Machine Learning Classification and Feature Extraction of Arrhythmic ECG Data

Sumanta Kuila, Sayandeep Maity, Suman Kumar Mal, Subhankar Joardar

Abstract: Electrocardiogram (ECG) is the analysis of the electrical movement of the heart over a period of time. The detailed information about the condition of the heart is measured by analyzing the ECG signal. Wavelet transform, fast Fourier transform are the different methods to disorganize cardiac disease. The paper elaborates the survey on ECG signal analysis and related study on arrhythmic and non arrhythmic data. Here we discuss the efficient feature extraction process for electrocardiogram, where based on position and priority six best P-QRS-T fragments are studied. This survey examines the the outcome of the system by using various Machine learning classification algorithms for feature extraction and analysis of ECG Signals. Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN) are the most important algorithms used here for this purpose. There are several publicly available data sets which are used for arrhythmia analysis and among them MIT-BIH ECG-ID database is mostly used. The drawbacks and limitations are also discussed here and from there future challenges and concluding remarks can be done.

Keywords : Electrocardiogram, Machine learning , Classification , Arrhythmia Database ,Physionet.

I. INTRODUCTION

In today’s world automation in healthcare domain is the business reality. The biometric recognition is obviously an important and mandatory idea in the field of information and security. The physiological and behavioral characteristics of the of healthcare data will give the support to the medical practitioners for the treatment of the patients. In this paper we studied raw Electrocardiogram (ECG) signal and various technique to determine a variety of heart disease, which is called heart arrhythmia in the medical term. There are different types of arrhythmias, like morphological which is generated by single irregular heartbeat and the other category consists set of irregular heartbeats which is called rhythmic arrhythmia, where each type of arrhythmia is related with certain pattern. Identifying and classifying the pattern, the nature of arrhythmia is identified and based on that the physician can initiate the treatment. In this survey the classification and analysis of heartbeats, study of wave frequency of the ECG signal and the process of identifying as well as classifying arrhythmias are discussed. Today a lot of work and research is going on machine learning and feature extraction, and from there intelligent decision making is also a goal[1] [2]. The structure of the paper contains different sections. Section II describes the ECG signal and the basic characteristics of the signal for feature extraction, Section III describes the details of MIT-BIH database and its characteristics, Section IV describes methodologies for ECG data acquisition , Section V describes different techniques and algorithms used for signal classification and feature extraction and last two section contains conclusion and references.

Manuscript received on May 25, 2020.
Revised Manuscript received on June 29, 2020.
Manuscript published on July 30, 2020.
***** Correspondence Author

Sumanta Kuila, Dept. of Computer Sc. & Engineering , Haldia Institute of Technology, Haldia-721657, West Bengal, India.

Sayandeep Maity, Dept. of Computer Sc. & Engineering, Haldia Institute of Technology, Haldia-721657, West Bengal, India. sayandeepmaity328@gmail.com

Suman Kumar Mal, Dept. of Computer Sc. & Engineering, Haldia Institute of Technology, Haldia-721657, West Bengal, India. sumank343@gmail.com

Subhankar Joardar, Dept. of Computer Sc. & Engineering, Haldia Institute of Technology, Haldia-721657, West Bengal, India

II. ECG SIGNAL

This survey work is based on ECG signal extraction and classification. Here we have described the nature of the raw signal through wavelet transform and also described the Arrhythmic ECG data which diagnose the heart disease.

2.1 Depolarization

ECG Signal is an important component to check the cardiovascular status of human beings. The recording of the heart’s electrical movement is expressed by ECG. Different cardiac disorders is indicated by the differences of usual electrical pattern displayed in the electrocardiogram. Normally cardiac cells are electrically polarized, whose inside portions are negatively charged with compare to the outside portion. The primary activity of heart is depolarization, it is a process by which cardiac cells drop their standard negativity[3].

2.2 Elementary waveforms

For the treatment of advanced cardiac disease, ECG signal analysis is very important because from there the physician can detect and diagnose the heart condition. The modern treatment is supported by ambulatory electrocardiography (known as AECG) and it affords accurate and strong information for clinical diagnosis. Automatic analysis is an important criteria for AECG and it helped the physician to analyze the data during short span of time, i.e between 24 and 48 hours[4].

2.3. Description of the ECG signal

The Figure 1. demonstrate the details of heartbeat and its concern waveform which describes the electrical activity of the heart relax and contract.

The waveform patterns and amplitudes are the key factor for the analysis of ECG signal. As the figure (Fig. 1) shows the waveform of ECG has some properties in its shape. The frequency range of a standard ECG signal lies between 0.05 HZ to 100 HZ. It contains five different peak names, they are P,Q,R,S,T. Apart from that T wave, P wave and QRS complex wave are there [5]. The P wave signifies the electrical signature of the passed current in atria for re-polarization, and T wave signifies re-polarization of ventricles.

For the left and right ventricles QRS wave are created. RR-interval is the time distance among two successive QRS-complex. In the horizontal axis the width of the wave determines the time, depth and height of wave quantify the voltage [6][7].

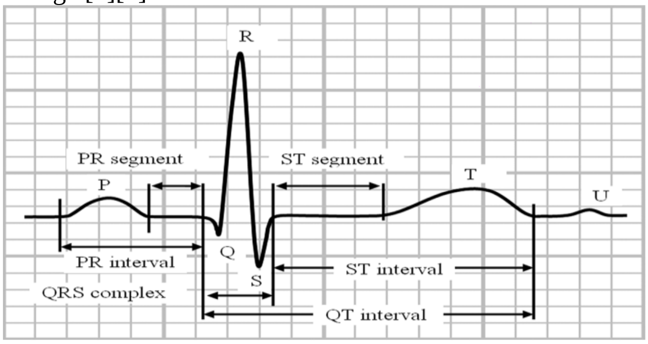

Figure 1. : Detailed ECG diagram [4]

2.4. Arrhythmic ECG Data

Heart arrhythmia is the problem related to rhythm of the heart and it feels when electrical impulses malfunction to our heart. The electrical impulses are responsible to coordinate our heartbeats and it diagnose the condition of the heart whether it is too slow, too fast or irregular. So cardiac Arrhythmia describes the heart related disease where by analyzing the ECG signal physicians perform treatment to the patients[8]. The goal of the study is to discuss the different techniques which will help to develop an efficient algorithm for the automation of arrhythmia detection. Patients having the problem with ventricular premature contraction (PVC) and atrial premature contraction (APC) both will be treated arrhythmia detection [9].

III. MIT-BIH DATABASE & AAMI STANDARD

From 1975 Beth Israel Hospital of Boston started work on ECG arrhythmia analysis and related subjects in collaboration with MIT. The MIT-BIH Arrhythmia Database was the first key product from that initiative and in 1980 it was completed. Later the database was generally accessible a set of standard experiment material for estimation of arrhythmia detection. The collaboration between American Heart Association database and MIT-BIH Arrhythmia database plays an important role to analyze arrhythmia and its characteristics and it becomes the standard database for heart treatment and medical research[10].

3.1. Characteristics

The various abnormalities of cardiac pulse and related problems are described in arrhythmia and a standard record set is required to analyze the problem for future treatment is MIT-BIH Database. Since 1970 the research work is going on to automate the ECG arrhythmia data. The signal is recorded such that it takes large variability of ECG beats and the detail facts of waveform are available there. This database supports different algorithms of automated arrhythmia which was developed between 1960 to 1970. Performance of arrhythmia detection algorithm depends upon the data and strong database gives the most effective and smart result for decision making on the concern medical area[11]. The modern standard of the database support most of different types of arrhythmia analysis and it takes almost five years long time duration to complete the database. Here Del Mar Avionics model 445 were used for ECG data recording, it also uses double channel Holter records for data recording and the playback unit of Del Mar Avionics model 660 generates the analog signal for the purpose of digitization[12].

3.2. Database Analysis

The study of MIT-BIH Database proceeds episode-by episode and beat-by-beat and the algorithms used to analyze the database uses some basic method which works to assess the performance of ventricular arrhythmia detectors .This was sponsored by Association for the Advancement of Medical Instrumentation (AAMI), an well known medical device manufacturer during the years between 1984 and 1987. The data added from ambulatory electrocardiographs marked with American National Standards and it is used for the classification of arrhythmia and ST segment measurement[4][13]. At the time of processing and testing the data for the comparison with different case studies the testing algorithm procedures reference annotation files for that purpose. The performance measurements and the comparison between different case studies are done by analyzing the annotated files where it uses standard comparison software are used. The concern agencies and the technicians of arrhythmia analyzers examine and verify the test results by using the comparison software, test data and accessory hardware devices which produces the required annotation files[14].

3.3. Database

The MIT-BIH database is created between 1975 to 1979 and with proper editing and modification the final version comes in late 1980. During the time period between 1981 and 1989, a CD-ROM is created and it contains huge amount of supplementary recording used for various research and development projects. The CD ROM has the approximate data of 600 MB which is totally in digitized form. The total duration of recording is more than 200 hours and the important criteria is that it has bear-by-beat annotations with the collected data. This data is an important raw data for data analytics as well as big data analytics[15]. The CD-ROM contains bulk amount of ECG data for arrhythmia which checks variety of waveforms morphology, with accurate cardiac pulses, with external noise and artifact. The modern version of CD ROM contains High Sierra format standard and it is commercially available with standard price and the database works mainly with format in addition with other format also[16].

IV. METHODOLOGY

The acquired ECG data is processed with certain techniques and methodologies which helps proper signal classification and feature extraction.

4.1. ECG data Acquisition

The data acquisition from the raw ECG signal uses the baseline noise reduction technique. The original raw ECG data describes the exact heart condition of a patient. So it contains irregular, uneven peak form unequal distance between peaks.

Due to abnormal partial breathing it generates different low frequency components. The particular low frequency component should be removed to minimize the influence of those features. The method of fast Fourier transform is used to minimize the low frequency and by using inverse fast Fourier transform the original ECG data is restored. Segmentation of ECG beat was created to analyze the data and to reduce the baseline noise[17]. The important concept about ECG signal is Wavelet transform, where it is concurrently observed at three different stages of focus to operate. Here the classification of beat and its characteristics described as premature ventricular contraction (V) and normal(N). According to the research on ECG signal the most frequent class used for supra-ventricular arrhythmia (S) are atrial flutter (AFL), normal rhythm (N) and atrial fibrillation (AF). The rigorous study and analysis are going on MIT-BIH Arrhythmia database which is publicly available database for arrhythmia detection. Apart from that several data acquisition system on minimum cost, The American Heart Association (AHA), The European Society of cardiology database (EDB), The Noise Street Test database (NST) and the Arrhythmia database of Creighton University (CU) are the popular and publicly available Arrhythmia database for research and analysis. The databases are created with the help of standard data acquisition system which consists of different modules like power supply, amplifiers, analog to digital converter (ADC), filters, isolators and interfacing circuits. The frequency range between 0.05 Hz and 113 Hz is the ideal range to analyze the ECG signal to collect the data accordingly. Important clinical application such as detection of cardiovascular disease (CVD), SCA prediction is done through the support of the frequency information of ECG signal. There are several ways of data acquisition and maintenance, one of the significant process is to use the NI6008 data acquisition card which is used save the acquired data. The conversion of ECG analog signal to digital signal is done through data acquisition interface and it is then saved for further use of medical investigation[18].

4.1.1. Train Set

The training data set is extracted from the central data repository and analyzing the dataset decision algorithm can be made. The references of different set works with different annotations and the training set is used accordingly. By the help of the annotated data the location and characteristics of the observed QRS complex data for the ECG signal is decided. Apart from the train data set for analyzing the ECG Signal a separate hidden data set is also available which determines the ranking or priority of the problem to solve the arrhythmia challenges. From the MIT-BIH arrhythmia database where Physionet provides train and test classification model for comparing the results from there we can make some new classification model[19]. Through the training set data and its analytical result gives the opportunity to the physicians to bring the necessary and accurate arrhythmia treatment[20].

4.1.2 Test Set

The Test set with reference to MIT-BIH arrhythmia database, AAMI suggested that the five different classifications are there, and according to AAMI 102,104,107,217 are the records with paced beats which are not considered for classification. The characteristic of the database says that it is highly imbalance (around 90%) if it belongs to class N, where as it is just 3% for SVEB, 6% for VER, and 1% for F classes respectively. So in the test set the work with different samples and its classifications is important. With compare to train set the test set includes wide variety of signals and test set signal sampling rate belongs between 120 and 1000 sample/second [21].

4.2. Data Preprocessing

The ECG is the video footage of the electrical characteristics of the heart over a certain period. The importance of the signal is that, analyzing and drawing the interface from signal the medical practitioners diagnose some disease. The obtained ECG signal usually effected by various types of artifacts and noise which resides within the frequency band of the generated ECG signal. This often changes the attributes of the ECG signal and for that reason extracting useful and correct information become difficult. Baseline wander noise, Electromyographic Noise , Power line Interference are the three major noise which are responsible for the corrupt and incorrect ECG signal. The data processing technique will lead the removal of noise from ECG signal by applying cascaded FIR filters. Figure 2 describes the configuration of a cascaded digital filter where Raw ECG data is inserted as an input, data then filtered and in the output De-Noise ECG signal will come. The three major noises that effects the ECG signal can be removed by cascaded digital filtering. The Kaiser window based filters are the important issues for the De-Notice ECG filters [22] [23].

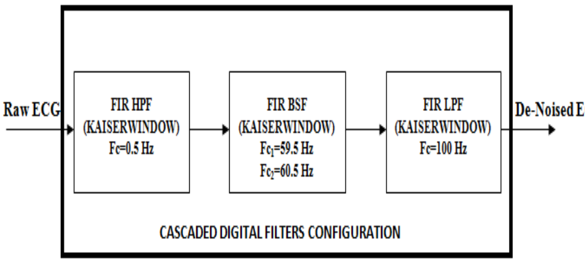

Figure 2 : Configuration of digital filter [23]

4.3. Segmentation

Detection of the QRS complex or R peak is the key issue for heart beat segmentation. The heartbeat segmentation accuracy made it easy for the researcher to classify the status of the heart signal. The accuracy of the heart beat segmentation depends upon two significant measure, positive predictivity and sensitivity, denoted as Positive Predictivity $_{(SEG)}$ = TP/(TP+FP) and Sensitivity $_{(SEG)}$ = TP/(TP+FN). False Negative(FN), False Positive(FP) and True Positive(TP) defines the heart beat number and the correct segmentation of the number[24].

This model can be defined and explained with the outcome of some standard medical diagnostic tests done on heart disease [25]. The accuracy of the result can be resolute from specificity and sensitivity. The concept describes the specificity, sensitivity and accuracy of the standard test result.

Outcome of diagnostic Test	Standard of Truth		
	Positive	Negative	Row Total
Positive	TP	FP	
Negative	FN	TN	
Column Total	TP+FN (Total number of subjects with given condition	FP+TN (Total number of subjects without given condition)	N = TP+TN+FP+FN (Total number of subjects in study)

Table-1: Terms used to define sensitivity, specificity and accuracy[26]

Here the table data describes total number of subjects with given condition and without given condition. For arrhythmia analysis which focuses on heart beat segmentation where the comparison of methods uses MIT-BIH database which is publicly available and it uses AAMI standards. Several sophisticated procedure is used to implement the segmentation such as wavelet transform, genetic algorithms, Quad Level vector , filter banks etc. Apart from these algorithms several other algorithms are also there to recognize different waves related to heartbeat, such as T wave and P wave. Detection of QRS complex and R peak is the important issue for the heart beat segmentation, and the database is the key repository to supply the information related to that. Here automation of the arrhythmia classification is done by using different segmentation algorithms. So segmentation has the direct impact on automation and for feature extraction the this impact works with exact result[26].

4.4. Feature Extracion

Feature extraction starts from calculating heartbeat interval or cardiac rhythm. Here RR interval works an important role as it reflects the variations in width at the time of classification. Analyzing the feature the pacemakers are allotted to the concern patients, where proper measure of RR interval is studied and this data indicates arrhythmia. RR interval have the huge competence to differentiate the types of heartbeat and feature extraction and reduction of noise interference. It has great impact on heart arrhythmia when we are calculating the RR interval average against certain time interval for a particular patient. Different experiment result shows that the classification results are significantly improved by normalized RR interval[27]. For feature extraction QRS interval and the QRS complex duration is utilized maximum times to extract high quality selective features. Determination of fiducial points are also important and several algorithms are available to identify it. Now if we analyze the samples of the curve we understand that it produces the vector and the feature of which expresses the high dimensions. Performance will improve if the dimension of feature vector is reduced and to do that several techniques is used upon the samples generated from the heartbeat pulses such as Independent component analysis (ICA) and Principal component analysis (PCA). Analyzing the signal new coefficient are generated which will represent the heartbeat signal. The aforesaid two techniques PCA and LDA study reveals that, to extract features ICA is better and for ECG signal to decrease the noise and related artifacts PCA is more productive. To enable statistically disconnected mixed signal, ICA technique is used and it mainly works with the signal of different arrhythmia class which is the combination of different action potentials[28]. The energy input to the signal depends on separate sources and this classification is done by PCA technique. For feature extraction on ECG classification the combined effect for ICA and PCA is wide and apart from that another technique Kernel Principal Component Analysis (KPCA) is more powerful than PCA because of its nonlinear structure[29].

4.5. Selecting Feature

Feature selection is important part after feature extraction. The significant method used for feature selection is floating sequential search and it is widely used in arrhythmia classification. Using the technique of backward and forward search most of the useful characteristics are obtained and it obtained better results by using eight selected feature of ECG signal. The number of features and its accuracy is important when analyzing the different options for feature selection. To bring various advantages in classification method different feature selection techniques and algorithms are used which reduce the computational cost and whose target is to use less number of features to built final structure[30]. For feature selection, filter feature selection method and wrapper feature selection method are used which analyzes around 200 different dimensions for ECG data classification. Weighted LD model supports the wrapper feature selection method and this procedure works closely with forward and backward searching method. Here support vector machine has an important role related to filter technique as it fetches mutual information related to combination with the help of ranking approach. For feature selection the study of R-R interval and for T wave, length and amplitude are analyzed. Here the characteristics of mutual information is the leading implementation of feature selection. The ultimate goal of feature selection is to analyze the heart disease and the proper diagnosis of the disease by using machine learning approaches. Several modern state-of-art procedures such as particle swarm optimization (PSO), Genetic Algorithms(GA) makes the classification result accurate which trend to detect heart arrhythmia and gives the promising results[31].

V. TECHNIQUES USED

Several algorithms and procedures are used to create the mathematical model of the classification technique, from there the automated systems can be developed.

5.1 Artificial Neural networks (ANN)

For arrhythmia classification Artificial Nural Network(ANN) is an important tool. Probabilistic Neural Network (PNN) and Multilayer Perceptron are the key features for the said classification. Using PNN multi-classification is done and with the support of that the set of data points which will create the data samples from the given set data points. If we analyze PNN we will realize that the basic building block of the structure of PNN consists of sub-networks, and it splits the ECG signal data with other layers of sub-networks.

Input layer, hidden layer and output layer are the three layers of PNN, This measures the gap between the training input vector and generates the vector which identifies the vector and it measures how close the input and training input vector works[32].

5.2 Classifier Models

There are several techniques used to classify the input data generated by ECG signal.

5.2.1. Linear discriminants (LD)

It is a statistic method where different discriminant functions are used. The training set of data is used to estimate the functions and it is linearly divide the feature vector where bias and weight vector are the two components which adjust feature vector. These are the classifiers used the methods and schema and for ECG arrhythmia detection it follows the standard of Association for the Advancement of Medical Instrumentation(AAMI). For a arrhythmia analysis imbalance of training set may arise problem which is difficult to solve by SVM and by using linear discriminant we can overcome it. The Support vector machine(SVM) and Multilayer perceptron (MLP) directly support the LD classifier[33]. LD divide the feature space into separate class and categories which uses different set of hyper planes.

5.2.2. Neural networks (NN)

Neural network implement logistic discriminants where by applying few conditions upon feature space, it extracts feature data for class distribution which has direct relation with exponential distribution. This features includes many of common distributions such as binomial, Gaussion, Poissan and Bernoulli. This model uses the optimization technique based on iterative numerical and here no direct optimization model is possible. Gradient-Descent algorithm helps optimization of parameter and adaptive learning rate upon the training sample, and the training ends when consecutive iteration has no longer effect on it. With several classifications the concluding classification is attained by selecting the class with maximum probability gained [34].

5.3. Support vector machine (SVM)

In the field of machine learning, classification and feature extraction, Support Vector Machine(SVMs) are very powerful and well known tool. It has the power of high reservoir computing with the support of logistic regression of dimensional data. Different data mining applications and pattern recognition like image classification, text categorization, 3D object detection, phoneme recognition, bioinformatics are supported by SVMs and for arrhythmia classification these properties are useful. The SVM uses different classification problems like binary (two-class), multi-class etc. The multiclass problem is neither unique or straightforward and several methods are there to support SVM classifier for multiclass classification [35]. SVM works with the ECG classifier training data set and it checks it is not linearly separable, the desired classifier many not perform well and the generalization ability may not touch at pick. The performance depends on linear separability where the unique input space is collaborated with dot product space which is often called feature space[36].

5.4 Reservoir computing (RC)

The reservoir computing models works dynamically and its aim is to produce the time series signals. It works into two parts, in first phase it represents the signal through a non-adaptable active reservoir and the second phase it makes a active readout through the reservoir. For heartbeat classification it uses a straightforward nonlinear active element which delivers to a delayed feedback where using a binary random mask every peak of ECG signal is sampled and the sampled data is held during the delay times. The logistic regression is accomplished by learning process of the ECG signal classification. This technique uses a huge dataset to classify the model and at the same time the computational cost of this process is low, so the hardware implementation of is system is profitable[37].

5.5 Supportive Techniques

In the previous section several impotent techniques are discussed for ECG signal processing and based on that the arrhythmia classifications are done. Several another important techniques are used to make the classification more logical and accurate.

5.5.1. Clustering

In data classification and analysis the clustering analysis technique is widely used and several clustering techniques are EM algorithms, hierarchical, k-means, Self organization Maps. Among the clustering algorithms k-means is mostly used and it works in two stages iterative algorithms where in the first stage it reduces the sum point-to-centroid length and in the second phase it sums all k clusters. In the clustering first the cluster batches are updated and in each iteration it reads the nearest cluster centroid and from there it recalculates the another cluster centroid. The second part is the automatic update where the clustering points are independently reassigned to decrease the sum of distances and the cluster centroids are recompiled after each reassignment. Using k-means clustering the ECG signal classification is done, and the popular Euclidean method is used under k-means method[38].

5.5.2. Decision Trees

For arrhythmia heart beat recognition and classification, decision tree has an important role as the analysis based on MIT-BIH Arrhythmia database shows that, There is a decision support system works and implements a tree like model or graph which take the decision of certain classification and their probable consequences. The Bagged Decision Tree (BDT) is a special type of classifier which is used for group learning for the recognition of arrhythmia heart beat. Here the decision tree contains three different types of nodes which are decision nodes, chance nodes and end nodes respectively. For operation management and operation research decision trees are commonly used and from there decision on classification can be taken[39].

5.5.3. Hidden Markov models

Hidden Markov modeling (HMM) is an unique approach to describe and analyze ECG arrhythmia. This technique is followed for arrhythmia analysis since middle of 1970 and it is mainly used to convert routine speech recognition to model speech waveform. Different classifications of ventricular arrhythmias are classified by analyzing and detecting the QRS complex and the RR intervals (as mentioned in Fig. 2). It combines statistical & structural knowledge where using training data model parameters are predicted from training data which uses iterative re-estimated algorithms. These algorithms suggests to implement supra ventricular ECG arrhythmia analysis [40].

Retrieval Number: B3548079220/2020©BEIESP
DOI:10.35940/ijrte.B3548.079220
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

5.5.4. Rules-based models

In rules based modeling approach the model condition is brief and it implements the implicit model specifications where specific and specialized rule based models use powerful software to implement ECG arrhythmia analysis and classification.

The rule based model includes a set of rules that can be executed by general purpose replication and examination of ECG training set data. The set of rules supported by Rules based models contributes the power for feature extraction and classification[41].

5.5.5. Optimum-path forest

Optimum–path Forest is a graph based technique for pattern recognition. The analytical power of Optimum–path Forest(OPF) is more than traditional technique of supervised pattern recognition like SVM, Multilayer Perceptrons, ANN etc., when the execution time and accuracy is considered. The patterns of the classification generally described by feature vectors which is obtained from samples of the available data set. The pattern recognition of OPF has two different fundamental problems, the first one is unsupervised classification where the natural groups or the clusters are compiled by samples and by homogeneous pattern. The second one is supervised one, where each sample is classified in one of the available classes and that class is implemented with specific labels. Here the approach of supervised learning is used[42].

5.5.6. Conditional random fields

This framework builds probabilistic models to create the segment and also label the sequence of data. It also avoids the primary restrictions of utmost entropy Markov models and also other distinct Markov models which are basically directed graph model. The basic limitations of different Markov models like Hidden Markov models (HMMs), maximum entropy Markov models (MEMMs) are overcome by using directed graph models and iterative parameter estimation algorithms which supports synthetic and natural-language data. Information extraction, syntactic disambiguation, topic segmentation and the tagging of part of speech(POS) are the important criteria supported by conditional random fields for the computational linguistics which support the classification and feature extraction of the ECG data[43].

5.5.7. Nearest neighbors

Classification and feature extraction needs data extraction from nearest neighbors and various algorithms are used to deal this approach. k-Nearest Neighbor is the popular and well known algorithm used for non-parametric classification where no former information is necessary to predict the class label. The weight nearest neighbor, fuzzy nearest neighbor, feature selection methods, different genetic algorithm based nearest neighbor and classifier algorithms etc., are the various techniques used in nearest neighbor applications and these classification techniques are the organized initiatives for various classifications for feature extraction[44].

VI. CONCLUSION

This paper describes the feature extraction and classification of ECG signal. Most of the researchers use publicly available MIT-BIH database for the ECG arrhythmia classification. The MIT-BIH database has certain limitations as the database is very unbalanced and the extracted features from the database sometimes produce erroneous results. So updated version of the database is always required to produce correct classifications. Continuous research is going on and the effort is there to reduce the diversity and complexity where the size of the database will increase and this will give more classifications patterns which will maintain typical evaluation protocols . Several popular machine learning techniques and algorithms such as PCA, LDA ,ICA are used to create the classification framework. The goal of the research work is to make fully automated heartbeat classification and from there the physicians can detect arrhythmia more accurately.

REFERENCES

1. Qiao Li, Cadathur Rajagopalan,Gari D. Clifford, "Ventricular Fibrillation and Tachycardia Classification Using a Machine Learning Approach" *IEEE Transactions on Biomedical Engineering*, Vol. 61, No. 6, pp 1607-1613, June 2014.
2. B Pyakillya, N Kazachenko and Nikhailovsky, "Deep Learning for ECG Classification" , *IOP Conf. Series: Journal of Physics: Conf. Series 913 (2017) 012004*, doi :10.1088/1742- 6596/913/1/012004 , pp 2-5.
3. C. Saritha, V. Sukanya, Y. Narasimha Murthy, "ECG Signal Analysis Using Wavelet Transforms" , *Bulg. J. Phys,* PACS number: 87.85.J; 02.30.Nw, 16 February 2008 , pp 68-77.
4. Rodrigo V. Andreão, Bernadette Dorizzi, Jérôme Boudy, "ECG Signal Analysis Through Hidden Markov Models" , *IEEE Transactions on Biomedical engineering,* Vol 53, No. 8, August 2006, pp 1541-1549.
5. Priyarani S. Jagatap and Rupali R. Jagtap, "Electrocardiogram (ECG) Signal Analysis and Feature Extraction: A Survey", *International Journal of Computer Sciences and Engineering,* Vol.-2(5), May 2014, pp 1-3.
6. P. S. Hamilton, W. J. Tompkins, "Quantitative Investigation of QRS Detection Rules Using MIT/BIH Arrhythmia Database", *IEEE Transactions on Biomedical Engineering,* Vol. 31, No.3, March 2007, pp. 1157-1165.
7. V.K.Srivastava, Dr. Devendra Prasad, "Dwt - Based Feature Extraction from ecg Signal ", *American Journal of Engineering Research (AJER),* Volume-02, Issue-03,2013, pp 44-50.
8. Abhinav Vishwa, Mohit K. Lal, Sharad Dixit, Pritish Vardwaj, "Clasification Of Arrhythmic ECG Data Using Machine Learning", *International Journal of Artificial Intelligence and Interactive Multimedia*, Vol 1, No. 4, DOI: 10.9781/ijimai.2011.1411, 2011,pp 68-71.
9. A. Selcuk Adabag, Barry J. Maron,Evan Appelbaum, Caitlin J. Harrigan, "Occurrence and Frequency of Arrhythmias in Hypertrophic Cardiomyopathy in Relation to Delayed Enhancement on Cardiovascular Magnetic Resonance", *Journal of the American College of Cardiology,* Vol. 51, No. 14, 2008, pp 1369-1374.
10. P.Chazal, M. O'Dwyer, and R. B. Reilly, " Auto-matic classification of heartbeats using ECG morphology and heartbeat interval features", *IEEE Trans. Biomedical Engineering*, 2004, pp 1196–1206.
11. R.G. Mark, P.S. Schluter, G.B. Moody, "An annotated ECG database for evaluating arrhythmia
12. Detectors", *4th Annual Conf. IEEE EMBS. Long Beach,* Frontiers of Engineering in Health Care–1982, pp. 205-210.
13. V.Mahesh, A. Kandaswamy, C. Vimal, B. Sathish, "ECG arrhythmia classification based on logistic model tree", *J. Biomedical Science and Engineering 2",* Vol.2, No.6, 2009,pp 405-41.
14. American National Standard for Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms. AMI/ANSI Standard EC57: 2012. 18 December 2012.
15. Jinkwon Kim, Se Dong Min and Myoungho Lee, "An arrhythmia classification algorithm using a dedicated wavelet adapted to different subjects", *BioMedical Engineering OnLine,* 2011 , http://www.biomedical-engineering- online.com/content/10/1/56.
16. George B.Moody and Roger G.Mark, " The MIT-BIH Arrhythmia database on CD-ROM AND Software for use with-it", *IEEE conference,* 0276-6574/91/00000/0185$01.00 , 1991, pp 185-188.
17. K.L Ripley and G.C.Oliver , "Development of an ECG database for arrhythmia detector evaluation" Computers in cardiology,1997. pp 203-209.

18. Kiran Kumar Patro, P.Rajesh Kumar, "Machine Learning Classification Approaches for Biometric Recognition System using ECG Signals", *Journal of Engineering Science and Technology Review*, doi:10.25103/jestr.106.01 , December 7 , 2017, pp 1-8.
19. M Murugappan, Reena Thirumani, Mohd Iqbal Omar,Subbulakshmi Murugappan, "Development of Cost Effective ECG Data Acquisition System for Clinical Applications using LabVIEW", *IEEE 10th International Colloquium on Signal Processing & its Applications*, March 2014, pp 100-105.
20. V. Mondéjar-Guerraa,J. Novoa, J. Roucoa, M.G. Penedoa, M. Ortegaa, "Heartbeat classification fusing temporal and morphological information of ECGs via ensemble of classifiers", *Biomedical Signal Processing and Control,* 1746-8094/© 2018 Published by Elsevier Ltd. , pp 41-48.
21. G.B. Moody, R.G. Mark, "The impact of the MIT-BIH arrhythmia database", IEEE Engineering, Medical ,Biol. Mag. 20 (3) http://dx.doi.org/10.1109/51.932724 , 2001, pp 45–50.
22. E.J. da S. Luz, W.R. Schwartz, G. Cmara-Chvez, D. Menotti, "ECG-basedheartbeat classification for arrhythmia detection: a survey", *Computer Methods and Programs in Biomedicine,* 2016 , pp 144-164, ://dx.doi.org/10.1016/j.cmpb.2015.12.008.
23. Kiran Kumar Patro, Dr.P.Rajesh Kumar , "De-Noising of ECG raw Signal by Cascaded Window based Digital filters Configuration", *IEEE Power, Communication and Information Technology Conference,* October 2015.
24. D. Jeyarani, T. J. Singh, "Analysis of Noise Reduction Techniques on QRS ECG Waveform - by Applying Different Filters", *IEEE conference on Recent Advances in Space Technology Services and Climate Change (RSTSCC),* Chennai, 2010.
25. R. de F. Dalvi, G. T. Zago, R. V. Andreão, "Heartbeat classification system based on neural
26. networks and dimensionality reduction", *Research on Biomedical Engineering*, Vol. 32, No.4 Rio de Janeiro Oct./Dec. 2016 Epub Jan 12, 2017.
27. Wen Zhu, Nancy Zeng, Ning Wang, "Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical" , *Health care Life Sciences*, NESUG 2010, pp 1-9.
28. S. Thulasi Prasad, S. Varadarajan, "ECG Signal Analysis: Different Approaches" , *International Journal of Engineering Trends and Technology* , Vol.7, No. 5, Jan 2014, pp 212-216.
29. G. Doquire, G. de Lannoy, D. Franc¸ois, M. Verleysen, "Feature selection for inter patient supervised heart beat classification", Computational Intelligence and Neuroscienc, 2011, pp 1–9.
30. C Alexakis, HO Nyongesa, R Saatchi, ND Harris, C Davies, C Emery, RH Ireland, SR Heller,"Feature Extraction and Classification of Electrocardiogram (ECG) Signals Related to Hypoglycaemia" , *Computers in Cardiology*, 2003, pp537−540.
31. L. Kanaan, D. Merheb, M. Kallas, C. Francis, H. Amoud, P.Honeine, "PCA and KPCA of ECG signals with binary SVM classification", *IEEE Workshop on Signal Processing Systems*, 2011, pp. 344–348.
32. E.E.M.Bolumu, "Feature Selection for ECG Beat Classification using Genetic Algorithms with A Multi-objective Approach", *26th Signal Processing and Communications Applications Conference,* 2-5 May 2018.
33. A.Elsayyad, M.Al-Dhaifallah,A.M.Nassef," Feature Selection for Arrhythmia Diagnosis using
34. Relief-F Algorithm and Support Vector Machine", *14th International Multi-Conference on Systems, Signals & Devices (SSD),* 28-31 March, 2017, pp 462-468.
35. 32. Y.H.Hu, W.J Tompkins,Q. Xue, "Artificial Neural Network for ECG Arrhythmia Monitoring", 0-7803-0559-0 /92 33.00 8, IEEE 1992, pp 987-992.
36. P. de Chazal, R. B. Reilly, "Automatic Classification of ECG Beats using Waveform Shape and Heart Beat Interval Features", *International Conference on Acoustics, Speech and Signal Processing,* 2003, pp 269-272.
37. 34. The CSE Working Party, "Recommendations for measurement standards in quantitative Electrocardiography", *European Heart Journal,* 1985 , pp 815-825.
38. Mi Hye Song, Jeon Lee, Sung Pil Cho, Kyoung Joung Lee, and Sun Kook Yoo, "Support Vector Machine Based Arrhythmia Classification Using Reduced Features", *International Journal of Control, Automation, and Systems*, vol. 3, no. 4, December 2005 , pp. 571-579.
39. Narendra Kohli, Nishchal K. Verma, "Arrhythmia classification using SVM with selected features", *International Journal of Engineering, Science and Technology,* Vol. 3, No. 8, 2011, pp. 122-131.
40. M.A. Escalona-Moran, M.C. Soriano, I. Fischer, C.R. Mirasso, Electrocardiogram classification using reservoir computing with logistic regression, *IEEE Journal of Biomedical and Health Informatics*, Vol. 11, No. 4, December 2012, pp 122-131.
41. Manpreet Kaur, A.S.Arora, " Unsupervised Analysis of Arrhythmias using K-means Clustering",
42. *International Journal of Computer Science and Information Technologies*, Vol. 1 (5) , 2010, pp 417-419.
43. Ahmet Mert, Niyazi Kilic, Aydın Akan, "ECG Signal Classification Using Ensemble Decision Tree", *Journal of Trends in the Development of Machinery and Associated Technology,* Vol. 16, No. 1, 2012, pp. 179-182.
44. D.A.Coast, R.M.Stren,G.G.Cano,S.A.Briller, "An Approach to Cardiac Arrhythmia Analysis Using Hidden Markov Models" , *IEEE Transactions on Biomedical Engineering*, Vol 37, No. 9, September 1990, pp 826-836.
45. Leonard A. Harris,Chang-Shung Tung,James R. Faeder,Carlos F. Lopez,William S. Hlavacek, "Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems", *Wiley Interdiscip Rev Syst Biol Med. Author manuscript*; January 01. 2015.
46. Joao Paulo Papa, Alexandre Xavier Falcao, "Optimum-Path Forest: A Novel and Powerful Framework for Supervised Graph-based Pattern Recognition Techniques", pp 41-48.
47. 43.John Lafferty,Andrew McCallum,Fernando C.N. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data", *WhizBang! Labs–Research,*
48. *4616 Henry Street, Pittsburgh, PA 15213 USA* , June 2001.
49. Rashmi Agrawal, "Extensions of k-Nearest Neighbor Algorithm" ,*Research Journal of Applied Sciences, Engineering and Technology* , July 2016, pp 24-29.

AUTHORS PROFILE

Mr. Sumanta Kuila is currently working as an Assistant Professor in the Deptartment of Computer Science & Engineering, Haldia Institute of Technology, Haldia-721657, India. He has 16 years of experience in the field of Computer Science & Engineering. He is doing Ph.D in Computer Science & Engineering from Amity University Uttar Pradesh, Lucknow campus. He did his masters (M.E) from west Bengal University of Technology in 2005. He has done 6 Technical certifications (Sun & Oracle) on Java/J2EE & Web Technology. His current research interest includes Machine Learning, Data Analytics, Big Data, Internet of Things etc.

Mr. Sayandeep Maity is currently a student of Haldia Institute of Technology, Haldia-721657, India, pursuing Bachelor of Technology in Computer Science and Engineering. He completed his Secondary and Higher Secondary from St. Xaviers School, Haldia in 2015 and 2017 respectively. He has 3 Technical certifications (IBM) on Python and Data Science. His current research interest includes Machine Learning, Data Science, Data Analysis, Artificial Intelligence, etc.

Suman Kumar Mal is currently pursuing his Bachelor's Degree in Computer Science and Engineering from Haldia Institute of Technology, Haldia-721657. He has completed his secondary and higher secondary school from Holy Cross School, Bokaro-827010. He has a Professional Certification in Data Science from IBM.His research interests are in the field of Data Science, Machine Learning and Decentralized Applications.

Dr. Subhankar Joardar is presently Professor & Head in the Department of Computer Science and Engineering, Haldia Institute of Technology, Haldia-721657, India. He received his Ph.D degree from Birla Institute of Technology, Mesra, Ranchi, India in 2016. He did his masters (M. Tech and MCA) both from BIT, Mesra, Ranchi in 2009 and 2002 respectively. He has published more than 14 technical papers in the referred journals/conferences. He has served as Organizing Chair of international conference (ICITAM 2017). He is also Program Committee member of International Conferences (ICCDC 2019). He is a member of Computer Society of India. His current research interests include Swarm intelligence, Routing in Mobile Ad Hoc Networks, Machine Learning.