Visvesvaraya Technological University, Belagavi.

PROJECT REPORT

on

"ANALYSIS OF ECG FOR BIOMETRIC IDENTIFICATION"

Project Report submitted in partial fulfillment of the requirement for the award of the degree of

Bachelor of Engineering

in

Electronics and Communication Engineering

For the academic year 2019-20

Submitted by

USN	Name
1CR16EC118	PRIYANKA G
1CR16EC131	REHANA SULTHANA RS
1CR16EC132	REKHA N

Under the guidance of

Dr. Binish Fatimah

Assistant Professor

Department of ECE

CMRIT, Bengaluru

Department of Electronics and Communication Engineering

CMR Institute of Technology, Bengaluru – 560 037

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to Certify that the dissertation work **"Analysis of ECG for biometric identification"** carried out by Student Priyanka G, Rehana Sulthana RS, Rekha N USN: 1CR16EC118, 1CR16EC131, 1CR16EC132 bonafide students of **CMRIT** in partial fulfillment for the award of **Bachelor of Engineering** in **Electronics and Communication Engineering** of the **Visvesvaraya Technological University, Belagavi,** during the academic year **2019-20**. It is certified that all corrections/suggestions indicated for internal assessment have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the said degree.

Signature of Guide	Signature of HOD	Signature of Principal
________________	________________	________________
Dr. Binish Fatimah,	Dr. R. Elumalai	Dr. Sanjay Jain
Associate Professor	Head of the Department,	Principal,
Dept. of ECE.,	Dept. of ECE.,	CMRIT,
CMRIT, Bengaluru.	CMRIT, Bengaluru.	Bengaluru.

External Viva

Name of Examiners	Signature & date
1.	
2	

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose consistent guidance and encouragement crowned our efforts with success.

We consider it as our privilege to express the gratitude to all those who guided in the completion of the project.

We express my gratitude to Principal, **Dr. Sanjay Jain,** for having provided me the golden opportunity to undertake this project work in their esteemed organization.

We sincerely thank Dr. R. Elumalai**,** HOD**,** Department of Electronics and Communication Engineering, CMR Institute of Technology for the immense support given to me.

We express my gratitude to our project guide **Dr. Binish Fatimah,** Assistant professor, her support, guidance and suggestions throughout the project work.

Last but not the least, heartful thanks to our parents and friends for their support.

Above all, We thank the Lord Almighty for His grace on us to succeed in this endeavor.

ABSTRACT

In recent years biometric systems have gained an impetus and are being used almost everywhere for authentication and validation. The need for highly reliable security systems is dictated by the modern life style. Biometric recognition using ECG is an alternative solution where the level of information security is very high. It can be used in places where security is the main concern. Starting from airports and crossing through key administration buildings and ending at transportation stations, all of these facilities are frequented by a large crowd. Hence, a need for identity verification by means of automatic biometric systems becomes a must in order to speed up the verification process and prevent criminals from penetration to determine the potential use of ECG as a biometric. To increase their security feature, many applications have also used biometric data for encryption purposes. Studies conducted have shown that there are various types of biometric traits such as fingerprint, iris scan, face and voice recognition, all of which are susceptible to threats. To overcome the drawbacks of these biometric traits we go for biometric system using ECG signals. ECG signals are used as they are found to be unique for each individual. Biometric recognition is an important field of information and security. There are two main categories of biometric characteristics: Behavioral and Physiological. ECG signal biometric comes under behavioral characteristics. ECG biometric is more reliable and satisfies most of the conditions to become an ideal biometric. ECG signals are unique due to difference in morphology or difference in the structure of heart of individuals. This paper presents a methodology to detect individual by first identifying the R-peak using Pan-Tompkins algorithm, computing fiducial and non-fiducial features followed by feature selection which is then used for identification of individuals by training models such as Support Vector machine (SVM), Ensemble Bagged Trees (EBT) and K Nearest Neighbor (KNN)

Table of contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Biometric Identification is a method used to identify a person either by detecting a specific behavioral characteristic or physiological feature. ECG was earlier used for medical purposes, which is now used for biometric purposes. Studies conducted have shown that there are various types of biometric traits such as fingerprint, iris scan, face and voice recognition all of which are susceptible to threats. In order to overcome the drawbacks of these biometric traits we go for biometric system using ECG signals. The signals are specific to each individual and can be used to prevent crimes based on forging.

One of the oldest and most widely used techniques is the Fingerprint Identification. Fingerprint of each individual is unique and unalterable. The steps involved in this technique are acquisition of image and feature extraction from the image. The decision on acceptance and rejection is made by matching the pattern of the finger impression with the database. Weak or fake fingerprint impressions cause trouble during the recognition.

Voice Biometrics is used to identify an individual based on their unique voice traits. This involves steps like voice signal acquisition through a microphone and pattern matching algorithm to find an exact match from the database. Drawbacks of these includes low accuracy, change in voice of the person due to illness such as cold, which then becomes difficult to recognize and also mimicking one’s voice for breaching security.

Iris scan based recognition involves pre- processing the image of the iris. Features from the iris are then compared in order to find the perfect match for the person from the database. Facial recognition involves steps similar to that of the iris scan. High-resolution image can be used to breach these systems.

ECG (Electrocardiography) based recognition involves the process of recording an individual’s heart beat by placing electrodes on the skin, to detect the electric signals produced by the heart muscles. The ECG indicates the recording of electrical activity of heart over a period of time. The shape of the waveform reveals the current state of the heart and it provides helpful information regarding the rhythm and function of the heart. One cardiac cycle in an ECG signal consists of the P - QRS - T characteristic wave.

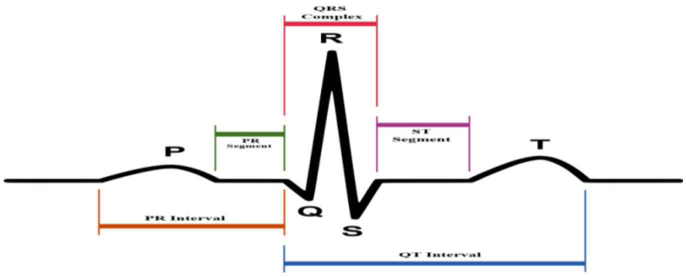

Fig 1.1 ECG Signal Pulse

A perfect biometric should have characteristics such as universality, easy measurability and uniqueness. The shape and position of the heart, and the presence and nature of pathologies govern the morphology and amplitudes of recorded cardiac complexes.

1.1 Why ECG?

The issues that arise in using EEG as biometric are:

- Data acquisition by placing the electrodes on scalp: the ECG are recorded from scalp electrodes, the synchronous activity of millions of neurons that have similar spatial orientation are the source of EEG .The EEG has significantly lower spatial resolution; therefore, the selection and placement of many electrodes requires special care in recording highly spatial resolution data, as inappropriate spatial separation of electrodes causes distortion of the estimated brain surface potential. Furthermore, with regard to this reference electrode, any single electrode can be chosen as the reference and the potential at another location recorded. Thus, inappropriate choice of reference often leads to misinterpretation of the source of the brain.

Variability of the brain wave patterns: The EEG varies with different brain states, such as walking, visualising, sleeping, thinking etc. EEG varies with age, adult EEG shows faster oscillation than childhood EEG, Therefore, a periodic

- Re - enrolment of an individual EEG could be one strategy to monitor the long-term variations.
- Scalability to a larger population: The potential of brain electrical activity as a biometric explored by most methods lacks the sample size needed to determine large-scale output. Therefore, a definitive conclusion on how to use the EEG as a biometric for identity verification is difficult to draw.

- Heritability of the EEG patterns: There were no significant differences in EEG spectral analysis among twins. Nevertheless, the twin spectra comparison has shown the intra-pair similarities that are significantly greater than the inter-individual similarity between unrelated individuals. Due to these drawbacks we go for ECG biometric.

1.2 Motivation

The need for highly reliable security systems is dictated by the modern life style. Biometric recognition using ECG is an alternative solution where the level of information security is very high. It can be used in places where security is the main concern. Starting from airports and crossing through key administration buildings and ending at transportation stations, all of these facilities are frequented by a large crowd. Hence, a need for identity verification by means of automatic biometric systems becomes a must in order to speed up the verification process and prevent criminals from penetration to determine the potential use of ECG as a biometric, it is necessary to evaluate how ECG satisfies the requirements for biometric characteristics. [1]A "perfect" biometric characteristic should be universal, i.e., each individual should have this characteristic, easily measureable, i.e., it is quite easy to measure the signal and convenient for an individual to obtain the characteristic, unique, i.e., no two individuals will have identical characteristics, and permanent, i.e., the characteristic does not change over time. Thus, good biometric characteristics can to a greater or lesser extent satisfy these requirements, depending on the purpose and application of biometric system.

1.3 Problem Statement

Most of the biometrics used in the present days is susceptible to threats and breaches. In order to overcome such threats, there is need for a reliable and efficient biometric system. Using ECG as biometric will help in increasing the security due to its characteristics such as uniqueness, easy measurability and universality.

1.4 Objective

The main intention of this project is to design a reliable security algorithm that uses ECG for individual identification. As mentioned earlier, ECG is unique to each individual and hence features extracted from them are also unique. Our main objective is to the perform feature extraction from the ECG signals and feed them to classifiers such Support Vector Machine (SVM), Ensemble Bagged Trees (EBT) and Random Forest, to train the classifiers using the training data set. Classification will be done on the test data set based on the training. Features selected could either be fiducial or non-fiducial in nature. Fiducial features include amplitude, location, duration and slope of the signal. Non-Fiducial features include energy, skewness, kurtosis and power of the signal. The algorithm developed should be able to identify individuals with good accuracy.

CHAPTER 2

LITERATURE REVIEW

2.1 Effective Feature Extraction of ECG for Biometric Application

In this article feature extraction involves four major steps Data Pre-processing, Feature Extraction, Effective feature extraction and Classification.

Data preprocessing: The ECG signals are noisy and contains a lot of distortions. The frequency components used to obtain the ECG signals may cause interference and adds noise to the signal. The noise present in the signal may tamper it and cause the data to be faulty. Appropriate filter has to be used in accordance to remove the noise. Digital cascaded filter configuration is used to remove three main noises of baseline drift, power line interference and EMG noise. After noise filtering, Smoothing filter is applied to the de-noised ECG signal. Smoothing the signal produces an approximate mechanism that attempts to catch essential trends in the data, while leaving out noise or other fine-scale structures / rapid phenomena.

Feature extraction: Each cardiac cycle consists of P-Q-R-S-T characteristic waves. Frequency Time based approach is used to detect the peak using wavelet decomposition. Majority of the information required for ECG biometric application is present in the intervals and amplitudes defined by its features. It is seen that QRS, ST and QT intervals change according to heart rate and physiology. In order to detect appropriate, the finest impulses must be detected out of the PQRST fragments. The priorities are selected based on different areas and distances in the signal such as QRS area, QT distance and ST distance. Only six best fragments are detected and used for further analysis.

The signal must be normalized about the origin before using it for further analysis and feature extraction mentioned above. In ECG diagnosis R-peak identification is more important. After finding out the R peak all the other peaks are detected in reference to this R peak by using frequency domain and pan-tompkins algorithm. Frequency domain approach is used for the detection of R-Peak only and after time domain approach is used for other characteristic wave’s detection. A total of 42 features were extracted. Here all 72 features were calculated separately for 7 pulses.

The ECG signal Features are extracted with a sampling frequency of 500hzThe best 6 PQRS fragments among the signal is chosen for data extraction ,these fragments are selected by their least difference from the mean of the Euclidian distance of the peaks.For identifying P-wave, a window in time domain is created with time gap limits from 65% of R-R interval to 95% R-R interval which is added to same R-peak location. The maximum value in that window will reflect P-wave. Effective feature extraction: The six PQRST fragments were computed along with their mean which constituted the 7 values for a particular individual. Assuming R-peak of the 6 fragments to be at the origin and other P, Q, R, S and T peaks are normalized at any standard point keeping the original aspect ratio of position unchanged. Thus features extracted from PQRST fragment were related to amplitude (y), time duration (x), difference between amplitude(amp), distance (dis), difference between times (time),slope, angle (angle), area and ratio of some features.

Classification: A classifier is used to classify data based on the training sets and the classifying attribute and uses it for classifying new data. The methodology used here includes Artificial Neural Networks for Pattern Recognition (ANN),Support Vector Machine(SVM) and KNN classifiers. There are 72 extracted features that are subjected to a two layered feed forward network, with Sigmoid hidden and Softmax output neurons, which can classify vectors given enough neurons in its hidden layers. 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 case of K-NN classifier, 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. This algorithm is implemented on MIT-BIH ECGID database which consists of 12 subjects, a total of 24 signals for testing and 36 signals for training.

This paper shows 98.14% over all accuracy for biometric system. Disadvantage is that fiducial features are costly and large number of features are used, advantage is that best PQRST fragment are used which is useful to reduce feature set (time efficient).

2.2 Biometric Identification from Raw ECG Signal Using Deep Learning Techniques

Introduction

Biometric involves 2 basic operating modes namely authentication mode and identification mode. In authentication mode the system compares each captured features with a specific template that is stored in biometric database by which it verifies the individual. In identification mode the system compares one to many template in a biometric database to determine the identity of an unknown individual.

In this paper, the process of identification of the ECG consists of the following phases: data acquisition, pre-processing, feature extraction, feature reduction and classification.

Data acquisition

The body monitoring can be conducted using 10 different sensors: pulse, oxygen in blood, airflow (breathing), body temperature, electrocardiogram, glucometer, galvanic skin response (sweating), blood pressure, patient position (accelerometer) and muscle/electromyography sensor.

Data acquisition was made using differential Op-Amp schema followed by 8-bit ADC operating at 277 Hz sampling rate. ADC data was transferred to PC via COM port using PySerial library. Each measurement lasted approximately 10 seconds, which means that user records typically contains approximately 10 or more heart beats.

It was decided to use modified Lead I schema for to place electrodes on the body surface and record the ECG tracing. Modified schema requires user to touch the electrodes with two fingers of his left hand and one finger of the right as shown at the picture below. This method is very convenient and can be applied for user authorization in everyday life.

Signal preprocessing

ECG beats are split into fixed-length windows (windows may overlap). And, if standard deviation of at least one sample exceeds certain selected threshold, then all samples within the window are recognized as outliers. Samples outliers are replaced by the mean values obtained by averaging corresponding samples from other segments. This procedure is done iteratively until no more outliers can be detected. Afterwards, samples have been averaged across all heart beats for each record.

Classification

In contrast to existing approaches, the proposed method in this combines two-stages, functions – extraction and classification of features.

Preprocessing of the ECG included the selection of appropriate beats and the removal of various artefacts. To detect uniformity and outliers sliding window approach was used. These steps include, The ECG beats are divided into fixed-length windows. And if the standard deviation of at least one sample exceeds the selected threshold, then all samples within the window are recognized as outliers. Sample outliers are replaced by mean values obtained by an average of corresponding samples from other segments. This procedure is done iteratively until no outliers can be detected. Afterwards, the samples were averaged across all heart beats for each record.

For this analysis a deep feed forward neural network has been selected as the basic architecture where data set is randomly separated into train set of 70% and test set of 30%. The system performance is evaluated after the training stage. The data set was randomly split into train set (70%) and test set (30%).After training stage is finished the system performance should be evaluated with accuracy.

Experiments and Results.

The data set used for the experiment consists of 147 ECG records of 18 unique individuals. The records have been selected from Lviv Biometric Data Set (minimal number of records per person is 8).

The first experiment was conducted to check how stable the classification results are, this check is done due to the DNN weights being used which are randomly initialized. The models end up in different results although they are trained under the same conditions. For this reason, 100 times iteratively trained different (selected) architectures were used. The results showed the deviation is large for all architectures and hence the most suitable approach is to run the same model multiple times and pick-up the best one. For further experimentation purposes three hidden layers with 70, 50, 30 neurons were chosen.

To investigate the impact of the number of classes on the overall classification accuracy the second experiment was conducted. The results showed that the additional classes do not impact the systems' significantly. This is because, by default, the DNN classifier will assign the records of an unknown user to an existing class which is completely wrong for identifying an individual. To overcome this rejection option is required.

The third experiment included the best rejection threshold to get optimal relation between identification rate and classification accuracy. The results showed that the optimal threshold is 70%, for which the classification accuracy is 96%.

The study shows that the hidden layers, the number of neurons and the number of users identified have an impact on identification accuracy. To obtain best results the same experiment should be run multiple times and model. To ignore the identification results that has low confidence level rejection threshold should be used. The results obtained are not as encouraging as expected, the reason this could be due to distortions that remain after filtering operation also less number of record per class. The results can be improved by using other DNN architectures, more advanced outlier correction algorithms, data augmentation that uses generative models and other techniques.

2.3 ECG as Biometric for Individual Identification.

In real time human recognition a new approach to processing the ECG signal and extracting the feature is used. Multi resolution wavelet analysis is used in this method to remove the special sections of the ECG signal beginning from the P wave, the QRS complex and the ending T wavelets are finite support length waveforms with an average zero value. Wavelet analysis decomposes signals using an orthonormal family of Base functions, they are well suited for analyzing transient,timevarying signals and are also well suited to ECG signals. Wavelet transformation is determined by transforming the signal under f(t) processing and the wavelet function φ(t).The discrete wavelet transform (DWT) is given by the below equation (1):

$$X_{j,k} = \int_{-\infty}^{\infty} f(t)\varphi_{j,k}(t) \quad (1)$$

The transforming wavelet decomposes a signal into two sub signal. Detail signal includes the upper half of the frequency elements, and the lower half of the approximation signal. The second description and approximation signal can be obtained with further decomposition. Therefore, multi resolution analysis can be conducted in a discrete wavelet domain. ECG feature extraction is done with Daubechies 8 (Db8) wavelet. Due to the structural similarity with QRS complex and energy spectrums being concentrated around low frequencies, Daubechies wavelet is expected to provide detail coefficients from multiresolution decomposition and shows better resemblance with QRS complex of the ECG wave in time scale domain.

The first step according to the proposed scheme is ECG wave detection. R wave is the most important wave in the ECG signal. With reference to R-peak all other waves of the ECG signal including T, P, Q and S waves can be easily located. The high frequency and low frequency contents of the ECG waveform are given by first and eight level reconstruction coefficient respectively, which in most cases represent noise. The decomposition results clearly shows that QRS complex is concentrated at level 3,4 and 5 .The coefficients d3,d4 and d5 and exploited for detection of QRS complex.

$$f = d3 + d4 + d5 \quad (2)$$

$$g = f * f \quad (3)$$

On using adaptive thresholding and equations (2) and (3), the coefficients are reconstructed and the R peaks are identified as the maximum amplitude points.

Q and S points are identified after the detection of R peak. The coefficients from d2 to d5 are reconstructed using equation (4)

$$h = d2 + d3 + d4 + d5 \qquad (4)$$

$$h' = (-h(x+2) + 8h(x+1) + 8h(x-1) + h(x-2)) / 12 \qquad (5)$$

Equation (5) provides the five-point differentiation which is used to identify Q and S points as maximum amplitude points before and after zero crossing .For detection of T and P which are mainly at a level of 6,7 and 8 , the reconstruction coefficients d6 and d7 are selected as the baseline drift is serious at d8.
Feature selection is done based on these features PR,PQ,RS,PS,RQ,RT,TS,TQ,QS,PT amplitude as well as interval durations: QS,PS,PR,PQ,PT,RT,ST,QT. The last step of identification procedure is classification. Template matching is done, where in each and every input feature vectors are compared to those stored in the database, to find the best match, TM is based on correlation coefficient. This measure exclusively can't be sufficient enough to locate the subjects in the database and hence Radial Basis function network is used as a classifier. The architecture of RBF network consists of an output layer with one node per category or class of data, an input vector and one layer of RBF neurons. The activation function in the RBF neuron measures the similarity between input and its prototype vector. The similarity function used is the Gaussian distribution. The performance assessment was carried out on four ECG public databases with a total of 149 people subjected to various physical activities and heart problems, the preliminary findings suggest that the system achieved an accuracy of 90-93%.

2.4 A Novel Electrocardiogram Biometric Identification Method Based on Temporal-Frequency Auto Encoding

The work can be divided into two parts

(1) On the basis of prior knowledge, we propose a QRST-targeting selection method for feature initiation. The method aims to remove most noise interference by only preserving identity-related information which refers to the QRST characteristics.

(2) For further feature extraction, auto encoder (AE), a classical neural network, is adopted as the tool to discover discriminative data structure from the initial features selected by DWT. AE has long been applied to fault diagnosis of machinery and image classification

The proposed identification process is mainly composed of five parts:

(1) pre-processing, (2) feature selection, (3) feature learning, (4) single-heartbeat identification, (5) multiple-heartbeat identification.

Here, since the signal label is obtained based on the voting of its contained multiple heartbeats, we call the signal identification process Multiple-heartbeat identification in this paper.

After de-noising, R peak detection and heartbeat segmentation is performed.

In this work, the Pan–Tompkins (PT) algorithm was adopted for R point detection because of its good performance on noisy ECG signals and was taken as the dividing reference during segmentation.

As mentioned above, a long ECG signal can be segmented into a series of heartbeats which are comprise successive characteristic waveforms as P, Q, R, S, and T. These waveforms appear in time order, and have different contributions to ECG biometric identification.

Later, it was found that using only three QRS based features was sufficient to identify a subject, which further highlighted the importance of QRS complex.

Based on the above, we can reasonably conclude the feature information of QRS and T wave contributes the most to ECG biometric identification.

Based on this idea, a QRST-targeting feature selection method, which consists of two parts: decomposition and selection, is proposed.

To convincingly evaluate the proposed method, in the Two-recording scheme, two recordings of each subject from the ECG-ID were alternately taken as the training set, and the rest one was used as the testing set.

When it comes to the All-recording scheme, the difference is that when one subject has over 5 recordings, its number of recordings used for training became two. All the recordings of each subject take turns to serve as the training data, and the rest of the recordings were used for testing.

On the MIT-BIH-AHA, ECG-ID (Two-recording), and ECG-ID (All-recording) database, the proposed obtained 5.89%, 9.68% and 11.93% higher for single-heartbeat identification accuracy than the reference method, which further proves the effectiveness of the proposed method.

Based on the above, the main advantages of the proposed method are high-level identification accuracy, convenient noise removal, and only R point detection required.

Later, the whole signal was segmented into multiple heartbeats centered on the detected R points, and each heartbeat roughly covered a cardiac cycle length. Then obtained heartbeats were processed by the proposed QRST-targeting feature selection strategy and represented by wavelet coefficients corresponding to QRS and T. In this way, most noise interference is removed, and only identity related information is preserved for the subsequent steps.

Compared with other fiducial methods, it has no need for accurate detection of Q, S, and T point. In addition to that, this work also produces high single-heartbeat identification accuracy on both healthy and patient databases and thus, is potentially suitable for wearable devices which use few heartbeats for fast identification.

In this work, a novel method based on DWT and S-AE for ECG signal identification was proposed.

Compared with other previous studies, we obtained promising average multiple-heartbeat identification accuracies of 98.87%, 92.3%, and 96.82% on the ECGID (Two recording), ECG-ID (All recording), and MIT-BIH-AHA databases, respectively, by using the raw ECG signals.

CHAPTER 3

SOFTWARE

3.1 MATLAB

MATLAB (matrix laboratory) is a multi-paradigm numerical computing environment and proprietary programming language developed by MathWorks.

MATLAB allows operations such as matrix manipulations, plotting of functions and data implementation of algorithms, creation of user interfaces and interfacing with programs written in other languages such as C, C++, C#, Java, Fortran and Python.
It combines a desktop environment tuned for iterative analysis and design processes with a programming language that expresses matrix and array mathematics directly. It also includes a live editor for creating scripts that combine the code, output and formatted text in an executable notebook.
MATLAB apps let you see how different algorithms work with your data. Iterate until you've got the results you want, then automatically generate a MATLAB program to reproduce or automate your work. Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing abilities. An additional package, Simulink, adds graphical multi-domain simulation and model-based design for dynamic and embedded systems.

MATLAB is a high-performance language for technical computing. It also integrates computation, visualization and programming in an easy-to-use environment where problems and solutions are expressed in a familiar mathematical notation.
Using MATLAB, one can perform data analysis, develop algorithms and create models and applications.
The language, applications and built-in math functions enable you to quickly explore multiple approaches to arrive at a solution. MATLAB also lets you take your ideas from research to production by deploying to enterprise applications and embedded devices as well as integrating with Simulink® and Model-Based Design.

Also the Classification Learner app used here trains several models to classify the data. Using this app, one can explore supervised machine learning using various classifiers. one can explore data, select features, specify validation schemes, train models, and assess results.

Automated training to search for the best classification model type, including decision trees, discriminant analysis, support vector machines, logistic regression, nearest neighbours, naive Bayes and ensemble classification can also be performed.

Supervised machine learning by supplying a known set of input data (observations or examples) and known responses to the data (e.g., labels or classes) can be performed. One can use the data to train a model that generates predictions for the response to new data.

To use the model with new data, or to learn about programmatic classification, you can export the model to the workspace or generate MATLAB® code to recreate the trained model.

CHAPTER 4

METHODOLOGY

An identification algorithm based on both fiducial features and non-fiducial features are proposed in this project. Fig. 4.1 is the block diagram of the methodology.

The approach proposed includes four major steps: Data Pre-processing, Feature Extraction, Feature Selection and Classification.

Flow chart:

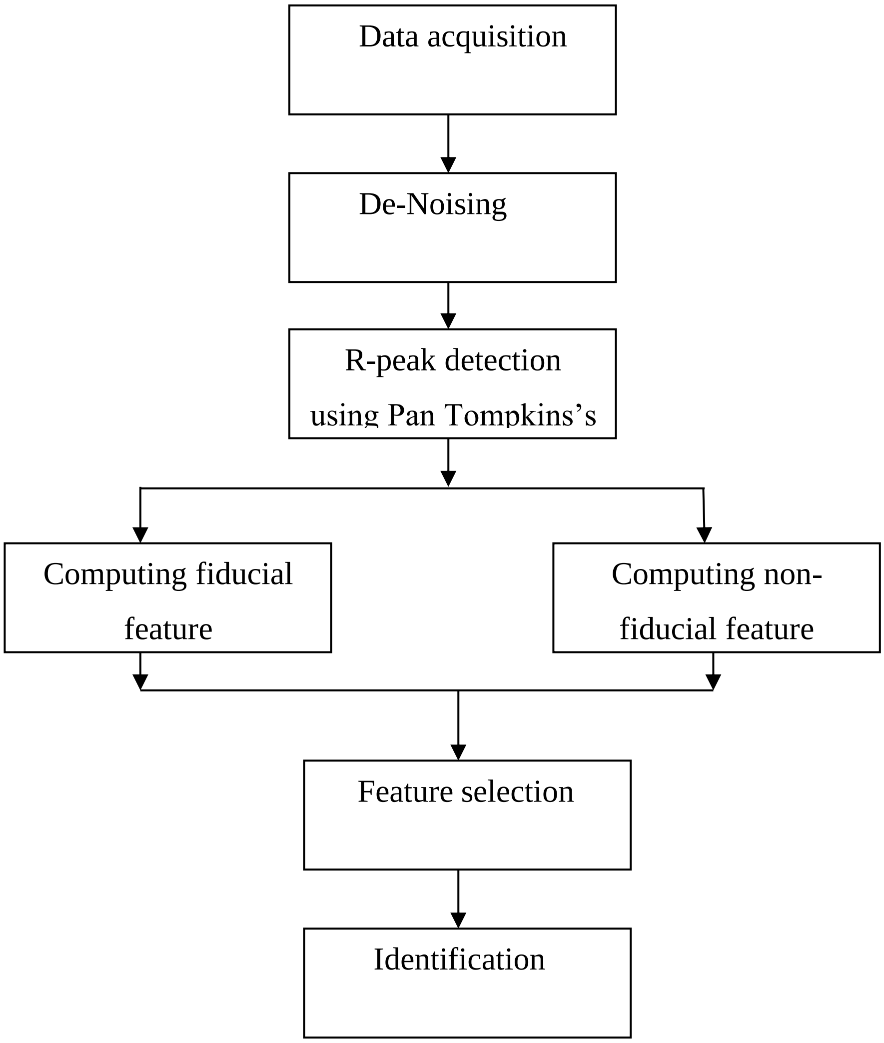

Fig: 4.1 Block Diagram of the methodology

4.1 Data Preprocessing

The recorder ECG data is corrupted with various kinds of noises such as the low frequency baseline wander, motion artefacts and power-line interference. These disturbances affects the important features of the ECG signal and lead to false peak detections. Therefore, before any features are extracted the given ECG signal must be de-noised.de-noising can be done by Subband decomposition or Fourier decomposition method.

4.1.1 Subband Decomposition

Wavelet Transform is a time-frequency analyzing method that the window of time and the window of frequency can change, which uses the multi-resolution feature to extract original signals from signals mixed with the noise interference, it has the unique advantage in dealing with non-stationary signals. Wavelet Transform has been used widely to process signals and images. Based on the nature of the quasiperiodic, the non-stationary and the sudden change of peaks of ECG signals, this paper selects Wavelet Transform to eliminate the noise interference of ECG signals. The basic idea is as follow: first of all, It uses Wavelet Transform to scale decompose ECG signals with noises into different frequency band signals, then, a variety of noises is removed, finally, the wavelet to reconstruct and restore useful signals to get ECG signals without noises.

If the function $\psi(t) \in L^2(R)$ (square integrable space) and meets $\int_{-\infty}^{+\infty} \psi(t)dt = 0$, then it is known as the mother wavelet or the basic wavelet. The basis function of Wavelet Transform is known as the wavelet function, for short the wavelet, which is expressed by $\psi_{a,b}(t)$, generated by the basic wavelet ψ (t) after companding and moving parallelly, as shown in the formula(4.1).

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \tag{4.1}$$

where a is the companding factor and b is the moving parallelly factor and a ≠ 0, a and b are the continuous quantities. $\bar{\psi}(t)$ is the conjugate of $\psi(t)$.

If the signal) f (t is reconstructed by the basic function of Wavelet Transform, its definition of inverse-transform is such as the formula (4.2)

$$\mathrm{f}\,(\mathrm{t}) = \frac{1}{C_\psi}\iint_\infty \frac{W_f \psi_{a,b}(t)}{a^2} dadb \tag{4.2}$$

where a ≠ 0 and a and t are the continuous quantities and $\bar{\psi}$(t) is the conjugate of ψ (t) . If the signal) f (t is reconstructed by the basic function of Wavelet Transform,its definition of inverse-transform is such as the formula(4.3).

$$C_{\psi} = \int_{+\infty}^{-\infty} \frac{|\widehat{\psi}|}{|w|} dw \qquad (4.3)$$

Where w is a continuous quantity and $\hat{\psi}($ is Fourier Transform of ψ (t). Further it can be represented in frequency domain also.

Wavelet Transform which decomposes the signal to the superimposition of a series of wavelet produced by the basic wavelet after companding and moving parallelly, is a localized time-frequency analysis method, that is, it has the lower temporal resolution and the higher frequency resolution in the low frequency part, and has the lower frequency resolution and the higher temporal resolution in the high frequency part. It has the automatic adaptive characteristics for signals, which is particularly suited to deal with ECG signals. But it gives rise to transition delay and phase delay, hence we go for FDM.

4.1.2 Fourier Decomposition Method

The FDM is an adaptive signal decomposition approach based on the Fourier theory. It maps the priori sine and cosine basis functions into a set of signal dependent basis functions. The signal is decomposed into a set of FIBFs which are local, adaptive, complete and energy-preserving (LACE). The FDM presents a generalized Fourier representation of a signal such that both amplitude and frequency are varying with time, i.e., each FIBF is an amplitude-modulated and frequency-modulated (AM-FM) signal, and thus FDM is suitable for a non-linear and non-stationary signal.

To preserve the energy of a signal s[n], the FDM decomposes it into a set of linearly independent non-orthogonal yet energy preserving (LINOEP) or orthogonal FIBFs with required frequency bands employing the following model of signal decomposition

$$\mathrm{s}[\mathrm{n}] = a_0 + \sum_{m=1}^{M} s_m[n] = a_0 + \sum_{m=1}^{M} v_m[n], \mathrm{m} = 1, 2, \cdots, \mathrm{M} \qquad (4.4)$$

where a_0 is the mean value of signal, s_m [n] and v_m[n] are mth orthogonal FIBFs and LINOEP, respectively. The FDM approach with discrete Fourier transform based zero-phase filter-bank, i.e., H_m[k]∈R_{0+}, ∀ m, k, is shown in Fig. 4.2. Moreover, the decomposition of signal s[n] into a set of desired orthogonal FIBFs {s_1[n], s_2[n], · · ·, s_m[n]} is also illustrated. The mth FIBF s_m [n] using inverse DFT can be derived as,

$$\boldsymbol{S_m[n]} = \sum_{k=0}^{N-1} [\, \boldsymbol{H_m[k] S[k] exp}(\frac{\boldsymbol{j2\pi kn}}{\boldsymbol{N}}) \qquad \textbf{(4.5)}$$

where Hm[k] denotes mth zero-phase filter and S[k] is the N length DFT of signal s[n]. To obtain a set of orthogonal FIBFs using the DFT based FDM, the frequency response of mth band can be defined as Hm[k] = 1 for the desired band of frequencies and zero otherwise, and thus these filter-banks are represented as,

$$\boldsymbol{H_m[k]} \begin{cases} = \mathbf{1}, & (\boldsymbol{k_{m-1}} + \mathbf{1}) \le \boldsymbol{k} \le \boldsymbol{K_m} \quad (\boldsymbol{N} - \boldsymbol{K_m}) \le \boldsymbol{k} \le (\boldsymbol{N} - \boldsymbol{K_{m-1}} - \mathbf{1}), \\ = \mathbf{0}, & \boldsymbol{otherwise} \end{cases} \qquad \textbf{(4.6)}$$

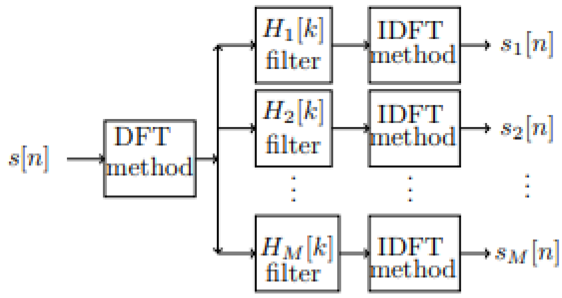

Fig.4.2 Block Diagram of the FDM

Using FDM, the ECG signal is decomposed into M FIBFs, where first FIBF (s_1[n]) captures low frequency baseline wander noise, another FIBF seizes and eliminates power-line interference, and all other FIBFs represent the bands of clean ECG signal. To obtain all the M FIBF components, one DFT and (M − 2) IDFT operations are used and complete method is implemented using the fast Fourier transform (FFT) algorithm which makes it computationally efficient. The FDM uses zero-phase filtering using non-negative real values (i.e. H_m[k] ≥ 0, ∀k, m) for defining the response of each filter. This zero-phase filtering preserves the salient features such as local maxima and minima (e.g. P, Q, R, S, T complexes) of the time-series, exactly at the same position where these features exist in the original time-series.

Authors in [1] used wavelet transform for the same. Other decompositions can also be used such as:

- Time frequency representation of a signal is used for analysis of time series signals. It maps a 1-D signal of time to 2-D signal of time and frequency.
- DFT sub-band and DWT sub-band are not suitable for nonlinear and non-stationary signals as they are valid under very general Dirichlet conditions.
- Empiric mode decomposition (EMD) is used for the analysis of such signals. It has drawbacks like aliasing.

Hence in this work, we have used FDM to remove baseline wander and power line interference, as discussed in [2]. FDM uses a bank of zero-phase filters which makes it a more appropriate de-nosing mechanism for algorithms using fiducial features.

4.2 R-peak Detection

Identification of R-peak is a crucial step in this algorithm as it is used for computing the rest of the fiducial features required for classification.

4.2.1 Pan-Tompkins algorithm

R peaks are obtained using Pan-Tompkins algorithm, which is a commonly used algorithm for detecting QRS complexes in ECG signals [3]. The QRS complexes are detected using 3 different types of processing steps: linear digital filtering, non-linear filter and decision rule algorithm. Linear processing includes a band pass filter, a derivative filter, and a moving window integrator. Amplitude squaring is performed in the non-linear transformation step. It squares the signal to amplify the QRS complex. Adaptive thresholds are used to detect the peaks of the filtered signal and discriminate the required information with T-waves. Combination of set of pre-processing methods are used to enhance the detection rate and the false detection of T-wave. Other waves such as P, Q, S and T are detected and calculated with respect to R being at the origin and using an algorithm given in [4](Implementation of an analog circuit or a real-time derivative algorithm providing slope information is straightforward. However, a derivative of its very nature amplifies the undesirable components of higher frequency noise. Also, due to its relatively low R wave slopes, many irregular QRS complexes with high amplitudes and long durations are missed in a strictly derivative approach. So the R-wave slope alone is insufficient to detect QRS properly. To achieve reliable performance, we need to extract other parameters such as amplitude, width and QRS energy from the signal.

Here, the QRS complexes have been identified using the area under non-linear phase space reconstruction. S and T waves are identified using state machines. QRS complexes are further identified using a digital filter bank and located using Multilevel Teager Energy Operator (MTEO)[5] which is computationally more efficient. One of the general methods used for peak detection, which also showed high performance for beat detection as well is the automatic Multiscale based peak detection.

4.2.2 Multilevel Teager Energy Operator (MTEO)

Among the existing unsupervised action potential detectors, the Teager energy operator (TEO) detector is reported to show relatively good performance at low SNRs. The TEO is a sort of time-frequency analyzer and has been used for many applications such as speech processing, image processing, and AM/FM demodulation. The k-TEO detector with the optimized shows good performance, but it works properly only if the optimal k is known prior to the detection. Since the optimal value of k varies with the shape of the action potentials to be detected, such prior knowledge is not available in general, and the k-TEO detector must work as a supervised detector.

A natural approach should be unsupervised combination of the outputs of a few k-TEOs with different resolution parameters k, which is, therefore, called the multiresolution TEO (MTEO). An MTEO detector that is used here is an improved version of the original MTEO in complexity and performance. Since the complexity of the detector is mostly concentrated on the TEOs and the smoothing windows, it is nearly proportional to the number of channels. The number of channels can be reduced if the sampling frequency of the input signal is known. Since the sampling frequency is the prior information of the recording system, not of the action potentials, it can be utilized in unsupervised systems. The MTEO detector needs about 6 or 7 -TEOs to cover the sampling frequencies from 10 kHz to 40 kHz. But if the sampling frequency is known, only three k-TEOs can achieve an equivalent performance.

Among the channels in the MTEO, small- channels tend to produce large peaks in response to high-frequency noise. Therefore, in an external noise environment, small k- channels with k less than that matched to the action potentials tend to cause many false alarms. In order to resolve this problem, we normalize the smoothing windows by the rms value of the noise. Having constant noise power in all channels implies relatively even distribution of noisy peaks over the channels, which, in turn, implies nearly constant contribution of each channel to the false alarm rate. This solves the problem of excessive false alarms caused by external noise in small k- channels. Thus the below fig.4.3 shows the new MTEO which perform as well as the optimized k-TEO without using any prior information of the signals.

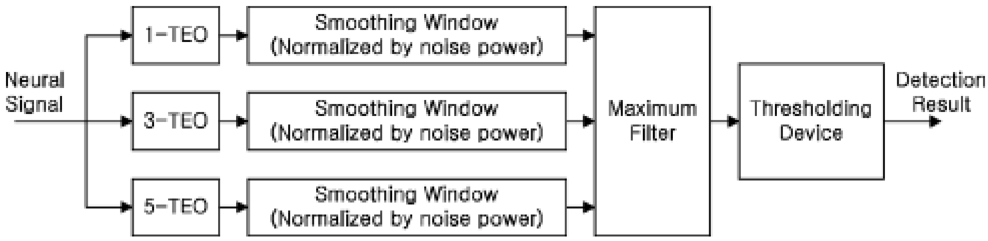

Fig.4.3 Multilevel Teager Energy Operator

4.3. Feature Extraction

ECG differs among individuals due to their specific anatomy and physiology of the heart. Thus, ECG attributes are represented by features unique to every individual. The internal features of the ECG can differ if an individual does any physical work and also changes over time. Such changes are not universal and differ from individual to individual and these effects are particularly reflected in P wave duration and PR interval.

In this project, a single ECG beat is used for identification and since the sampling rate is 500 Hz, 125 samples before the R peak and 200 samples after the peak has been considered as an average beat duration.

With respect to feature selection, existing approaches can be broadly classified as fiducial and non-fiducial.

Fiducial features are extracted from the P, Q, R, S and T waves onset/offset, boundaries, slopes, amplitude, peaks, angle and other measurements derived from the reference points as shown in Fig 4.4 , Fig:4.5, and Fig:4.6 [6]. Whereas, the non-fiducial feature extracted consists of AR model coefficients and L4 norm for each beat.

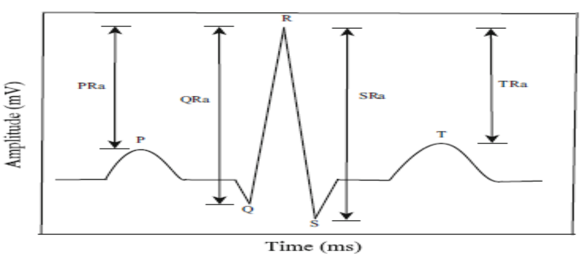

Fig: 4.4 Amplitude features of ECG waveform

The ECG trace for a single cardiac cycle includes following fiducial features:

- PR interval. The time between the beginning of the P wave and the beginning of the Q wave.
- P wave. Corresponds to atrial depolarisation.
- PR Segment. The time between the end of the P wave and the beginning of the Q wave.
- QRS complex. Corresponds to ventricular depolarisation.
- ST segment. The time between the end of the S wave and at the beginning of the T wave.
- T wave. Corresponds to ventricular repolarisation.
- QT interval. The time duration between the beginning of the Q wave and the end of the T wave.

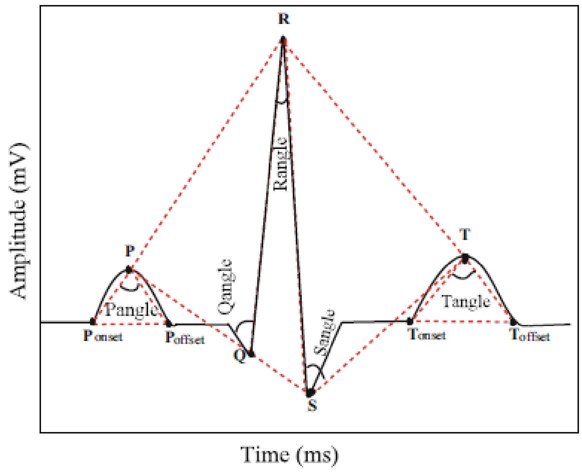

Fig: 4.5 Angle features of ECG waveform

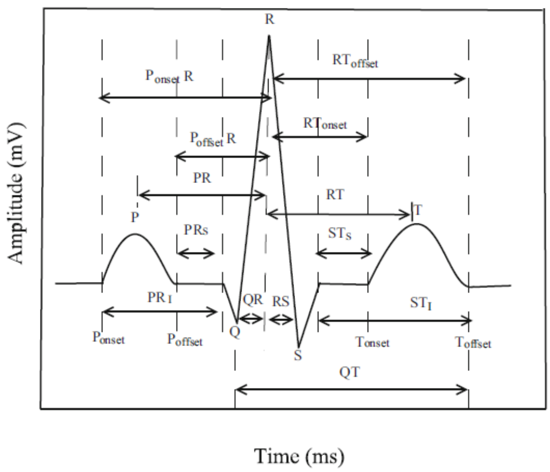

Fig.4.6 Interval features of ECG waveform

4.3.1 Outlier Removal

Even though noises are removed, it is almost guaranteed that some segments will be corrupted. In most cases, this is caused by noise dominating some parts of the ECG signal, for instance, if the subject performs an abrupt movement during the recording. In order to mitigate the effect of such interference, a simple outlier removal is done, which would drop the corrupt segments.

We can present all the heartbeat waveforms for a particular subject in the form of a matrix, where each row is a 250-dimensional vector representing a single heartbeat. As a first step, the algorithm creates a median waveform by taking median values for each of the 250 dimensions. The median was chosen, as it is not so affected by the outliers, as opposed to the mean.

Afterwards, the algorithm computes the distance between each heartbeat vector and the median. For this purpose, the Euclidian distance is used, though other metrics could be applied as well. Finally, the algorithm decreases 20 per cent of the most remote vectors.

In practice, this outlier removal procedure performed well even with ECG signals severely corrupted by noise. However, the setting of a hard cut-off point of 20 per cent has a significant drawback. For some ECG traces of high quality, more heartbeat waveforms were dropped than necessary. An improvement would be to use more advanced statistical methods to detect outliers, for instance, the Modified Z-scores or the IQR method

4.4. Feature Selection

The features considered in this project are given in Table I. In order to select only the relevant features from the extracted feature set, we compute the p-values of these features using Kruskal-Wallis test.

Feature	Feature Description
FIDUCIAL 1. Slope 2. Amplitude 3. Position 4. Distance	 -QS, QT, ST, RS, QR - P peak, R peak - P wave, Q wave, S wave, T wave - PQ,PT,ST, QT, QS
NON FIDUCIAL 1. Normal 2. Moment 3. Energy 4. Entropy 5. Skewness 6. Kurtosis 7. Auto regression	 - square root of its total energy - measure of the shape of a function -Area under squared magnitude -Measure of spectral power density -Distortion varies from normal distribution. -Measure of probability distribution of a real random variable evaluated -output variable depends linearly on its own previous values

Table I: List of Feature Extracted

Feature	p-value	Feature	p-value
F1	3.4645e-20	F15	1.0025e-138
F2	1.4995e-12	F16	2.7619e-146
F3	3 .8769e-16	F17	7.0231e-153
F4	3.2884e-25	F18	0.0014
F5	1.6701e-16	F19	8.1363e-04
F6	2.3736e-11	F20	8.2416e-04
F7	8.7324e-11	F21	7.6844e-04
F8	6.0260e-09	F22	5.4390e-04
F9	6.0260e-09	F23	6.7796e-54

F10	2.8671e-140	F24	9.1666e-70
F11	6.287e-87	F25	4.2488e-97
F12	4.0618e-141	F26	1.1758e-99
F13	3.528e-135	F27	6.2503e-93
F14	4.1855e-152		

Table II p-values of the Feature set obtained using Kruskal-Wallis method

4.4.1 Kruskal-Wallis test

The Kruskal-Wallis is a nonparametric statistical test that tests the differences between three or more independently sampled groups on a single continuous variable which is not normally distributed. For the Kruskal-Wallis test, non-normally distributed data (for example, ordinal or rank data) are suitable. In contrast, the one-way analysis of variance (ANOVA), which is a parametric test, usually used for a normally distributed continuous variable. The Kruskal-Wallis test is extended from Mann-Whitney U test. The Kruskal Wallis is therefore a more generalized form of the Mann-Whitney U test and is the non-parametric version of the ANOVA one-way test

The Kruskal-Wallis test is used to assess whether three or more groups differ on a single variable that fails to meet the normality assumptions of ANOVA. Since the variable of interest does not meet normality assumptions, we are not comparing by group means; instead we compare ranks.

The null hypothesis specifies that the groups are subsets from the same population (i.e., H_0: (a, b, c, ... , n) ⊆p). To test the null hypothesis, the groups are combined into a single group and rank the variable of interest. The new rank scores are summed by group (Ta, Tb, ... , Tn) and, along with group sample sizes, can be used to calculate the H statistic. H reflects the variance in ranks between groups and closely resembles the chi-square distribution. When testing the null, we can use H and refer to a chi-square table with degrees of freedom equal to n (number of groups) minus 1. If H exceeds a critical value, we may conclude that the groups do not come from the same population; many researchers often conclude that the test is an omnibus test for differences between groups.

The features with p-values < 0:05 are selected and used for identification of each ECG beat using popular machine learning classifiers like support vector machines (SVM), k nearest neighbour (kNN) and ensemble bagged trees.

4.5 Classification

The database used here includes 310 ECG recordings from ninety subjects. The number of records for each person vary from over 2 to 20 and also the recording duration varies from twice in one day to twenty times in six months.

Since we wanted to compare our results with state-of-art algorithms existing in the literature we have selected 12 subjects from the above-mentioned dataset as done in [12]. These records include person number 3, 10, 24, 25, 30, 32, 34, 36, 52, 53,59 and 72. The ECG-ID dataset includes 5 records of each of these subjects.

The data after denoising, detection of PQRST waves, statistical feature computation using Kruskal-Walis method, the features with p-values less than 0.05 are selected for classification of the given dataset.

Features	KNN	SVM	EBT
F1 to 9	75.2	59.4	62
F10	37.6	37.3	37.8
F11	23.2	35.7	35.7
F12	30.8	40.9	38.9
F13	39	30.4	36.6
F14	35	38.3	35.5
F15	42.1	37	41.5
F16	41.8	41.5	40.8
F17	32.9	37.3	32.8
F18	14.8	10.7	14.7
F19	14.1	8.8	13.7
F20	13.2	9.8	12.9
F21	13.7	9	13.4
F22	12.5	6	12.8
F23	27.3	42	46.9
F24	20.6	27	25.4
F25	23.6	29.3	26.7
F26	23.8	28.8	25.5
F27	29.1	28.8	28.9

Table III: Performance comparison of feature set

To further illustrate the contribution of each feature in the algorithm, we have computed the identification accuracy obtained for each feature for three machine learning classifiers, the results are given in Table III

The identification accuracy is computed as the total number of correct identification beats divided by the total number of beats.

The classification accuracies obtained with the complete selected feature set for kNN, SVM and ensemble bagged trees are given in Table IV.

Classifier	Accuracy (%)	AUC
KNN	97.99 ± 0.18	98.73
SVM	98.04	99.79
EBT	94.34	99.88

Table IV: Classification accuracies obtained for beat identification using machine learning algorithms.

SVM as a classifier is used for identification as it produces the highest accuracy of 98.04%.

4.5.1 KNN

K nearest neighbors is a simple Machine learning algorithm that stores all available cases and classifies new cases based on a similarity measure.

In KNN, K is the number of nearest neighbors. The number of neighbors is the main deciding factor. It is generally an odd number if the number of classes is 2. For instance, when K=1, then the algorithm is known as the nearest neighbor algorithmic.

It uses an independent variable (or set of independent variables) and a dependent variable (the thing we are trying to guess given our independent variables).

KNN works by finding the distances between a query and all the examples in the data selecting the specified number examples(K) closest to the query and then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression).

4.4.7 SVM

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression problems. However, it is essentially used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is the number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.

The SVM algorithm has a good feature to ignore outliers and then find the hyper-plane that has the maximum margin. Hence, we can say, SVM classification is robust to outliers.

Support vector machines have the following advantages:

- Useful in high dimensional spaces.
- Effective in cases where number of dimensions is greater than the number of samples.
- It is memory efficient as it uses a subset of training points in the decision function (called support vectors).

However a few disadvantages include the following:

- If the number of features is much greater than the number of samples, avoid over-fitting in choosing Kernel functions and regularization term is crucial.
- SVMs do not directly provide probability estimates therefore these are calculated using an expensive five-fold cross-validation.

4.4.8 Ensemble bagged trees (EBT)

Ensemble methods on the other hand combine several decision trees classifiers to produce a better predictive performance than a single decision tree classifier. The main principle behind the ensemble model is that a group of weak learners come together to form a strong learner, thereby increasing the accuracy of the model.

Combinations of multiple classifiers tend to decrease variance especially in the case of unstable classifiers and therefore may produce a more reliable classification than a single classifier.

Random Forest is an extension over bagging. It takes one extra step where in addition to taking the random subset of data, it also takes the random selection of features rather than using all features to grow trees. When you have many random trees, it's called Random Forest.

The steps required to implement random forest are:

1. Suppose there are N observations and M features in the training data set. First, a sample from training data set is taken randomly with replacement.

2. A subset of M features are selected randomly and whichever feature gives the best split is used to split the node progressively.

3. The tree is grown to the largest.

4. The above steps are repeated and prediction is given based on the aggregation of predictions from n number of trees. Advantages of using Random Forest technique include handling of higher dimensionality data essentially and handling missing values and maintaining accuracy for missing data.

Drawbacks of using Random Forest technique:

Since final prediction is based on the mean predictions from subset trees, it doesn't give precise values for the regression model.

CHAPTER 5

RESULTS

The performance of the proposed algorithm is compared with existing methodologies for ECG biometric systems in Table V and it can be observed that the proposed work gives promising results as compared to the contemporary literature

Authors Accuracy (%)	Subjects considered	Number of features	Accuracy (%)
Biel et al. [1]	20	30	98
Shen et al. [2]	20	7	95
Lugovaya T.S. [3]	90	250	96
Patro et al. [4]	20	72	99.13
Abdekar Sellami [5]	40	-	92
Patro et al. [6]	12	72	98.14
Proposed work	12	6	99.4

Table V: Comparison with state-of-art-algorithms

CHAPTER 6

ADVANTAGES AND APPLICATION

6.1 ADVANTAGES

- ECG signals are immune to attacks and hence suitable for biometric.
- ECG based biometrics are hard to steal or fake.
- Reliable - Passwords and PINs are forgettable, whereas biometrics are unforgettable, non-transferable and also has robust security features.
- Convenient - Biometric systems can be used to recognize the right person amidst hundreds of individuals or authenticate the right individual in a second or less.

6.2 APPLICATIONS

- UAV or unmanned vehicles are military vehicles that are subject to attack and hacking. These vehicles use autonomous behavior whose communication system is highly susceptible to potential cyber-attacks. To avoid such breaching there is need for reliable biometric systems. In such cases ECG can be used as biometric.
- Biometric Identification at Airports, key administration buildings and transportation stations is highly crucial and hence finds its application in this area.
- Data centers are store house of large computer systems and associated units such as telecommunication systems that holds huge amount of data and is of high importance for IT operations. As data is of high importance for such businesses there is need for highly reliable security system.

CHAPTER 7

CONCLUSIONS AND SCOPE FOR FUTURE WORK

7.1 Scope for future work

The proposed algorithm currently makes use of 12 subjects obtained from the publicly available dataset from Physionet database.

The dataset used above consists of 90 subjects, unlike the proposed algorithm that utilizes only 12 subjects, the complete dataset can also be used.

Different publicly available datasets such as PTB-XL, Arrhythmia dataset etc. can be used to carry out the proposed algorithm using different number of subjects to obtain identification accuracies closer to or even higher than the proposed methodology.

7.2 Conclusion

We proposed an identification algorithm based on ECG biometric system. Since algorithm depends on QRS complex and accurate R peak detection is vital, noise is removed in pre-processing using FDM which gives zero phase difference.

The algorithm used both fiducial and non-fiducial features. The performance of the proposed method has been validated on a publicly available dataset and compared with existing algorithms; we have obtained better results using less number of features. SVM classifier is used for identification and 98.5% of accuracy is obtained. As fiducial approach is computationally more expensive, we have also used only non-fiducial features as well to obtain an accuracy of 99.4%.

REFRENCES

1. A. Sellami, A. Zouaghi, and A. Daamouche, “Ecg as a biometric for individual’s identification,” in 2017 5th International Conference on Electrical Engineering - Boumerdes (ICEE-B), Oct 2017, pp. 1–6.
2. A. Singhal, P. Singh, B. Fatimah, and R. B. Pachori, “An efficient removal of power-line interference and baseline wander from ecg signals by employing fourier decomposition technique,” Biomedical Signal Processing and Control, vol. 57, p. 101741, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1746809419303222
3. J. Pan and W. J. Tompkins, “A real-time qrs detection algorithm,” IEEE Transactions on Biomedical Engineering, vol. BME-32, no. 3, pp. 230– 236, March 1985.
4. H. Sedghamiz, “Biosigkit: A matlab toolbox and interface for analysis of biosignals,” Journal of Open Source Software, vol. 3, no. 30, p. 671, 2018. [Online]. Available: https://doi.org/10.21105/joss.00671
5. Choi JH, Jung HK, Kim T. A new action potential detector using the MTEO and its effects on spike sorting systems at low signal-to-noise ratios. *IEEE Trans Biomed Eng.* 2006;53(4):738-746. doi:10.1109/TBME.2006.870239
6. N. Samarin and D. Sannella, “A key to your heart: Biometric authentication based on ecg signals,” 06 2019.
7. For textbooks – A.V. Oppenheim and R.W. Schafer, Digital Signal Processing, Englewood, N.J., Prentice Hall, 3 Edition, 1975.
8. For papers – Devid, Insulation design to combat pollution problem, Proc of IEEE, PAS, Vol 71, Aug 1981, pp 1901-1907