of this window, the samples that have minimum amplitude value on the left side of R_{peak} is the Q_{peak} location. The S_{peak} is detected by setting the window width of 95 ms. This window starts at 5 ms on the right of R_{peak} and ends at 100 ms away from R_{peak} . Within the boundary of this window, the samples that have minimum amplitude value on the right side of R_{peak} is the S_{peak} location. Fig. 5 shows detected $QRS_{complex}$ in ECG signal.

2) *P-peak Detection*: We set the width of window size 155 ms. This window extends from 200 to 45 ms to the left of R_{peak} [12]. Within the boundary of this window, the location that have maximum amplitude value is the P_{peak} location. P_{on} and P_{off} are detected by setting the window width of 300 ms. This window starts at 150 ms on the left of P_{peak} and have minimum amplitude value is P_{on} location, whereas it ends at 150 ms on the right of P_{peak} as well as it also have minimum amplitude value is P_{off} location.

3) *T-peak Detection Algorithm*: T_{peak} is the next prominent peak after R_{peak} . T_{peak} is detected by setting the window size of width 300 ms. This window starts at 100 ms on the

right of R_{peak} and ends at 400 ms away from R_{peak} . Within the boundary of this window, the samples that have maximum amplitude value on the right side of R_{peak} is the T_{peak} location [13]. T_{on} and T_{off} are detected by setting the window width of 300 ms. This window starts at 150 ms on the left of T_{peak} and have minimum amplitude value is T_{on} location, whereas it ends at 150 ms on the right of T_{peak} as well as it also have minimum amplitude value is T_{off} location.

F. Feature Extraction

Once known the limitation and peak of the QRS-complex, P wave and T wave of ECG signals, 36 features are extracted from each heartbeat those derive from one of the classes.

1) *Interval Features:* The interval features as shown in Table I related to heartbeat are computed as follows:

TABLE I. INTERVAL FEATURES CLASS ARE SELECTED FROM EACH HEARTBEAT

Interval Features	Representation
PR interval	PR_I
PR segment	PR_S
QT interval	QT_I
ST segment	ST_S
ST interval	ST_I
P_{onset} to R_{peak} segment	$P_{on}R_S$
P_{peak} to R_{peak} segment	PR_S
P_{offset} to R_{peak} segment	$P_{off}R_S$
Q_{peak} to R_{peak} segment	QR_S
R_{peak} to S_{peak} segment	RS_S
R_{peak} to T_{peak} segment	RT_S
R_{peak} to T_{onset} segment	$RT_{on}S$
R_{peak} to T_{offset} segment	$RT_{off}S$
RR interval	RR_I
PP interval	PP_I
TT interval	TT_I

The interval features are calculated with respect to P_{peak} are PR_I is the time difference between P_{peak} to R_{peak} fiducials and PR_S is the time difference between P_{off} to QRS_{on} fiducials. The QT is the corrected time difference between QRS_{on} to T_{off} fiducials. The ST_S is the time difference from QRS_{off} to T_{on} fiducials and ST_I is the time difference from QRS_{off} to T_{off} fiducials [21]. The interval feature class are calculated with respect to R_{peak} fiducial. PR is the time difference from P_{peak} to R_{peak} fiducials, $P_{off}R$ is the time difference P_{off} to R_{peak} and $P_{on}R$ is the time difference from P_{on} to R_{peak} fiducials. The time difference from R_{peak} to Q_{peak} fiducials and R_{peak} to S_{peak} fiducials is define as QR and RS . The time difference from R_{peak} to T_{on} fiducials, R_{peak} to T_{peak} fiducials and R_{peak} to T_{off} fiducials are defined as RT_{on} , RT and RT_{off} respectively [2]. The calculated time difference features are shown in Fig. 6. With these interval features within a heartbeat three interbeat interval feature RR , PP and TT are also extracted. RR is defined as the time difference between two consecutive R_{peaks} , similarly PP and TT are also detected [2] [22] [23]. The interval features between two ECG signals is shown in Fig. 7.

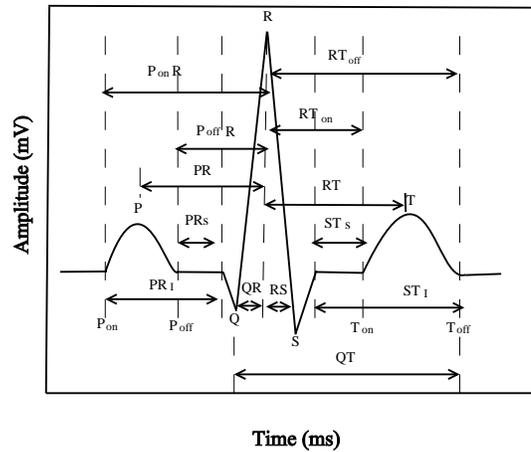


Fig. 6. Interval features class are selected from ECG fiducials.

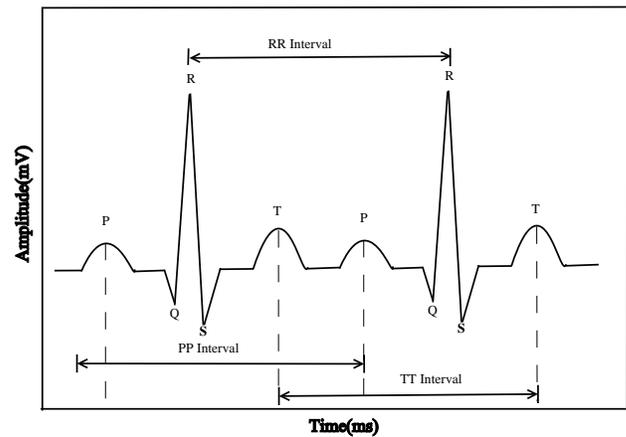


Fig. 7. Interval features between two heartbeats.

2) *Amplitude Features:* Amplitude is a measure of its change over a solitary period. The amplitude features as shown in Table II are calculated with respect to the amplitude of R_{peak} . The change in the heart rate is dependent on QRS-Complex. The feature QR_a is defined as a difference between the amplitude of R_{peak} and Q_{peak} . Feature SR_a is defined as a difference between the amplitude of R_{peak} and S_{peak} . Likewise, the variance in amplitude of P_{peak} to R_{peak} and T_{peak} to R_{peak} are characterized as PR_a and TR_a , respectively [2] [3] [25] [24]. These amplitude features are shown in Fig. 8.

TABLE II. AMPLITUDE FEATURES CLASS ARE CONSIDERED FROM EACH HEARTBEAT

Amplitude Features	Representation
QR amplitude	QR_a
PR amplitude	PR_a
RS amplitude	SR_a
RT amplitude	TR_a

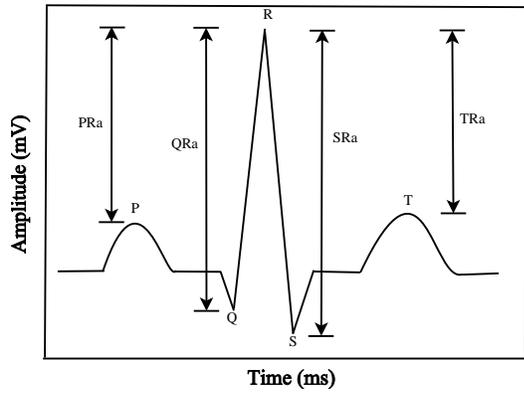


Fig. 8. Amplitude features class are considered from ECG dominant fiducials.

3) *Angle Features*: Following features as shown in Table III related to different peak fiducials of P_{on} , P_{peak} , P_{off} , Q_{peak} , R_{peak} , S_{peak} , T_{on} , T_{peak} and T_{off} waves are extracted from each heartbeat. The $\angle P$ is computed between the lines joining from P_{on} to P_{peak} and P_{peak} to P_{off} fiducials [26]. Let a, b, c are the sides of a triangle then using Cosine rule [26] we can find

$$\cos(A) = \frac{b^2 + c^2 - a^2}{2 * b * c}$$

Therefore, we can compute angle features as follows:

Angle P:

$$\cos(P) = \frac{P_{on}P^2 + PP_{off}^2 - P_{on}P_{off}^2}{2 * (P_{on}P) * PP_{off}}$$

where,

$$P_{on}P = (P_{peak} - P_{on}),$$

$$PP_{off} = (P_{off} - P_{peak}), \text{ and}$$

$$P_{on}P_{off} = (P_{off} - P_{on})$$

Angle Q:

$$\cos(Q) = \frac{(PQ^2 + QR^2 - RP^2)}{(2 * PQ * QR)}$$

where,

$$PQ = (P_{peak} - Q_{peak}),$$

$$QR = (R_{peak} - Q_{peak}), \text{ and}$$

$$RP = (R_{peak} - P_{peak})$$

Angle R:

$$\cos(QRS) = \frac{(SR^2 + QR^2 - QS^2)}{(2 * SR * QR)}$$

where,

$$SR = (S_{peak} - R_{peak}),$$

$$QR = (R_{peak} - Q_{peak}), \text{ and}$$

$$QS = (S_{peak} - Q_{peak})$$

Angle S:

$$\cos(S) = \frac{(SR^2 + ST^2 - TR^2)}{(2 * SR * ST)}$$

where,

$$SR = (R_{peak} - S_{peak}),$$

$$ST = (T_{peak} - S_{peak}), \text{ and}$$

$$TR = (R_{peak} - T_{peak})$$

Angle T:

$$\cos(T) = \frac{(T_{on}T^2 + TT_{off}^2 - T_{on}T_{off}^2)}{(2 * T_{on}T * TT_{off})}$$

where,

$$T_{on}T = (T_{peak} - T_{on}),$$

$$T_{on}T_{off} = (T_{off} - T_{on}), \text{ and}$$

$$TT_{off} = (T_{off} - T_{peak})$$

$\angle R$ is computed between the directed lines joining from Q_{peak} to R_{peak} and from R_{peak} to S_{peak} fiducials. Similarly, $\angle S$ is computed between lines joining from R_{peak} to S_{peak} and from S_{peak} to T_{peak} fiducials. $\angle T$ is computed between lines joining from T_{on} to T_{peak} and from T_{peak} to T_{off} fiducials. Angle features are shown in Fig. 9.

TABLE III. ANGLE FEATURES ARE SELECTED FROM EACH HEARTBEAT

Angle Features	Representation
Angle P	$\angle P$
Angle Q	$\angle Q$
Angle R	$\angle R$
Angle S	$\angle S$
Angle T	$\angle T$

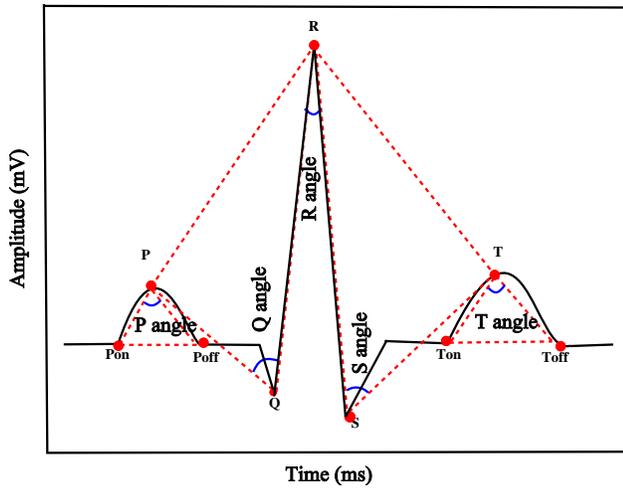


Fig. 9. Angle features class are considered from ECG dominant fiducials

4) *Area under curve method*: Area under curve features are shown in Table IV. We can compute a set of features called area under curve features are formed between the fiducial points of ECG signals are as follows: The area of a triangle where sides have lengths a , b and c can be computed using Heron's formula [27] as follows,

$$\text{Area of triangle} = \sqrt{s(s-a)(s-b)(s-c)},$$

where s is the semiperimeter of the triangle i.e

$$s = \frac{a + b + c}{2}$$

For example, Area of triangle QRS can be computed as,

$$\text{Area of triangle } QRS = \frac{QR_a * QR_I + SR_a * SR_I + SQ_a * SQ_I}{2}$$

where,

$$QR_I = (RS_I - SQ_I),$$

$$SR_I = (SQ_I - QR_I), \text{ and}$$

$$SQ_I = (QR_I - RS_I)$$

where QR_a, SR_a, SQ_a are the amplitude values and QR_I, SR_I, SQ_I are the interval values. It calculates the area of a triangle whose all three vertices are known. Similarly we can compute ten other area under curve features that are $PP_{off}R, PP_{on}P_{off}, PP_{off}Q, P_{off}RQ, RST, RST_{on}, RST_{off}, TT_{on}T_{off}, TT_{on}S$ and PRQ . These features are shown in Fig. 10.

III. FEATURE SELECTION

Feature selection is the process of selecting a subset of relevant features from the feature vector collected from ECG identification model.

The necessity of feature selection for data preprocessing is to facilitate the data management and classification. The purpose is to represent the data in a low-dimensional space that captures the inherent nature of the data. In this paper we

TABLE IV. AREA UNDER CURVE FEATURES ARE SELECTED FROM EACH HEARTBEAT

Area Features	Representation
Area $P_{on}PP_{off}$	area of $\Delta P_{on}PP_{off}$
Area RST_{on}	area of ΔRST_{on}
Area QRS	area of ΔQRS
Area RST	area of ΔRST
Area $T_{on}TT_{off}$	area of $\Delta T_{on}TT_{off}$
Area $PP_{off}R$	area of $\Delta PP_{off}R$
Area $P_{off}RQ$	area of $\Delta P_{off}RQ$
Area $PP_{off}Q$	area of $\Delta PP_{off}Q$
Area QST	area of ΔQST
Area STT_{off}	area of ΔSTT_{off}
Area $RT_{on}T$	area of $\Delta RT_{on}T$

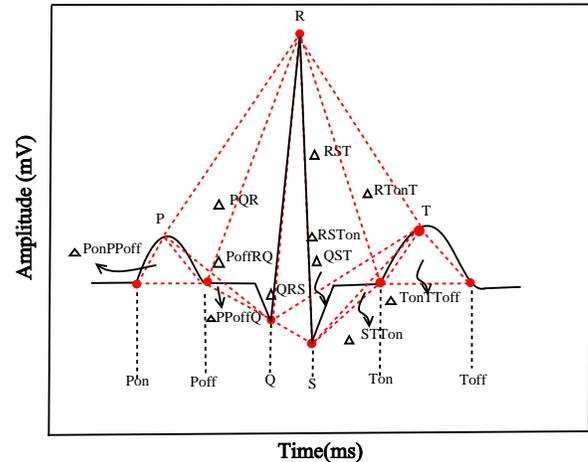


Fig. 10. Selected class of area under curve features from ECG dominant fiducials.

use Kernel principal component analysis. Principal component analysis gives a linear projection that best fits the data set in the least-square manner and due to its computational and analytical simplicity it is widely used [28]. Eigenvector is a well known application of principal component analysis for biometric recognition. Most current literature focus on the limitation of linear principal component analysis specifically, its ability to capture the nonlinear structure of the complex data set that is beyond second order statistics [28]. The nonlinear technique has been proposed that is Kernel principal component analysis (KPCA) [28].

1) *Kernel principal component analysis*: The kernel substitution method is applied in principal component analysis to get a nonlinear generalization solution that is known as kernel principal component analysis [19]. Suppose a nonlinear transformation $\phi(z)$ into an M -dimensional feature space from D -dimensional feature space, where $M > D$ and D is the original dimension of feature space. Projecting each data point (z_i) onto a point $\phi(z_i)$ [30].

Assume that the projected data set has zero mean that is

$$\frac{1}{N} \sum_{i=1}^N \phi(z_i) = 0 \quad (14)$$

The covariance matrix of projected feature space in $M \times M$ sample is defined as

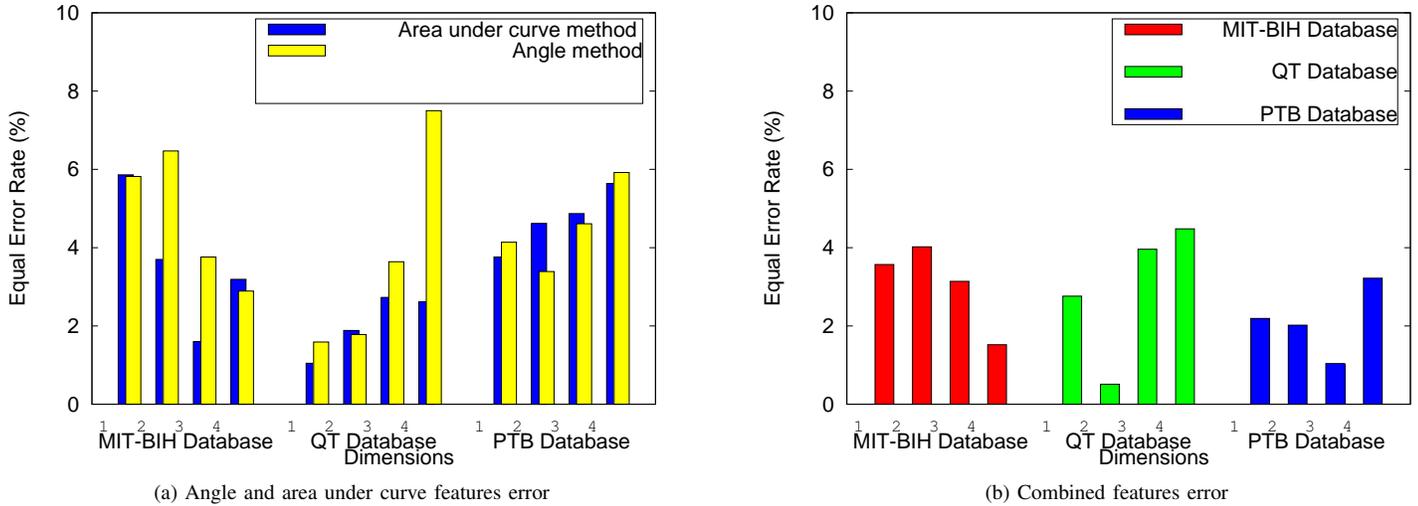


Fig. 11. The error rates on different dimensions on different databases (a) Angle and area under curve features error are evaluated on MIT-BIH , QT and PTB database, (b) Combined features error(i.e amplitude, interval, angle and area under curve features) are evaluated on MIT-BIH , QT and PTB database.

$$C = \frac{1}{N} \sum_{i=1}^N \phi(z_i) \phi(z_i)^T \quad (15)$$

The eigenvectors and eigenvalues are given by

$$Cv_n = \lambda_n v_n \quad (16)$$

where $n = 1, \dots, M$, putting the value of Eq. 15 in Eq. 16 we get eigenvector

$$C = \frac{1}{N} \sum_{i=1}^N \phi(z_i) \{ \phi(z_i)^T v_n \} = \lambda_n v_n \quad (17)$$

by simplifying we get

$$v_n = \sum_{i=1}^N a_{ni} \phi(z_i) \quad (18)$$

By substituting the value of v_n from Eq. 18 to Eq. 17,

$$\frac{1}{N} \sum_{i=1}^N \phi(z_i) \phi(z_i)^T \sum_{j=1}^N a_{nj} \phi(z_j) = \lambda_n \sum_{i=1}^N a_{ni} \phi(z_i) \quad (19)$$

The kernel function is defined as,

$$\kappa(z_i, z_j) = \phi(z_i)^T \phi(z_j) \quad (20)$$

By multiply $\phi(z_i)^T$ on both sides of Eq.19 to give

$$\frac{1}{N} \sum_{i=1}^N \kappa(z_i, z_i) \sum_{j=1}^N a_{nj} \kappa(z_i, z_j) = \lambda_n \sum_{i=1}^N a_{ni} \kappa(z_i, z_i) \quad (21)$$

The above equation is written in matrix notation are as follows

$$K^2 A_k = \lambda_n N K A_k \quad (22)$$

where A_k is N-dimensional column vector of A_k : $A_k = [A_{k1} A_{k2} \dots A_{kN}]^T$ simplifying the above equation

$$K A_k = \lambda_n N A_k \quad (23)$$

The point z is projected onto eigenvector n , the principal component of the projection in terms of kernel function, is represented as

$$y_n(z) = \phi(z)^T v_n = \sum_{i=1}^N a_{ni} \kappa(z, z_i) \quad (24)$$

The case when projected data set $\phi(z_i)$ does not have zero mean. So, we cannot simply compute and then subtract off the mean, since we formulate the algorithm in term of the kernel function. We calculate the Gram matrix \tilde{K} by substituting the kernel function K . The Gram matrix is represented as

$$\tilde{K} = K + 1_N K 1_N - K 1_N - 1_N K \quad (25)$$

where 1_N is $N \times N$ matrix in which every element has $1/N$ value, \tilde{K} is the kernel function by which we can calculate eigenvectors and eigenvalues. The linear kernel is

$$\kappa(z, z') = z^T z'$$

The Gaussian kernel is

$$\kappa(z, z') = \exp(-\|z - z'\|^2 / 2\sigma^2)$$

The performance of the proposed ECG biometric system is determined by authentication processes on different feature detection method using equal error rate (EER). The EER can be calculated by genuine acceptance rate and false acceptance rate. The EER result shows on kernel PCA feature selection techniques on detecting feature vector and in different database are shown in Fig. 11. In MIT-BIH arrhythmia database [31] while using area under curve feature detection method, the EER at dimensions 1, 2, 3 and 4 are reported as 5.86%, 3.7%, 1.6% and 3.19%, respectively. The EER results for angle feature vector of dimensions 1, 2, 3 and 4 are found to be 5.82%,

6.47%, 3.76% and 2.89%, respectively. On QT database by

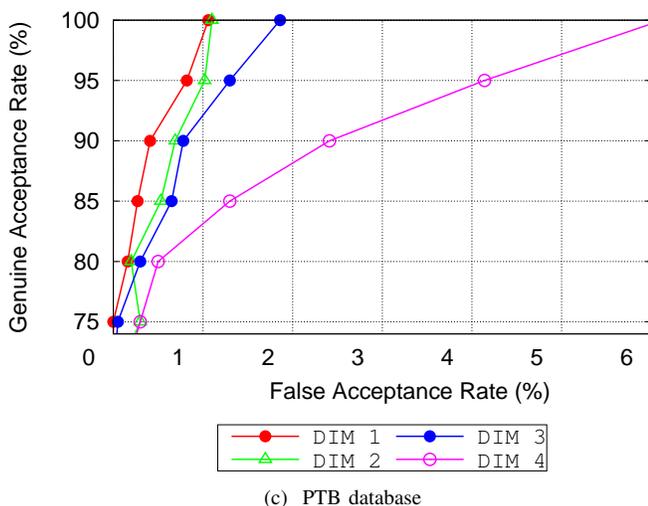
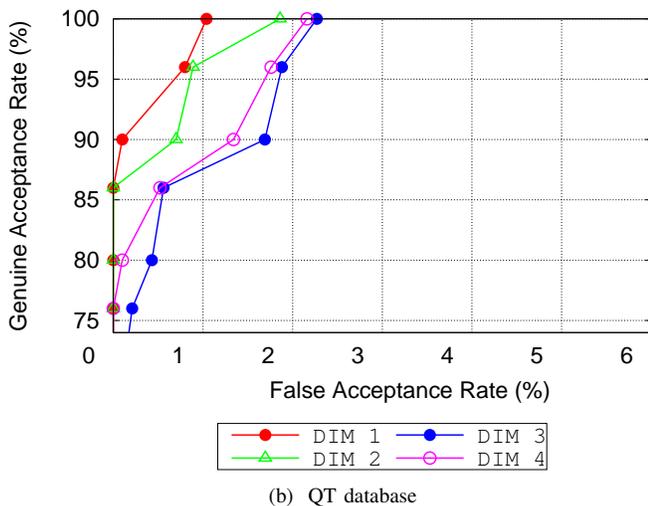
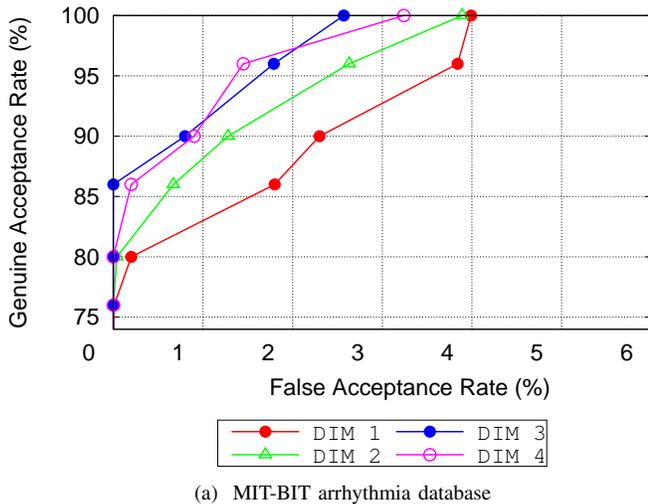


Fig. 12. ROC curve of area under curve features on different dimensions and on different database (a) MIT-BIT arrhythmia database, (b) QT database (c) PTB database.

using area under curve feature vector, the EER are reported to be 1.047%, 1.884%, 2.729% and 2.615% respectively, at dimensions 1, 2, 3 and 4. At the same dimensions the EER using angle feature vector are reported as 1.59%, 1.78%, 3.64% and 7.5%. In PTB database the EER using area under curve feature vector are reported to be 3.76%, 4.62%, 4.87% and 5.64% respectively, at dimensions 1, 2, 3 and 4, whereas on same dimension the EER using angle feature vector are reported as 4.14%, 3.39%, 4.61% and 5.92% respectively.

The equal error rate for the proposed biometric system by combing the interval features, amplitude features, angle features and area under curve features using kernel PCA transformation techniques on MIT-BIH arrhythmia database, QT database and PTB database are shown in Fig. 11. In MIT-BIH arrhythmia database, the EER at dimensions 1, 2, 3 and 4 is 3.57%, 4.02%, 3.14% and 1.52%, respectively. The QT database reported the EER 2.76%, 0.51%, 3.96% and 4.48% respectively, at dimensions 1, 2, 3 and 4. The PTB database reported the EER 2.19%, 2.02%, 1.04% and 3.22% respectively, at dimensions 1, 2, 3 and 4. The proposed biometric system performs better on the QT database as shown in Fig. 11. The best results are reported to be 0.51% at dimension 2 on QT database, 1.04% at dimension 3 on PTB database and 1.52% at dimension 3 on the MIT-BIH arrhythmia database, respectively

The results of error rates often summarized in receiver operating characteristic (ROC) curve. The ROC curve between FAR and GAR for different dimensions using kernel PCA and feature extraction techniques are shown in Fig. 12 and Fig. 13, respectively, for MIT-BIH arrhythmia database, QT database and PTB database. The ROC curve for the area under curve feature vector in MIT-BIH arrhythmia database is shown in Fig. 12a. It can be seen that the GAR is found to be 76%, 73%, 70% and 68% at dimensions 1, 2, 3 and 4, respectively when FAR is zero. The 100% GAR can be achieved with the FAR value of 3.99%, 3.89%, 2.57%, and 3.24% at dimensions 1, 2, 3 and 4, respectively. The ROC curve for area under curve feature vector on the QT database in Fig. 12b, shows that 100% GAR can be achieved for the FAR of 1.04%, 1.86%, 2.27% and 2.16% at dimensions 1, 2, 3 and 4, respectively. Whereas 0% FAR, the GAR is found to be 86%, 87%, 68% and 77% at dimensions 1, 2, 3 and 4, respectively. The ROC curve for area under curve feature vectors on the PTB database in Fig. 12c, shows that 100% GAR can be achieved for the FAR of 1.06%, 1.1%, 1.86% and 6.12% at dimensions 1, 2, 3 and 4, respectively. Whereas without allowing false acceptance ie. 0% FAR, the GAR is found to be 75%, 72%, 70% and 68% at dimensions 1, 2, 3 and 4, respectively.

The ROC curve for the angle feature vectors on MIT-BIH arrhythmia database is shown in Fig. 13a. It can be seen that the GAR is found to be 70%, 65%, 70% and 78% at dimensions 1, 2, 3 and 4, respectively when FAR is zero. The 100% GAR can be achieved with the FAR value of 3.21%, 3.96%, 2.17% and 2.0% at dimensions 1, 2, 3 and 4, respectively. The ROC curve for angle feature vectors on the QT database in Fig. 13b, shows that 100% GAR can be achieved for the FAR of 1.65%, 1.68%, 2.9% and 2.75% at dimensions 1, 2, 3 and 4, respectively. Whereas without allowing false acceptance ie. 0% FAR, the GAR is found to be 78%, 75%, 76% and 71% at dimensions 1, 2, 3 and

4, respectively. The ROC curve for angle feature vectors on the PTB database in Fig. 12c, shows that 100% GAR can be achieved for the FAR of 1.29%, 0.96%, 1.88% and 2.5% at dimensions 1, 2, 3 and 4, respectively. Whereas 0% FAR, the GAR is found to be 75%, 80%, 70% and 66% at dimensions 1, 2, 3 and 4, respectively.

The ROC curve for combining the interval features, amplitude feature, angle features and area under curve features on MIT-BIH arrhythmia database is shown in Fig. 14a. It can be seen that the GAR is found to be 71%, 68%, 69% and 80% at dimensions 1, 2, 3 and 4, respectively when FAR is zero. The 100% GAR can be achieved with the FAR value of 4.4%, 4.86%, 3.23% and 1.65% at dimensions 1, 2, 3 and 4, respectively. The ROC curve for interval features, amplitude feature, angle features and area under curve features on QT database in Fig. 14b, shows that 100% GAR can be achieved for the FAR of 1.9%, 0.56%, 2.86% and 3.56% at dimensions 1, 2, 3 and 4, respectively. Whereas without allowing false acceptance ie. 0% FAR, the GAR is found to be 80%, 86%, 76% and 72% at dimensions 1, 2, 3 and 4, respectively. The ROC curve for interval features, amplitude feature, angle features and area under curve features on PTB database in Fig. 14c, shows that 100% GAR can be achieved for the FAR of , 0.9%, 0.36%, 0.13% and 1.86% at dimensions 1, 2, 3 and 4, respectively. Whereas without allowing false acceptance ie. 0% FAR, the GAR is found to be 76%, 76%, 90% and 60% at dimensions 1, 2, 3 and 4, respectively.

A. Comparison of Present ECG Biometric Systems

The comparison of ECG biometric system for proposed and existing method with the help of the authentication process and identification process as shown in Table V. The proposed method performs the other existing methods of ECG biometric authentication in terms of EER and size of samples. The experimental results show the robustness of the proposed method over a larger data set and produces better EER than the existing method of [8], [2] and [33]. Although the EER of the proposed method is greater than the method of [8], [2] and [33], but tested on a larger data set.

TABLE V. COMPARISON OF EXISTING METHOD WITH PROPOSED METHOD ON THE BASIS OF FIDUCIAL POINTS

Method	Identification rate	Equal error rate(%)	Sample size
Wubbeler et. al [8]	98%	2.8	74
Singh and Gupta [2]	99%	0.76	85
Odinaka et. al [33]	99%	0.37	269
Biel et. al [6]	100%	-	20
Kyoso et. al [35]	> 90%	-	9
Shen et. al [11]	100%	-	20
Irvine et. al [36]	91%	0.01	104
Palaniappan and Krishnun [32]	97.6%	-	10
Israel et. al [9]	100%	-	10
Shen et. al [10]	95.3%	-	29
Wang and Zhang [34]	97.4%	-	168
Silva et. al [37]	99.97%	-	520
Proposed method	99.26%	0.13	38
	99.5%	0.56	30
	98.5%	1.65	248

IV. SPORT/EXERCISE ISSUES

The big changes in the heart beat of individual subject after exercise or sport, the individual verification and identification can disturb the ECG signal [38]. The main challenges are as follows:

- 1 Baseline shifted due to deeper breath.
- 2 The heart rate become high because of heart activity.

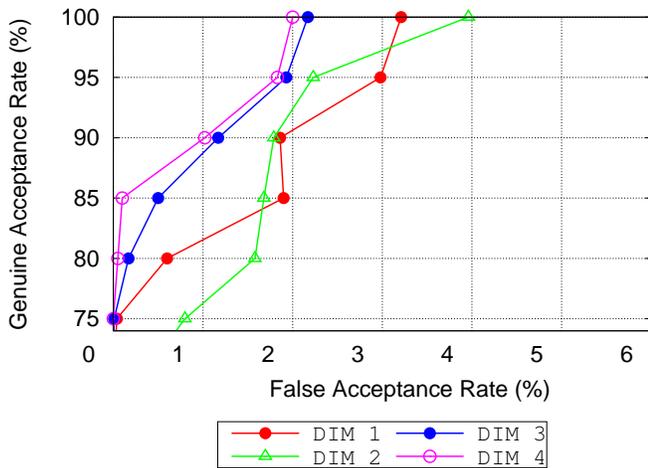
By measuring the ECG signal after exercise or sports the heart rate increases up to 45% to 55% in comparing to resting state. This will create a problem in ECG biometric identification.

V. CONCLUSION

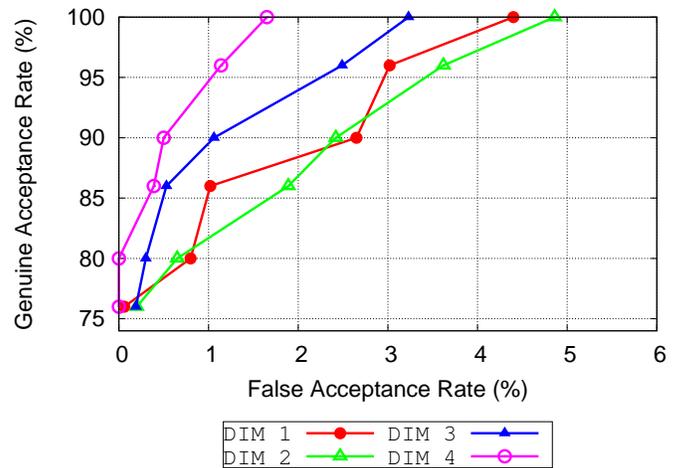
This examination has proposed a technique for biometric recognition of individuals using their heartbeats. The method has delineated the dominant fiducials of ECG waveform and after that interval, amplitude, angle and area under curve features are computed. The recognition results also demonstrated that the proposed technique for ECG biometric recognition differentiates the heartbeats of normal and the impatient subjects. All human beings have a heart, every individual has its own unique set of heart beat features.

Finally, the proposed strategies can be utilized as a potential biometric for human recognition, which is very secure and robust from falsification. The ECG biometric based on fiducial points and advantage of various different types of morphological features. That are interval, amplitude, angle and area under curve features are significantly different between individual subjects. These fiducial points have their onset and offset. With the help of these features we reduce the dimension of feature vector by kernel principal component analysis. The Euclidean distance that calculates the matching score shows better results for authentication as well as identification on the QT database by combining all the features. It reports 0.51%, 1.04% and 1.52% error rate authentication and 99.49%, 98.96% and 98.48% classification accuracy for identification on three different databases i.e QT database, PTB database and MIT-BIH arrhythmia database. It also differentiates healthy subjects as well as impatient subjects on the basis of heart beat and deducted feature vector. The result shows that the proposed method works well on QT database in comparison to MIT-BIH arrhythmia database and PTB database.

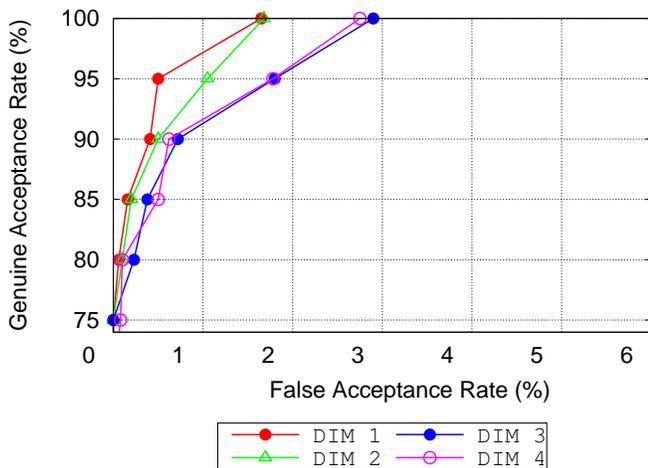
In the future try to automatically diagnose the arrhythmia diseases and find the respective medicines for unhealthy individual.



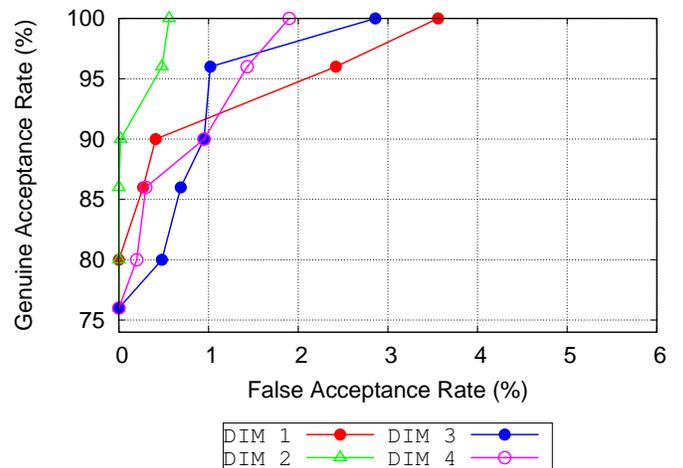
(a) MIT-BIT arrhythmia database



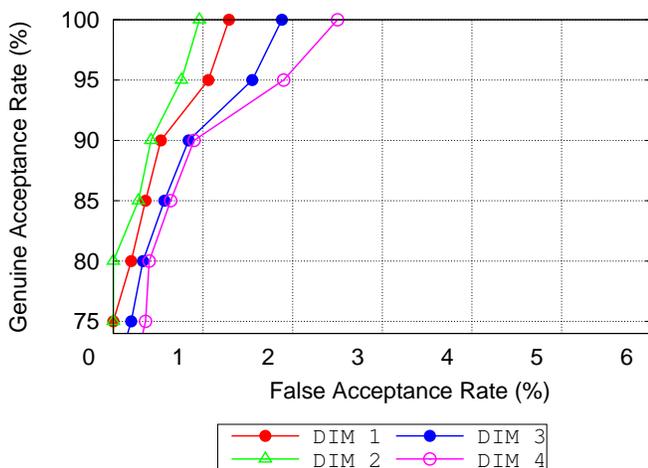
(a) MIT-BIH arrhythmia database



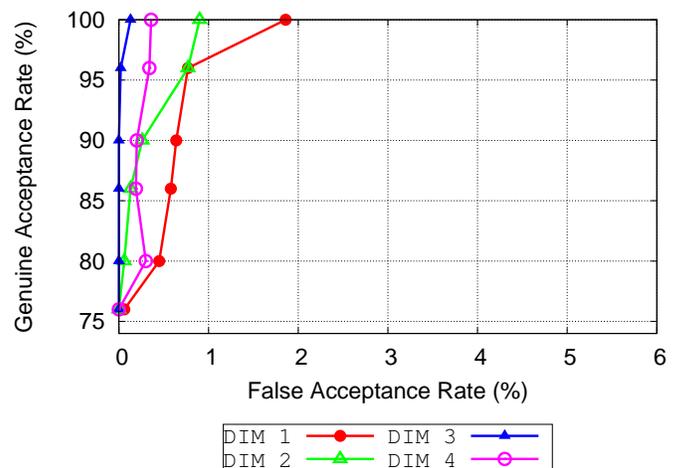
(b) QT database



(b) QT database



(c) PTB database



(c) PTB database

Fig. 13. ROC curve of angle features on different dimensions and on different database (a) MIT-BIT arrhythmia database, (b) QT database (c) PTB database.

Fig. 14. ROC curve of combined features(i.e amplitude, interval, angle and area under curve features) on different dimensions and on different database (a) MIT-BIT arrhythmia database, (b) QT database (c) PTB database.

REFERENCES

- [1] Y. N. Singh. Human recognition using fishers discriminant analysis of heartbeat interval features and ECG morphology. *Neurocomputing*, 2015, pp. 322–335.
- [2] Y. N. Singh and P. Gupta. ECG to individual identification. *Proceedings 2nd IEEE International Conference on Biometrics Theory, Applications and Systems*, 2008, pp. 1–8.
- [3] Y.N. Singh, P. Gupta. Biometric method for human identification using electrocardiogram. *Lecture notes of computer science, Springer-Verlag* **5558**, 2009, pp. 1270–1279.
- [4] ECG library, <https://ecglibrary.com/ecghist.html>, accessed on July 2017.
- [5] J. G. Webster . Medical Instrumentation: Application and Design. *John Wiley and Sons, Philadelphia, USA*, 1997.
- [6] L. Biel, O. Pettersson, L. Philipson, and P. Wide. ECG analysis: A new approach in human identification. *IEEE Transactions Instrument and Measurement* **50**, 3, 2001, pp. 808–812.
- [7] M.Yochum, C. Renaud and S. Jacquir. Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT. *Biomedical Signal Processing and Control, Elsevier*, 2016, pp. 1–8.
- [8] G. Wubbeler, M. Stavridis, D. Kreiseler, R. Bousseljot and C. Elster. Verification of humans using the Electrocardiogram. *Pattern Recognition Letters* **28(10)**, 2007, pp. 1172–1175.
- [9] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold. ECG to identify individuals. *Pattern Recognition* **38**, 2005, pp. 133–142.
- [10] T. W. D. Shen, W. J. Tompkins, Y. H. Hu. Implementation of a one-lead ECG human identification system on a normal population. *J. Eng. Comput. Innovations*, 2011, pp. 12–21.
- [11] A.D.C. Chan, M.M. Hamdy, A. Badre. Wavelet distance measure for person identification using Electrocardiograms. *IEEE Trans. Instrument*, 2008, pp. 248–253.
- [12] Y.N. Singh, P. Gupta. A robust delineation approach of Electrocardiographic P Waves. In *Proc. 2009 IEEE Symposium on Industrial Electronics and Applications (ISIEA)* **2**, 2009, pp. 846–849.
- [13] Y.N. Singh, P. Gupta. A robust and efficient technique of T wave delineation from electrocardiogram. In *Proc. of Second Internat. Conf. on Bioinspired Systems and Signal Processing (BIOSIGNALS), IEEE-EMB*, 2009, pp. 146–154.
- [14] S.K.Mitra. Digital Signal Processing. *Tata McGraw-Hill Publishing Co. Ltd, New Delhi*, 1998.
- [15] MATLAB software, <http://in.mathworks.com/help/signal/ref/fdesign.highpass.html>, accessed on Mar 2016.
- [16] Y. Luo, R. H. Hargraves, O. Bai, K. R. Ward. A Hierarchical Method for Removal of Baseline Drift from Biomedical Signals Application in ECG Analysis. *The Scientific World Journal* **2013**, pp. 1-2.
- [17] MATLAB software, <http://in.mathworks.com/help/signal/ref/fdesign.lowpass.html>, accessed on Mar 2016.
- [18] R. S. Stankovic, B. J. Falkowski. The Haar wavelet transform: its status and achievements. *Computers and Electrical Engineering*, 2003, pp. 25–44.
- [19] B.Scholkopf , A.Smola , KR. Muller. Kernel principal component analysis. In *Gerstner W., Germond A., Hasler M., Nicoud JD. (eds) Artificial Neural Networks, ICANN 1997. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg* **1327**, 1997.
- [20] Maths open Reference, <http://in.mathworks.com/help/wavelet/ref/wavedec.html>, accessed on June 2017.
- [21] Y. N. Singh and S. K. Singh. Identifying individuals using Eigenbeat features of Electrocardiogram. In *Journal of Engineering*, 2013, pp. 1–8.
- [22] Y. N. Singh and S. K. Singh. The state of information security. In *Proceedings of the Artificial Intelligence and Agents: Theory and Applications*, 2011, pp. 363–367 .
- [23] Y. N. Singh and S. K. Singh. A taxonomy of biometric system vulnerabilities and defences. *Journal of Biometrics* **5**, 2013, pp. 137–159.
- [24] Y. N. Singh. Challenges of UID environment. In *Proceedings of the UID National Conference on Impact of Aadhaar in Governance*, 2011, pp. 37–45.
- [25] Y.N. Singh, P. Gupta. Correlation based classification of heartbeats for individual identification. *J. Soft Comput.* **15**, 2011, pp. 449–460.
- [26] Maths open Reference, <http://www.mathopenref.com/lawof-cosines.html>, accessed on June 2017.
- [27] Maths open Reference, <http://www.mathopenref.com/coordtrianglearea.html>, accessed on June 2017.
- [28] Richard O. Duda and Hart, Peter E. and Stork, David G. Pattern Classification (2Nd Edition). *Wiley-Interscience*, 2000.
- [29] Irena Jekova and Giovanni Bortolan. Personal Verification/Identification via Analysis of the Peripheral ECG Leads: Influence of the Personal Health Status on the Accuracy. *Hindawi Publication Corporation, BioMed Research International*, 2015, pp. 1–13.
- [30] Christopher M. Bishop. Pattern Recognition and Machine Learning (Information Science and Statistics). *Springer-Verlag New York, Inc., Secaucus, NJ, USA*, 2006.
- [31] Physionet, "Physiobank Archives", Physikalisch-Technische Bundesanstalt, Abbestrasse 2-12, 10587 Berlin, Germany, <https://www.physionet.org/physiobank/database/ptbdb>, accessed on 2016.
- [32] R. Palaniappan and S. M. Krishnun. Identifying individuals using ECG beats. *Proceedings International Conference on Signal Processing and Communications*, 2004, pp. 569–572.
- [33] I. Odina, P.-H. Lai, A. D. Kaplan, J. A. O. Sullivan, E. J. Sirevaag, S. D. Kristjansson, A. K. Sheffield, J. W. Rohrbaugh. ECG biometrics: A robust short-time frequency analysis. *Proceedings 2010 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2010, pp. 1–6.
- [34] Z.Wang and Y.Zhang. Research on ECG Biometric in Cardiac Irregularity Condition. *Proceedings of IEEE International Conference on Medical Biometrics(ICMB)*, 2014, pp. 157–163.
- [35] M. Kyoso, A. Uchiyama. Development of an ECG identification system. *Proc. 23rd Ann. EMBS Int. Conf., Istanbul, Turkey*, 2001.
- [36] J. Irvine, S. Israel, M. Wiederhold. A new biometric: Human identification from circulatory function. *Proc. Joint Statistical Meetings of the American Statistical Association, San Francisco, CA*, 2003.
- [37] H. Silva, A. Fred, A. Lourenco. Finger ECG signal for user authentication: Usability and performance. *IEEE Sixth International conference on Biometrics: Theory, Applications and Systems BTAS, USA*, 2013.
- [38] Kuo-Kun Tseng, Fufu Zeng, W. H. Ip, and C. H. Wu. ECG Sensor Verification System with Mean-Interval Algorithm for Handling Sport Issue. *Journal of Sensors*, 2016.