

Machine Learning Classification Approaches for Biometric Recognition System using ECG Signals

Kiran Kumar Patro* and P.Rajesh Kumar

Andhra University, Department of Electronics & Communication Engineering Visakhapatnam, A.P, India

Received 5 October 2017; Accepted 7 December 2017

Abstract

Person identification is an essential task in defense and forensic applications such as surveillance and criminal investigation activities. Studies conducted in the past included different types of regular biometric traits, namely fingerprint, face, iris, and voice have a limitation of falsification, and they do not fit guarantee of liveness of the subject. In this context, Electrocardiogram based Biometric recognition is an alternative solution, where the security of the information is very high level. This research aims to provide with a complete systematic approach to ECG based Biometric recognition, which contains primarily the processing of raw signal through noise elimination filters and a time domain analysis is carried for all ECG characteristic waves detection. Subsequently, an effective feature extraction method for ECG is developed, which extracts six best P-QRS-T fragments based on priority and their positions are normalized. Also, 72-time domain features are calculated. These features are formed into feature vector corresponding to each signal separately for both train and test data sets. To analyze the performance of the system, the feature vectors are trained with various Machine learning classification algorithms like Artificial Neural Network (ANN), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN). Finally, the proposed system is tested with a challenging, public available MIT-BIH ECG-ID Database. A comparative analysis using performance parameters is made with different classifiers, and the obtained results show that SVM classifier provides 93.709% overall classification accuracy when compared to previous literature.

Keywords: Electrocardiogram (ECG); Biometric; Machine Learning; Pattern Recognition; Support Vector Machine (SVM); K-Nearest Neighbor (KNN); MIT-BIH ECG-ID

1. Introduction

In the current day's world, biometric recognition is undoubtedly an essential field of information and security. Generally, biometric characteristics are divided into two main categories: behavioural, and physiological [1]. Physiological characteristics are associated with the shape of the body. Few examples of physiological based biometrics are fingerprint, face recognition, DNA, hand and palm geometry, and iris recognition. Behavioural based biometrics depends on the behaviour of a person. Examples include typing rhythm, gait, and voice. However, there exists a chance for these systems to get falsified, stolen and forged. For instance face recognition can be fooled by a picture, fingerprints can be recreated and voice can be imitated or pre-recorded [2-3]. Furthermore, Security systems being used currently rely on passwords or pins, which can be easily shared, observed, forgotten or stolen, for authentication [4].

• ECG as Biometric

Electrocardiography (ECG) is the process of recording the electrical activity of the heart over a period of time using electrodes placed on the skin, which are regularly used for

medical purposes. Currently, ECG is not only being used as a diagnostic tool for clinical purposes, but also as a new biometric tool for high level security systems [4].

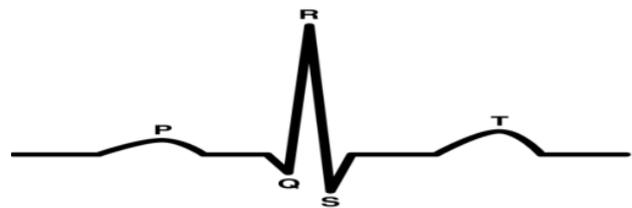


Fig. 1. ECG beat

Figure 1 shows a sample ECG beat with its characteristic points. The electrodes detect the tiny electrical charges on the skin that arise from the heart muscles' electrophysiological pattern of depolarizing during each heartbeat [5]. Compared to traditional biometrics, ECG obtains a more reliable biometric solution and it satisfies most of the conditions required to become an ideal biometric. ECG signals of every individual contain a unique pattern due to the differences in morphology or the change in the structure of the heart among various individuals. ECG signal is a life indicator, and it can be used as a tool for detecting life. It also has various significant characteristics to substantiate its use as a biometric system, such as universality, uniqueness, permanence, collectability, circumvention, etc. Unlike other Biometric technologies, ECG has the inherent property of life detection for identification [6].

*E-mail address: kirankumarpatro446@gmail.com

ISSN: 1791-2377 © 2017 Eastern Macedonia and Thrace Institute of Technology. All rights reserved.
doi:10.25103/jestr.106.01

In the recent past, many studies have been used for ECG based biometric recognition systems [7-11]. In the literature, different techniques have been used for ECG biometrics. Initially, Biel et al. [12] conducted biometric experiments on ECG in which 12 features with a specific hardware were proposed. The proposed methodology was tested by using 3-lead ECG data from 20 subjects at rest. Principal Component Analysis with SIMCA (Soft Independent Modeling of Class Analogy) model for the classification was adopted and finally 95% recognition rate was reported.

Md Najam Dar et al. [13] proposed biometric identification of the general population using multi resolution analysis of DWT based features. Feature extraction was performed by the extraction of cardiac cycle followed by discrete wavelet transform (DWT) to extract wavelet coefficients as a feature vector. Classification was performed using single nearest neighbour classifier and this method reported an accuracy of 93.1% with ECG ID database. Chen et al. [14] proposed a chaotic feature extraction along with back propagation neural network (BPNN) for biometric identification using lead-I ECG of 19 subjects and reported an identification rate of 91%. Boumbarov et al. [15] used linear space projection with Principle Component Analysis (PCA) along with Radial Basis Function Neural Network (RBFNN) classifier for biometric identification with 9 subjects and reported an accuracy of 86.1%. X.Tang et al. [16] used Quantum neural network classifier for identification with MIT-BIH arrhythmia open source database and achieved a recognition rate of 91.7%.

This paper proposes an effective framework for the identification of an individual with the analysis of lead-I ECG. The proposed system adopts a robust cascaded digital filters configuration for noise elimination so that the raw ECG signal can directly be applied to the system. A new approach for the detection of characteristic points of ECG has been employed. Furthermore, an effective feature extraction method has been developed to extract fiducial based features for identification. Finally, the obtained features have been trained with the best suitable higher order classifiers for recognition. The entire individual stages have been combined to strengthen the performance of the biometric system which overcomes the drawbacks of existing literature.

The rest of the paper is outlined as follows: Section-2 describes the detailed methodology of stages for the biometric recognition system. Section-3 discusses the application of different classification algorithms and identification results. Finally, Section-4 includes the final conclusion.

2. Methodology

The proposed system includes the following processes: ECG data acquisition, preprocessing, P-QRS-T detection, feature extraction, cardiac cycle classification and finally, ECG signal recognition. Figure 2 indicates the methodology of an ECG biometric recognition system.

• ECG data Acquisition

Phsiobank contains over 90,000 recordings and the website is supported by National Institute of General Medical and Sciences (NIGMS) and National Institute of Biomedical Imaging and Bioengineering (NIBIB). MIT-BIH ECG ID database was used with Lead-I configuration from 44 men

and 46 women. Records were obtained from persons aged between 13 and 75 years. Each recording was recorded for 20 seconds, digitized at 500 Hz, with 12 bit resolution over a nominal ± 10 mV range. The raw signal was extracted and collected over a period of 6 months. The proposed methodology was tested on 20 subjects with five records each [17].

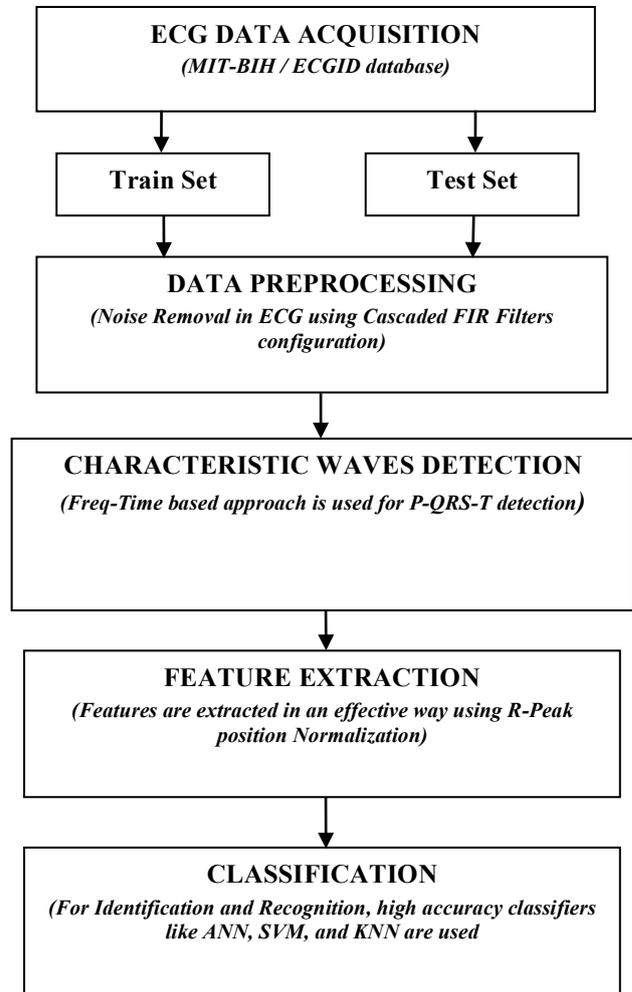


Fig. 2. Methodology of an ECG Biometric Recognition System

• Data Preprocessing

The raw ECG is rather noisy and contains distortions of various origins. The different frequency components used in acquiring the ECG signal may produce interference, and such interference can add noise to the ECG signal. This added noise may tamper with the original information and lead to false ECG data. For each possible case, corruption in real ECG signal made by the introduction of unwanted noise filtering methods has to be used. An appropriate filter has to be selected on the basis of the possible type of noise present in the signal. For suppressing three major noises of baseline drift, power line Interference and EMG noise, a Cascaded digital filters configuration is used [18]. The cascaded filter structure is designed using an FIR High pass filter, an FIR Band stop filter and an FIR Low pass filter in addition to Kaiser window as shown in Figure 3.

• Characteristic Waves detection

The electrical activity during the cardiac cycle is characterized by five separate characteristic waves that are P, Q, R, S and T. For identifying characteristic waves, Frequency-Time based approach is adopted. In this method,

primarily the R-peak is detected using wavelet decomposition (Symlet) method and other peaks are detected using the R-peak locations as a reference and thereby creating time based adaptive windows [19]. All the detected peaks of an ECG signal are shown in Figure 4.

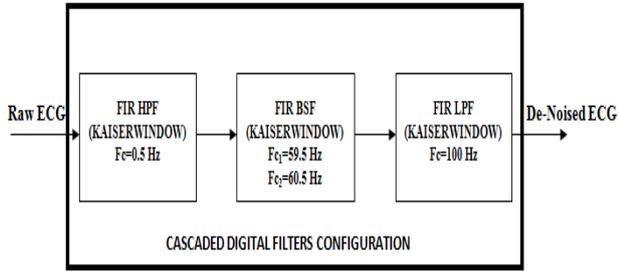


Fig. 3. Cascaded Filter configuration [18]

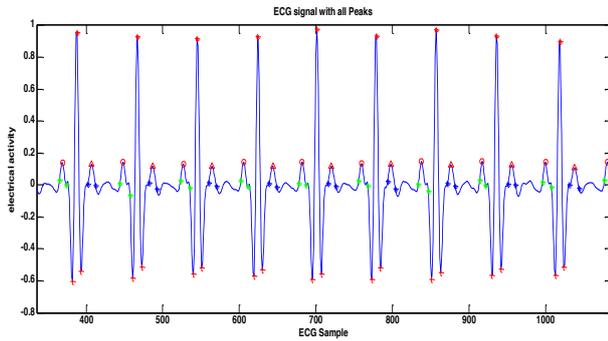


Fig. 4. ECG signal with Characteristic waves

• Feature Extraction

The Majority of the information in the ECG is originated in the intervals and amplitudes defined by its features. The biometric application of ECG mainly depends on the nature of its features. After detecting all the peaks of P, Q, R, S and T in an electrocardiogram signal, it is identified that QRS, ST and especially QT intervals change according to heart rate and physiology. So to extract appropriate features for biometric system, the finest cardiac impulses must be detected, out of all PQRST fragments. The selection of best fragments is based on priorities of QRS area, QT distance and ST distance. These priorities are considered under standard rules of normal ECG ranges [5]. Out of all PQRST fragments, the top six finest fragments are selected on the basis of their difference (the lower, the better) from the mean value of the individual priorities [20-21]. Only these 6 best fragments are used for further analysis.

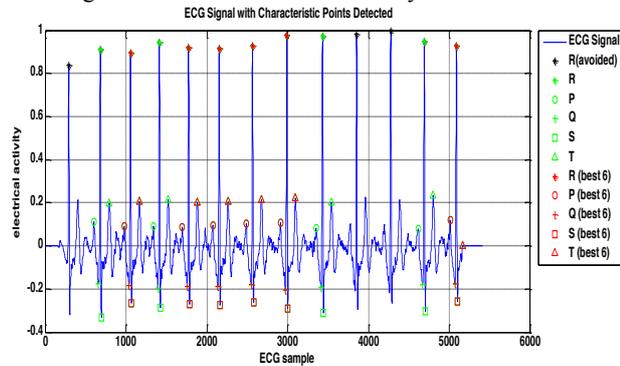


Fig. 5. ECG signal with peaks detected

The best 6 PQRST fragments of the ECG signal selected are broken into independent pulses so that the data in each

pulse can be compared with that of the other pulses. Each pulse was broken ranging from peak 'P' to peak 'T' (Beat). Selected best peaks are marked with 'red' in Figure 5.

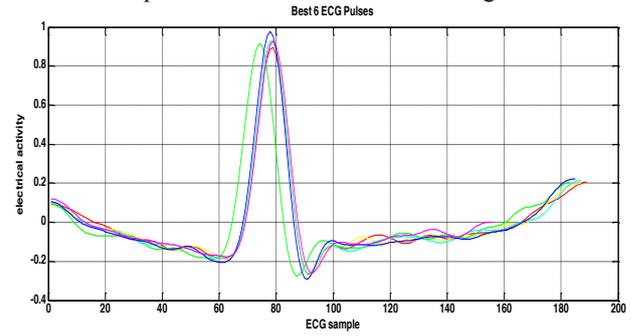


Fig. 6. ECG pulse of best six PQRST fragment

The broken pulse (PQRST beat) in Figure 6 is still not appropriate enough to compare data. Before extracting data from such pulses, their position has to be normalized taking a standard origin point common to all of them as shown in Figure 7. Thus, R Peaks of all such fragments were assumed to be at the origin and the positions of their respective P, Q, S and T peaks were translated to new coordinates keeping their original aspect ratio of position unchanged.

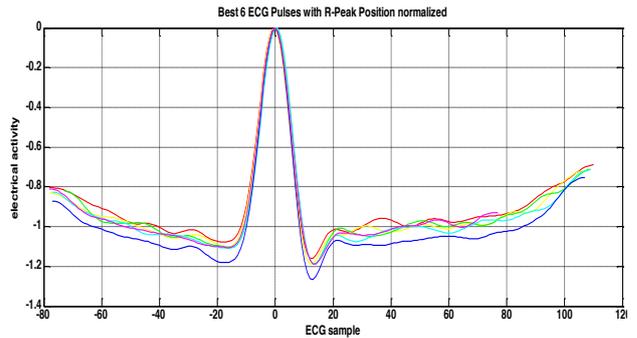


Fig. 7. ECG pulse of best six PQRST fragments with the position normalized with respect to R.

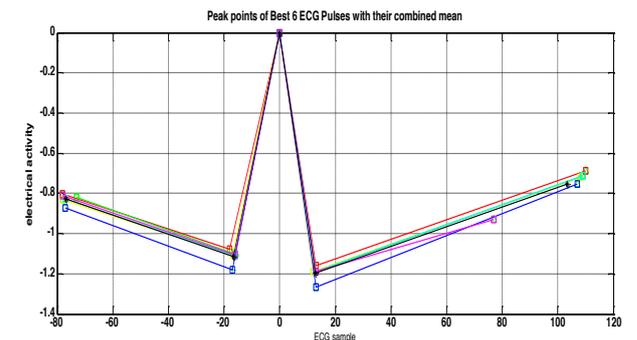


Fig. 8. Best six PQRST fragment with their combined mean of ECG pulses

Figure 8 shows the ECG pulse of best 6 PQRST fragments with their combined mean. All the amplitudes and time durations were calculated taking the position of R as origin. Based on the relation between time and amplitude of the peaks of ECG signal, a total of 72 features were extracted [21-22] which are shown in Table I. In this work, all the 72 features were calculated separately for 7 pulses, that is 6 best PQRST fragments along with their mean pulse of each ECG record.

Table 1 The set of features (Time, Amplitude) extracted with respect to the position of ‘R’ as origin

Features	Feature Description	Label	Features	Feature Description	Label		
Time Features	Time difference of P w.r.t. R (X-axis)	Px	Distance Features	Distance between P and Q	PQdis		
	Time difference of Q w.r.t. R (X-axis)	Qx		Distance between Q and R	QRdis		
	Time difference of S w.r.t. R (X-axis)	Sx		Distance between R and S	RSdis		
	Time difference of T w.r.t. R (X-axis)	Tx		Distance between S and T	STdis		
	Time difference between P and Q	PQtime		Distance between Q and S	QSdis		
	Time difference between P and T	PTtime		Distance between P and R	PRdis		
	Time difference between Q and S	QStime		Ratio of STdis to QSdis	(ST/QS)dis		
	Time difference between Q and T	QTtime		Ratio of RSdis to QRdis	(RS/QR)dis		
	Time difference between S and T	STtime		Slope Features	Slope of line joining P and Q	PQslope	
	Time difference between P and S	PStime			Slope of line joining Q and R	QRSlope	
	Ratio of PTtime to QStime	(PT/QS)time	Slope of line joining R and S		RSslope		
	Ratio of QTtime to QStime	(QT/QS)time	Slope of line joining S and T		STSlope		
Amplitude Features	Amplitude difference of P w.r.t. R (Y-axis)		Angle Features	Angle PQR	PQRangle		
	Amplitude difference of Q w.r.t. R (Y-axis)			Angle QRS	QRSangle		
	Amplitude difference of S w.r.t. R (Y-axis)			Angle RST	RSTangle		
	Amplitude difference of T w.r.t. R (Y-axis)			Angle RQS	RQSangle		
	Amplitude difference between P and Q			Py	Angle RSQ	RSQangle	
	Amplitude difference between Q and R			Qy	Angle RTS	RTSangle	
	Amplitude difference between R and S			Sy	Miscellaneous Features	Area of Triangle QRS	
	Amplitude difference between S and T			Ty		Ratio of QRS area to square of RSamp	
	Amplitude difference between Q and S			PQamp		Ratio of QRS angle to RST angle	
	Amplitude difference between P and S			QRamp		Ratio of QRS angle to QStime	
	Amplitude difference between P and T	RSamp	Ratio of RST angle to QTtime				
	Amplitude difference between Q and T	STamp	Ratio of RST angle to PQdis				
	Ratio of STamp to QSamp	QSamp	Ratio of QRS angle to RQS angle				
	Ratio of RSamp to QRamp	PSamp	Ratio of QRS angle to RTS angle				
	Ratio of PQamp to QSamp	QTamp	Ratio of RQS angle to RTS angle				
	Ratio of PQamp to QTamp	(ST/QS)amp	Ratio of QRS area to QRamp				
	Ratio of PQamp to PSamp	(RS/QR)amp	Perimeter of triangle QRS				
	Ratio of PQamp to QRamp	(PQ/QS)amp	Incircle radius of QRS triangle				
	Ratio of PQamp to RSamp	(PQ/QT)amp	Centroid of QRS triangle (Time)				
	Ratio of RSamp to QSamp	(PQ/PS)amp	Centroid of QRS triangle (Voltage)				
	Ratio of RSamp to QTamp	(PQ/QR)amp					
	Ratio of STamp to PQamp	(PQ/RS)amp					
	Ratio of STamp to QTamp	(RS/QS)amp					
		(RS/QT)amp					
		(ST/PQ)amp					
		(ST/QT)amp					

3. Classification and Identification

One of the most important Machine learning operations is classification. A classifier classifies data based on training set and the values in a classifying attribute, and uses it in classifying new data. In this paper, proposed methodology is tested using Artificial Neural Networks for Pattern Recognition (ANN), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) classifiers.

• **Artificial Neural Networks (ANN)**

Artificial Neural Networks are class of statistical learning models in Machine Learning and stimulated from biological neuron networks. At the primary stage, the Neural Network

has to be trained with ECG data of different persons, and then the neural network generated from the training is used for biometric identification of the individuals [23-24].

MIT-BIH ECG ID Database was used for training and testing purposes. The database was chosen because it includes more than one record for some of its subjects. Each individual person has 5 records, and out of these, we have used 2 records as train sets and the remaining records as test sets. The biometric system is analyzed for 20 subjects of each 5 individual records. The ECG records for training and testing are tabulated in Table 2. For convenience, Person_01/rec_1 record of MIT-BIH ECG ID database has been renamed as P1(1).

Table 2 Train and Test signals of ECG

Train Set	Test Set
Includes record-1 and record-2 of each Subject	Includes record-3, record-4, record-5 of each subject
P1,P2,P3, P9,P10, P11,P24, P25, P26,P28,P30,P32, P34, P36, P52, P53, P59,P63,P71, P72	P1,P2,P3, P9,P10, P11,P24, P25, P26,P28,P30,P32, P34, P36, P52, P53, P59,P63,P71, P72

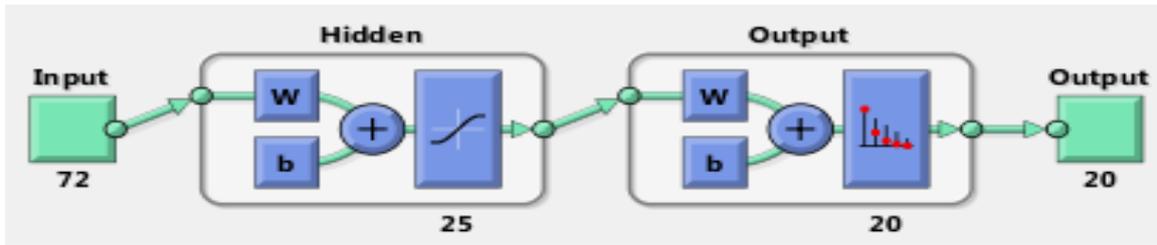


Fig. 9. Neural Network used for Pattern recognition

The 72 extracted features are subjected to a two layered feed forward network, with Sigmoid hidden and Softmax output neurons, which can classify vectors arbitrarily well, given enough neurons in its hidden layers [24]. From Figure 9, the 72 generated features were taken as input. The 25 hidden layers were taken for 20 output targets (persons) to be tested at that instant. Weight matrices connected to inputs are input weights and weight matrices connected to layers of outputs are layer weights.

In the recognition phase, the pattern recognition network created by the trained set of data is simulated over the test data containing the features of the ECG signal of a person arranged in the sequence as of the train data. The simulation generates the probabilistic matrix (Confusion Matrix) for each ECG signal for the entire set of target. The ECG signal is categorized as a particular target (person) for which the probability generated by the simulation is maximum.

Table 3 Confusion Matrix for True Classification of ECG Pulse data

	P3	P10	P25	P30	P32	P34	P36	P52	P53	P59	P72	P1	P2	P24	P28	P11	P26	P9	P63	P71
P3	17	0	0	0	3	3	0	0	0	0	0	0	0	8	0	0	0	0	0	0
P10	0	20	0	0	0	0	7	0	0	3	16	0	3	0	0	0	0	0	0	0
P25	0	0	20	0	0	8	0	0	0	0	0	0	0	1	0	0	0	0	0	0
P30	0	0	1	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
P32	2	0	0	0	15	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
P34	1	0	0	0	2	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P36	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0
P52	0	0	0	0	0	0	0	16	0	0	0	0	1	0	0	0	0	0	0	0
P53	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0
P59	0	1	0	0	0	0	2	0	0	7	0	0	0	0	0	0	0	0	0	3
P72	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	1	0	8
P1	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	2	0	0
P2	0	0	0	0	0	0	0	0	0	10	0	0	14	0	0	0	1	4	0	0
P24	1	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	1	0	0	0
P28	0	0	0	0	0	0	0	5	0	1	0	0	2	0	21	0	0	0	0	0
P11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0
P26	0	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	19	0	0	1
P9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	14	0	0
P63	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	2	0	0	20	6
P71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Match	81%	95.2%	95.2%	100%	71.4%	38.1%	57.1%	76.2%	100%	33.3%	23.8%	90.5%	66.7%	62.4%	100%	90.4%	90.4%	66.7%	100%	9.5%

For each ECG record, only 6 PQRST fragments along with their mean were taken. These best 6 with 1 mean altogether make the 7 sets of pulse data constitute one ECG record. Thus 3 ECG records used for testing have 21 sets of pulse data for each person included for testing. Table 3

reports True classification of ECG pulse data of persons used for testing the pattern recognition neural network. According to the confusion matrix obtained, the True Positive Rate (TPR) of over all True classification of pulses was found to be 71.20%.

Table 4 Confusion Matrix for True Identification of ECG Pulse data

	P3	P10	P25	P30	P32	P34	P36	P52	P53	P59	P72	P1	P2	P24	P28	P11	P26	P9	P63	P71
P3	21	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0
P10	0	21	0	0	0	0	14	0	0	0	14	0	0	0	0	0	0	0	0	0
P25	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P30	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P32	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P34	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P36	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0
P52	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0
P53	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0
P59	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0
P72	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0
P1	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0
P2	0	0	0	0	0	0	0	0	0	14	0	0	21	0	0	0	0	7	0	0
P24	0	0	0	0	7	0	0	0	0	0	0	0	0	14	0	0	0	0	0	0
P28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	0	0
P11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	0
P26	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	21	0	0	0
P9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0
P63	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	7
P71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14
Match	100%	100%	100%	100%	66.7%	100%	33.3%	100%	100%	33.3%	33.3%	100%	100%	66.7%	100%	100%	100%	66.7%	100%	66.7%

Table 5 Confusion Matrix for True Identification of Person (ANN)

	P3	P10	P25	P30	P32	P34	P36	P52	P53	P59	P72	P1	P2	P24	P28	P11	P26	P9	P63	P71
P3	3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
P10	0	3	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0
P25	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P30	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P32	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P34	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P36	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
P52	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
P53	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
P59	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P72	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
P1	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
P2	0	0	0	0	0	0	0	0	0	2	0	0	3	0	0	0	0	1	0	0
P24	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
P28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
P11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
P26	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	3	0	0	0
P9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
P63	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1
P71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Match	100%	100%	100%	100%	66.7%	100%	33.3%	100%	100%	33.3%	33.3%	100%	100%	66.7%	100%	100%	100%	66.7%	100%	66.7%

Table 4 represents True identification of pulses. The basic difference between True classification and True Identification is in terms of misclassified pulses. During classification, some of the pulses of a particular set may be misclassified. The process in which these misclassified pulses which are categorized for a true person for whom the true classified pulses out of 7 pulses in the pulse set is maximum is termed as True Identification. According to the confusion matrix obtained, the TPR of overall True identification of pulses was 81.01%, which is 9.81% more compared to the true classification of pulses.

Each set of pulse data contains 7 pulses, thus the 7 pulses represent one individual. So the number of persons involved in the test is 7 times less than the number of total pulses in each category. According to the confusion matrix, the TPR of overall true identification of person involved in testing session was 81.01%, shown in Table 5.

• **Support Vector Machines (SVM)**

Support Vector Machine is moderately a new Machine learning method used for binary classification. The biometric identification of ECG signal can be done by feeding the data of ECG features to Support Vector Machines (SVM). A set of known objects is called training set; each object of training set consists of a feature vector and a belonging class value. Based on the training data, the learning algorithm extracts a decision function to classify the unknown input data called test set. SVM continuously minimizes the empirical classification error and maximizes the geometric margin, so SVM is also called Maximum margin classifier. The proposed system uses multi class SVM, One against All (OAA) method with Linear Kernel [25]. Table 6 reports confusion matrix of True identification of a person involved in the testing session for SVM Classifier. It shows an overall True Positive rate of 82.758%.

Table 6. Confusion Matrix for True Identification of Person (SVM)

	P3	P10	P25	P30	P32	P34	P36	P52	P53	P59	P72	P1	P2	P24	P28	P11	P26	P9	P63	P71
P3	3	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
P10	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
P25	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P30	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P32	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P34	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P36	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
P52	0	0	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	0	0
P53	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
P59	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
P72	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1
P1	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	1
P2	0	0	0	0	0	0	0	0	0	1	0	0	3	0	0	0	0	1	0	0
P24	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
P28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
P11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
P26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
P9	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
P63	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
P71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Match	100%	66.7%	100%	100%	66.7%	66.7%	100%	100%	100%	33.3%	66.7%	100%	100%	66.7%	100%	100%	100%	66.7%	100%	33.3%

• **K-Nearest Neighbor (K-NN)**

K-NN is a type of instance based learning, where the function is only approximated locally and all computation is deferred until classification. Classification using an instance

based classifier can be a simple matter of locating the nearest neighbor in instance space and labeling the unknown instance with the same class label as that of the located (known) neighbor. In K-NN classification, the output is a

class membership. An object is classified by a majority vote of its neighbors, with the object being most common among its K-nearest neighbors being assigned to the class [26].

The features extracted from different subjects of each known ECG signals are formed into a matrix in which each row corresponds to a signal and each column corresponds to a feature vector of that signal. This vector is given as train data set to K-NN classifier. The unknown ECG signals which are to be classified are considered as test signals. The feature vector corresponding to the unknown ECG signals are known as test data. Based on train data set, the K-NN

predicts whether the test data is similar to the train data or not. K-NN classifier classifies test data set by calculating the distance between test samples and samples from train data set. For finding the distance, ‘hamming distance’ metrics are chosen here. In this work, ‘Nearest Neighbor Search method’ (NS) with exhaustive search algorithm is employed for classifying new points or query points (test data).

According to the confusion matrix from Table 7, the TPR of overall true identification of a person involved in testing session for KNN is 81.034%.

Table 7. Confusion Matrix for True Identification of Person (K-NN)

	P3	P10	P25	P30	P32	P34	P36	P52	P53	P59	P72	P1	P2	P24	P28	P11	P26	P9	P63	P71
P3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P10	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P25	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P30	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P32	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P34	1	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P36	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
P52	0	0	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	0	0
P53	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
P59	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
P72	0	1	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	1	0
P1	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
P2	0	0	0	0	0	0	0	0	0	1	0	0	3	0	0	0	0	1	0	0
P24	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
P28	0	0	0	0	0	0	0	0	1	0	0	0	0	0	3	0	0	0	0	0
P11	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
P26	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
P9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
P63	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
P71	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	3
Match	33.3%	66.7%	100%	33.3%	66.7%	100%	33.3%	100%	66.7%	66.7%	100%	100%	100%	100%	100%	100%	100%	100%	66.7%	100%

Table 8. The Summary Results of Classification

S.No	Performance Parameter	ANN	SVM	KNN
1	True_Classification Rate (TPR)	71.182	73.152	72.906
2	True_Identification Rate(TPR)	81.010	82.758	81.034
3	Over all Classification Accuracy	92.453	93.709	92.724
4	Recall	75.210	77.510	77.976
5	Specificity	95.130	96.793	96.273
6	Precision	78.642	83.263	79.263
7	F Score	74.212	77.841	76.851

The Performance of the three classifiers can be observed by examining the performance parameters shown in Table 8. All the performance parameters are calculated using Confusion matrix for true identification of a person. The results given in Table 8 indicate that SVM classifier is able to provide accuracy of 93.7094%, and it shows 1.2562%

improvement compared to ANN and KNN classifiers. Overall results of SVM showed improved results compared to other Machine Learning Classification Approaches. Table 9 shows comparison of proposed method with existing ECG based recognition techniques.

Table 9 Comparison with Similar studies

Author	Features	Techniques	Accuracy
Adrian, et. al 2008 [7]	Correlation coefficients from P-QRS-T Complexes	Percent Residual difference (PRD) Correlation Coefficient (CC)	70% (PRD) 80% (CC) (Lab database)
Irvine, et. al 2008 [8]	Discrete Wavelet	Minimum Distance Classifier (MD)	91% (Lab database)
Homer, et. al 2009 [9]	Frequency based features	K-NN	85.20% (MIT BIH)
F.Agrafioti et al. 2010 [10]	Fiducial based features	PCA with LDA	92.30% (PTB Database)
Md. Najam Dar et al. 2015 [11]	Discrete Wavelet Coefficients	Single Nearest neighbor classifier	82.30% (ECG ID Database)

Chen et al. 2014 [14]	Frequency features (Lyapunov exponents)	Back Propagation Neural Network (BPNN) with Chaotic features	91% (Lab database)
X.Tang, et. al 2014 [16]	Rough set	Quantum Neural Network (QNN)	91.70% (MIT BIH Arrhythmia)
Proposed Study*	Fiducial based features (N=72) with priority	ANN, SVM, K-NN classifiers	93.70 (SVM) (ECG ID database)

4. Conclusion

In this paper, a systematic methodology of ECG based biometric recognition has been proposed. The methodology includes ECG data acquisition, data preprocessing, P-QRS-T detection, feature extraction, cardiac cycle classification and finally ECG signal recognition. The ECG signal is universal and permanent in humans and results depict that it is also unique. Heart beat is an essential sign of life and cannot be mimicked by any other being, gives it the privilege of being a highly secured biometric feature. In this paper, the performance of proposed system has been tested with different Machine Learning Classification algorithms like

Artificial Neural Networks for Pattern Recognition (ANN), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) classifiers. The results obtained are very promising to support ECG as one of the biometric features. The achieved accuracy for different sets of ECG data were very appreciable. Finally, an overall accuracy of 93.7094% and improved results of performance parameters have been achieved with SVM classifier for biometric identification.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence



References

1. F.Agrafioti, "Robust Subject Recognition using Electrocardiogram," Master's thesis, University of Toronto, 2008.
2. R.Scott Blaul, John Chirillo, "Implementing Biometric Security", Indianapolis: Wiley publications 2003.
3. JH Connel, RM Bolle, Pankanti S, "Guide to Biometrics", Newyork: Springer Verlag 2003.
4. Y.Wang, F.Agrafoti, D.Hatzinakos, and K.N Plataniotis, "Analysis of human electrocardiogram for biometric recognition," EURASIP Journal on Advances in Signal Processing, pp.01-11, 2008.
5. R. Acharya, J. S. Suri, J. A.E. Spaan and S .M. Krishnan, "Advances in Cardiac Signal Processing", Springer Berlin Heidelberg, New York, ISBN: 3-540-36674-1, 2007.
6. Md.Najam Dar, M.Usman Akram, Arslan Shaukat and M.A Khan, "ECG based biometric identification for population with normal and cardiac anomalies using hybrid HRV and DWT features," IEEE International Conference on IT Convergence and Security, pp.01-05, 2015.
7. Adrina, D.C Chan, M Hamdy, "Wavelet distance measure for person identification using electrocardiograms", IEEE transactions on Instrumentation and measurement, vol. 57, no.2, pp.248-253, 2008.
8. Irvine JM, Israel SA, Scruggs WT," Robust human identification from Cardiovascular function", Pattern Recognition, vol.41, pp.3427-3435, 2008.
9. M. Homer, JM Irvine and S. Wendelken, "A Model based approach to human identification using ECG", in Proc.SPIE, Orlando, FL, 2009. Vol.7306.
10. F.Agrafioti and D. Hatzinakos " Signal validation for Cardiac Biometrics", in Proc.2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp.1734-1737, March 2010.
11. Md.Najam Dar, M.Usman Akram and S.A Khan, "ECG Biometric identification for general population using Multiresolution analysis of DWT based features", 2nd International Conference on Information and Cyber forensics (InfoSec), pp.05-10, 2015.
12. Biel L., Petterson O., Philipson L., Wide P, "ECG analysis: a new approach in human identification.", IEEE Transactions on Instrumentation and Measurement, vol.50, no.3, pp.808-812, 2001.
13. Muhammad Najam Dar, M. Usman Akram, Anam Usman and Shoab A. Khan "ECG Biometric Identification for General Population Using Multiresolution Analysis of DWT Based Features" IEEE International Conference on BioSignal Analysis, Processing and Systems (ICBAPS), pp.05-10, 2015.
14. C.K Chen, C.L Lin, S.L Lin, Y.M Chiu and C.T Chaing " A Chaotic theoretical approach to ecg based identity recognition, "Computational Intelligence Magazine, IEEE, vol.9, no.1. pp.53-63, 2014.
15. O. Boumbarov, Y.Velchev, S.Sokolov, " ECG Personal identification in subspaces using Radial Basis Neural Networks", IEEE Int. workshop on Intelligent data acquisition and Advanced Computing systems, pp.446-451, 2009.
16. X. Tan, L Shu, "Classification of Electrocardiogram signal with RS and Quantum Neural Networks", International Journal of Multimedia and Ubiquitous Engineering, vol.9, no.2, pp.363-372, 2014.
17. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 101(23):e215-e220 [Circulation Electronic Pages; <http://circ.ahajournals.org/cgi/content/full/101/23/e215>]; 2000 (June 13).
18. Kiran Kumar Patro, P. Rajesh Kumar, "De-Noising of ECG raw Signal by Cascaded Window based Digital filters Configuration" IEEE Power, Communication and Information Technology Conference, pp.120-124, 2015.
19. Kiran Kumar Patro, P. Rajesh Kumar, " A Novel Frequency-Time Based Approach for the Detection of Characteristic Waves in Electrocardiogram Signal", Springer India, vol.372, no.1, pp.57-67, 2015.
20. Lugovaya T.S., "Biometric human identification based on electrocardiogram", [Master's thesis] Faculty of Computing Technologies and Informatics, Electrotechnical University "LETI", Saint-Petersburg, Russian Federation; June 2005.
21. Kiran Kumar patro, P. Rajesh Kumar, " Effective Feature extraction of ECG for Biometric Application", Journal of Procedia of Computer Science, Elsevier vol. 115, pp.296-306, 2017.
22. Y.N Singh, Gupta, "ECG to Individual Identification" 2nd IEEE International Conference on Biometrics Theory, Applications and Systems, 2008. BTAS 2008.
23. Stevan A.Israek, Jhon M Irvine, "ECG to identify Individuals" Elsevier Pattern Recognition Journal vol.38, pp.133-142, 2005.
24. Ashutosh Gupta, Betsy Thomas, "Neural Network based indicative ECG Classification" 5th IEEE international Conference (Confluence), pp.277-279, 2014.
25. C. W. Hsu and C. J. Lin. A comparison of methods for multiclass support vector machines. IEEE Transactions on Neural Networks, vol.13, no.2, pp. 415-425, Mar 2002.
26. Nahid Ghofrani, Reza Bostani, "Reliable Features for an ECG-based Biometric System", Proceedings of the 17th Iranian Conference of Biomedical Engineering (ICBME2010), pp. 01-05, 3-4 November 2010.