Intelligent Arrhythmia Diagnostics System

Dipti D. Patil[1], Dhanashri Patil[2] , Sharlin Pandharpatte[3], Ruta Dhekane[4], Trupti Mohol[5] and Dr. V.M.Wadhai[6]

[1] PhD Student of Sant Gadgebaba Amravati Uiversity, PG Computer Science and Engg. Department, Amravati, India

[2,3,4,5] Student, MIT College Of Engineering, Computer Department, University of Pune, Pune,India

[6] Principal and Professor MIT College of Engineering, Pune, India

Abstract

Changes in the normal rhythm of a human heart may result in different cardiac arrhythmias, which may be immediately fatal or cause irreparable damage to the heart sustained over long periods of time. The ability to automatically identify arrhythmias from ECG recordings is important for clinical diagnosis and treatment. This paper discusses about how the intelligent signal processing method is used to detect and extract features of ECG signals and finally clustering these features to predict the type of arrhythmia by using data stream mining algorithms. The system discussed in this paper is integrated with Wireless Sensors Network (WSN) through which ECG signal of patient can be sensed and analyze in real time to diagnose presence of arrhythmia and classify them in 8 different types. The system is thus helpful to doctors for evaluation and monitoring of patient's heart condition at any time and any location.

Keywords: *Data Stream Mining, Preventive Health Care, Feature Extraction, ECG Signal Preprocessing*

1. Introduction

In today's world death rate due to heart diseases is a major threat to human beings. Arrhythmia is such a type of heart disease which is increasing rapidly and major deaths in cardiac patients are due to arrhythmia. An arrhythmia is any disorder of heart rate. The time interval between successive heart beats is called as rhythm and when the rhythm is not regular it is called as arrhythmia (a+ rhythm=arrhythmia).Symptoms of arrhythmia are fast or slow heart beat, skipping beats, chest pain etc that can be seen in the early stage.

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, it is the process of finding correlations or patterns among large number of collected data. It is the technique which learns from behavior of data. Data mining is playing vital role in many applications especially in healthcare. The data generated in healthcare system is vast and complex, so it becomes very difficult to process it. With the help of data mining one can transform this enormous amount of data into useful information for decision making.

This paper presents an online system which detects arrhythmia and predicts its type in real time. As arrhythmia is caused due to abnormal heart rate it can be easily diagnosed with the help of ECG signals. This system initially focuses on pre-processing and feature extraction. Then data mining algorithm is used to learn from the collected data of ECG and forms the rules which help to predict the type of arrhythmia.

2. Related Work

There are several methods for automatic detection and classification of cardiac arrhythmia like Artificial Immune Recognition system, Neural Network, Fuzzy Neural Network, Data Mining etc.

Artificial Immune Recognition system is a technique which uses a well-known neural network architecture named multi-layered perceptron (MLP) with back propagation training algorithm, and a new

fuzzy clustering NN architecture (FCNN) for early diagnosis. A fuzzy neural network was implemented using a multithreading approach for detection of atrial fibrillation, bigeminy, and normal sinus rhythm in the MIT-BIH Arrhythmia Database. The input data to the neural network consisted of nine inputs: Seven contiguous RR intervals, their average and their standard deviation. Wavelet-based algorithm is used for arrhythmia discrimination which analyses the Electrocardiograph (ECG) signal by using the continuous wavelet transform and its rule in different scales of variation and it can automatically distinguish arrhythmia. This method based on Fuzzy neural network (FNN) is developed to create fuzzy membership functions for classification of cardiac arrhythmia [4]

The proposed system is based on Data Stream Mining techniques which consist of different algorithms for classifying arrhythmia into seven types. They are namely Normal Beat(NB),Left Bundle Branch Block Beat(LBBB), Right Bundle Branch Block Beat(RBBB),Premature Ventricular Contraction(PVC),Fusion of ventricular and normal beat(FUSION), Atrial Premature Contraction(APC) and Paced Beat(PACE) [4].

Earlier there was K-mean algorithm used for stream data clustering which use single phase model but this algorithm was not able to detect the priority of data and it gave same priority to recent and historic data. In addition to this, K-mean algorithm was incapable of identifying arbitrary shapes of cluster and noise in signals. This algorithm also requires prior knowledge of no of clusters determined by K. In order to overcome these drawbacks Aggarwal et al. proposed 2-phase scheme which processes the raw data in micro-clusters in online mode and gives summary which is used by offline components. This design leads to Clu-Stream system. Another basic clustering algorithm called Density based clustering algorithm which is more general as it can identify arbitrary shapes clusters, handle noise and scan data only once so it can be applied to large volume of data stream. In this paper, various time domain feature are extracted from ECG and HRV signals then obtained data is clustered with help of Data Mining algorithm called D-stream algorithm

3. SIGNIFICANCE OF THE SYSTEM

The main concern in the proposed system is to detect the arrhythmia and its type. Though there are various types of diseases, the worldwide results [13] shown in figure 1 have showed that more death rates are caused due to the cardiovascular diseases and in that arrhythmia results are markable.

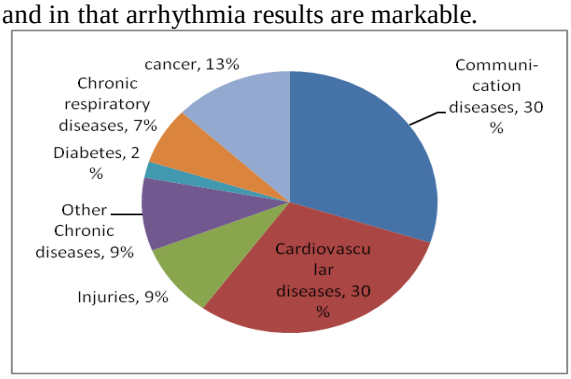

Fig 1. Principle causes of death worldwide [15]

4. PROPOSED SYSTEM

WSN is a network of small wireless sensor devices, deployed in an ad-hoc fashion to cooperate on sensing a physical phenomenon. It is collaborative network of large number of loosely connected nodes.WSN is a distributed system tasked to sample environment for sensory information. It combines and compresses small fragments of possibly inaccurate data. The WSN is built of "nodes" – from a few to several hundreds or even thousands, where each node is connected to one (or sometimes several) sensors. Each such sensor network node has typically several parts: a radio transceiver with an internal antenna or connection to an external antenna, a microcontroller, an electronic circuit for interfacing with the sensors and an energy source, usually a battery or an embedded form of energy harvesting.

The system shown in figure 2 takes the ECG signal and historical data of patient to predict the disease called arrhythmia and its type. The ECG signal which is taken as an input is passed through various stages like Feature Extraction, Pre-processing, Clustering of data and the formation of rules using data mining algorithms.

The system consists of two parts, online and offline processing of data. The offline mode consist of previously formed rules which are obtained from patients having different characteristics such as age, gender or any other symptoms. In online mode the continuous ECG signal is taken as input and it is passed to pre-processing module. In this module the noise is removed from ECG signal, then this filtered ECG signal is given to feature extraction module where all time domain features of signal are extracted like R-R interval, Root Mean Square Successive Difference(RMSSD) in 8 R-R

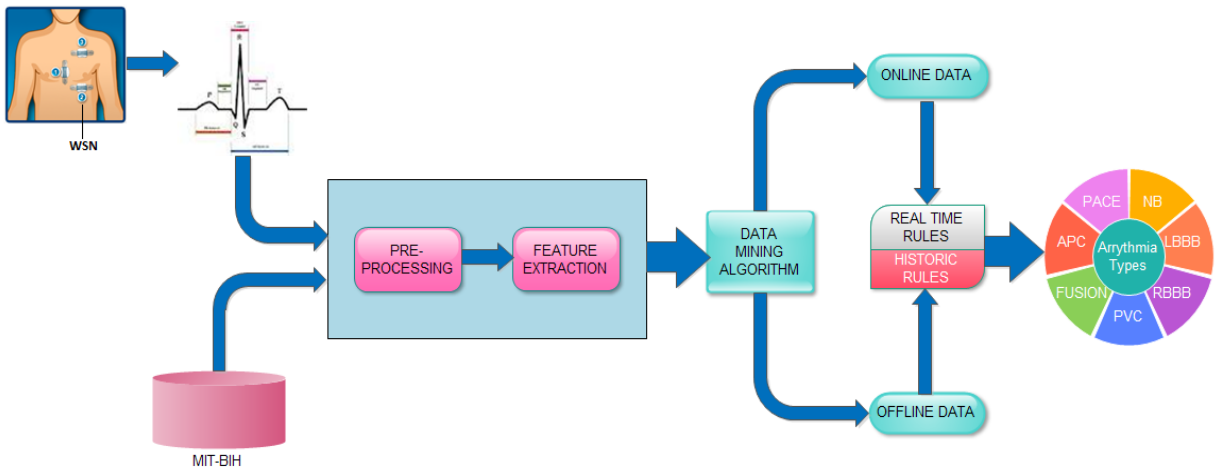

Fig 2. Proposed ECG Arrhythmia Dignostic SystemSystem

intervals, Standard Deviation(SDNN) of 8 R-R intervals, Fractal dimension, Lyapunov Exponent, Hurst Exponent etc. The data obtained from feature extraction module is given as an input to data mining algorithm. Then the data mining algorithm forms the clusters of data having similar pattern. These obtained clusters are used to form rules. The rules obtained in online and the offline module are used together to detect the type of arrhythmia.

5. SIGNAL PREPROCESSING

In general, the aim of preprocessing steps is to improve the general quality of the ECG for more accurate analysis and measurement. Noises may disturb the ECG to such an extent that measurements from the original signals are unreliable. The main categories of noise are: low frequency base line wander (BW) caused by respiration and body movements, high frequency random noises caused by mains interference (50 or 60Hz) and muscular activity and random shifts of the ECG signal amplitude caused by poor electrode contact and body movements.

6. ECG FEATURE EXTRACTION

The multi-resolution wavelet transform is one of the main methods of ECG feature extraction [2]. ECG signals from Modified Lead II (MLII) are chosen for processing as the peaks of the R waves in signals from the ML II lead have the largest amplitudes among other leads. The result of applying two Daubechies wavelet filters (D4 and D6) of different length on the signal is compared. The wavelet filter with scaling function more closely similar to the shape of the ECG signal achieved better detection [15]. Db wavelet family is similar in shape to QRS complex and their energy spectrums are concentrated around low frequencies. The signal is approximated by omitting the signal's high frequency components. The ECG signal and the details for eight wavelet scales are used which are scaled for better illustration. The detection of R peaks is shown in Fig-2.

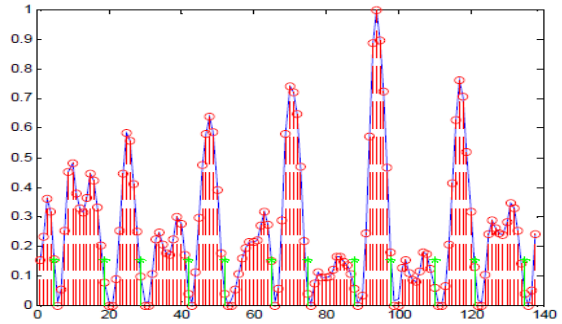

Fig 3. [2] R peak detection

7. TIME DOMAIN ANALYSIS FOR ECG

There are many methods to evaluate the variations in the heart rate. Time domain measures are the simplest to perform. With these methods either the heart rate at any point in time or the intervals between successive normal complexes are determined. In a continuous ECG record, each QRS complex is detected, and the so-called normal-to-normal (NN) intervals (that is all intervals between adjacent QRS complexes resulting from sinus node depolarization), or the instantaneous hear rate is determined, Simple time-domain variables that can be calculated include the mean NN interval, the mean heart rate, the difference between night and

day heart rate, etc. Other time-domain measurements that can be used are variation in instantaneous heart rate secondary to respiration, tilt, Valsalva manoeuvre, or secondary to phenylephrine infusion [2]. These differences can be described as either differences in heart rate or cycle length. From the original R-R intervals, four standard measures parameters used in this work are:

1. The standard deviation of the NN intervals (SDNN).
2. The standard deviation of differences between adjacent NN intervals (SDSD)
3. The root mean square successive difference of intervals (RMSSD).
4. The number of interval differences of successive NN intervals greater than 50ms (NN50) used for the proportion derived by dividing NN50 by the total number of NN intervals (pNN50). [2]

8. COMPARATIVE STUDY OF DATA CLUSTERING ALGORITHMS

8.1. Centroid-based clustering:

In centroid-based clustering, clusters are represented by a central vector, which may not necessarily be a member of the data set. When the number of clusters is fixed to k, k-means clustering gives a formal definition as an optimization problem: find the k-cluster centers and assign the objects to the nearest cluster center, such that the squared distances from the cluster are minimized.Most k-means-type algorithms require the number of clusters - k- to be specified in advance, which is considered to be one of the biggest drawbacks of these algorithms. Furthermore, the algorithms prefer clusters of approximately similar size, as they will always assign an object to the nearest centroid. This often leads to incorrectly cut borders in between of clusters

8.2. Connectivity based clustering (hierarchical clustering): Connectivity based clustering, also known as hierarchical clustering, is based on the core idea of objects being more related to nearby objects than to objects farther away. As such, these algorithms connect "objects" to form "clusters" based on their distance. A cluster can be described largely by the maximum distance needed to connect parts of the cluster. At different distances, different clusters will form, which can be represented using a dendrogram, which explains where the common name "hierarchical clustering" comes from: these algorithms do not provide a single partitioning of the data set, but instead provide an extensive hierarchy of clusters that merge with each other at certain distances. In a dendrogram, the y-axis marks the distance at which the clusters merge, while the objects are placed along the x-axis such that the clusters don't mix.

While these methods are fairly easy to understand, the results are not always easy to use, as they will not produce a unique partitioning of the data set, but a hierarchy the user still needs to choose appropriate clusters from. The methods are not very robust towards outliers, which will either show up as additional clusters or even cause other clusters to merge.

8.3. Density Based Clustering:

Density-based clustering has been long proposed as another major clustering algorithm [5,6]. It is found [1]that the density based method a natural and attractive basic clustering algorithm for data streams, because it can find arbitrarily shaped clusters, it can handle noises and is an one-scan algorithm that needs to examine the raw data only once. Further, it does not demand a prior knowledge of the number of clusters *k* as the *k*-means algorithm does.

9. ARRHYTHMIA CLASSIFICATION

In [2] the author has described the features that are needed to classify the arrhythmia .The following table gives the detailed description:

Types of Arrhythmia [4]:

1) Normal Beat (NB)
2) Left Bundle Branch Block Beat (LBBB)
3) Right Bundle Branch Block Beat (RBBB)
4) Premature Ventricular Contraction (PVC)
5) Fusion of Ventricular and Normal Beat (FUSION)
6) Atrial Premature Contraction (APC)
7) Pace Beat (PACE)

10. DATA STREAM MINING

"Data mining is the process of discovering meaningful patterns and relationships that lie hidden within very large databases"[15]. Data stream, an ordered sequence of instances is read only once or a few times with limited computing and storage capabilities. Mining of such data streams is a unique opportunity and a challenging task as opposed to the finite, statically stored data sets. This is a process of extracting knowledge structures from continuous, rapid data records. Analysing data streams helps in

Table 1. Comparison & Results using Weka Clustering Tool [3]

Name	Number of clusters	Cluster Instances	Number of Iterations	Within clusters sum of squared errors	Time taken to build model	Log likelihood	Unclustered Instances
K-Means Algorithm	2	0:254(42%) 1:346(58%)	4	2016.6752520938053	0.08 Seconds	-	0
Hierarchical Clustering	2	0:599(100%) 1:1 (0%)	-	-	1.16 Seconds	-	0
Density based Clusters	2	0:239(40%) 1:361(60%)	4	2016.6752520938053	0.06 Seconds	-22.04211	0

applications such as trade surveillance for security fraud and money laundering, network traffic monitoring, sensor network data analysis, power consumption measurement, and dynamic tracing of stock fluctuations.

Table 2 .Effective features that are selected to distinguish each arrhythmia from others**. [2]**

Features	NB	LBBB	RBBB	PVC	FUSION	APC	PACE
Mean	Yes		yes		yes		yes
RMSSD			yes				yes
SDNN		yes					
SDSD		yes		yes	yes	yes	yes
pNN50			yes	yes		yes	
pNN10	yes						
pNN5					yes		yes

Besides querying data streams, another important application is to mine data streams for interesting patterns or anomalies as they happen. Another related issue is privacy. The data streams may contain personal data that need to be protected.

Some anonymity or depersonalization mechanism needs to be applied to the data as it streams in, before the data is provided to the mining algorithms.

Where Data Stream Mining is the process of extracting knowledge structures from continuous, rapid data records. A data stream is an ordered sequence of instances that in many applications of data stream mining can be read only once or a small number of times using limited computing and storage capabilities. Data stream mining can be considered a subfield of data mining, machine learning, and knowledge discovery.

D-stream mining algorithm is one of the data stream mining algorithms. The ECG signals which are taken as input are continuous in nature, thus the input data is in the form of a stream so the system uses D-stream mining algorithm.

11. CONCLUSION

In this paper the system which is able to differentiate between the normal and abnormal ECG using intelligent techniques is discussed. It predicts whether a person is suffering from arrhythmia or not and if yes then it detects the appropriate type. Due to continuous monitoring of the heart of the cardiac patient, the system helps in early detection of the disease. This helps the doctor to provide better diagnosis and treatment to the patient. It also provides mobility to the patient who requires continuous monitoring in the hospital. It reduces the death rate as arrhythmia is one of the leading causes of death in cardiac patients. In case of emergencies, like stroke or cardiac arrest, the system will detect the changes in ECG and immediate treatment can be provided to the patient. Thus it will reduce the overall death rate in the cardiac patients.

The system is a real time application, which uses data stream mining algorithms like D-stream those provide dynamic processing of the data in real time bringing the uniqueness in a system.

Acknowledgment

This is the research work associated to the PhD work going on Under Sant Gadge Baba Amravati University (SGBAU) and PG Department of Computer Science and Engineering, Amravati. My sincere thanks to the centre Head Dr. V. M. Thakre and the faculty for their extensive support.

References

[1] Yixin Chen, Li Tu", Density-Based Clustering for Real-Time Stream Data",in KDD'07, August 12–15, 2007, San Jose, California, USA, ACM 978-1-59593-609-7/07/0008

[2] B. Anuradha, K. Suresh Kumar and V. C. Veera Reddy " Classification Of Cardiac Signals Using Time Domain Method" in ARPN Journal of Engineering and Applied Sciences , VOL. 3, NO. 3, JUNE 2008 ,ISSN 1819-6608

[3] Manish Verma, Mauly Srivastava, Neha Chack, Atul Kumar Diswar, Nidhi Gupta" A Comparative Study of Various Clustering Algorithms in Data Mining". In GLNA Institute of Technology, Mathura , Vol. 2, Issue 3, May-Jun 2012, pp.1379-1384

[4] Masih Tavassolia, Mohammad Mehdi Ebadzadeha, Hamed Malek b "Classification of cardiac arrhythmia with respect to ECG and HRV signal by genetic programming ",in Canadian Journal on Artificial Intelligence, Machine Learning and Pattern Recognition Vol. 3 No. 1, January 2012,Iran

[5] J. Sander, M. Ester, H. Kriegel, and X. Xu."Density-based clustering in spatial databases: The algorithm gdbscan and its applications.". Data Min.Knowl. Discov. 2(2):169–194, 1998.

[6] S. Subramaniam, T. Palpanas, D. Papadopoulos,V. Kalogeraki, and D. Gunopulos. "Online outlier detection in sensor data using non-parametric models". In Proc. VLDB, pages 187–198, 2006.

[7] Dipti Patil, Bhagyashree Agrawal,Snehal Andhalkar, Richa Biyani, Mayuri Gund, Dr. V.M.Wadhai." An Adaptive parameter free data mining approach for healthcare application". In (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 1, 2012.

[8] Zetao Lin, Yaozheng Ge*, Guoliang Tao," Algorithm for Clustering Analysis of ECG Data",in Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference ,Shanghai, China, September 1-4, 2005,PP-0-7803-8740-6/05

[9] Douglas A. Coast, Richard M. Sten, Stanley A. Briller, "An Approach to Cardiac Arrhythmia Analysis Using Hidden Markov Models"in IEEE Transactions On Biomedical Engineering. Vol. 37. No. 9. September 1990, Pp- 001 8-9294/90/0900-0826

[10] V.Ilankumaran ,Dr.S.ThamaraiSelvi," Ventricular Arrhythmias Detection using Wavelet Decomposition",in International Journal of Computer Applications (0975 – 8887)Volume 20– No.1, April 2011.

[11] Miad Faezipour, Adnan Saeed, Suma Chandrika Bulusu,Mehrdad Nourani,, Hlaing Minn," A Patient-Adaptive Profiling Scheme for ECG Beat Classification"in IEEE Transactions On Information Technology In Biomedicine, Vol. 14, No. 5, September 2010,PP- 1089-7771.

[12] Mohd Afzan Othman, Norlaili Mat Safri, Rubita Sudirman "Characterization of Ventricular Arrhythmias in Electrocardiogram Signal using Semantic Mining Algorithm"in 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation,PP- 978-0-7695-4062-7/10

[13] Jimena Rodríguez, Alfredo Goñi, and Arantza Illarramendi," Real-Time Classification of ECGs on a PDA" in IEEE Transactions On Information Technology In Biomedicine, Vol. 9, No. 1, March 2005, Pp-1089-7771

[14] J. Hill, R. Szewczyk, A, Woo, S. Hollar, D. Culler, and K. Pister, System Architecture Directions for Networked Sensors, ASPLOS, November 2000.

[15] Fortis Escorts Hospital, Jaipur 302017 India, "Burden of Cardiovascular Diseases in India".

[16] Claude Seidman. "Data Mining with Microsoft SQL Server 2000 Technical Reference", ISBN: 0-7356-1271-4,amazon.com/Mining-Microsoft-Server-Technical-Reference/dp/0735612714.

[17] Amara Graps. 1995. An Introduction to Wavelets. IEEE Comp.Sc. and Eng. 2(2).

[18] Dipti D. Patil, V.M.Wadhai, "Cardiac Health status implementation on Mobile", accepted for publication in IJCSI and in publishing process.

Dr. Vijay M.Wadhai received his B.E. from Nagpur University in 1986, M.E. from Gulbarga University in 1995 and Ph.D. degree from Amravati University in 2007. He has experience of 26 years which includes both academic (19 years) and research (7 years). He has been working as the Principal of MAEER's MITCOE, Pune since 2010. He had worked as Dean of Research, MITSOT, MAE, Pune (from 2009 to 2010) and handled the post of Director - Research and Development, Intelligent Radio Frequency (IRF) Group, Pune. He is currently guiding various numbers of students for their PhD work in both Computers and Electronics & Telecommunication area. His research interest includes Deductive Databases, Knowledge Discovery and Data Mining, Data Security, Natural Language processing, Cognitive Radio and Wireless Communication, Spectrum Management, Wireless Sensor Network, ASIC Design - VLSI, Advance Network Design. He is a member of LMISTE, MIETE, MIEEE, MIES and GISFI (Member Convergence Group), India. He has filed numbers of patents in his research areas and also published quality International journal papers.

Dipti D. Patil has received B.E. degree in Computer Engineering from Mumbai University in 2002 and M.E. degree in Computer Engineering from Mumbai University, India in 2008. She is pursuing her PhD from Sant Gadgebaba Amravati University in P.G. Department of Computer Science and Engineering, SGBAU, Amravati. She is currently working as Assistant Professor in MITCOE, Pune. She has filed patents in data mining area and published numbers of international papers. Her Research interests include machine learning, Medical Signal Processing and Wireless Body Area Networks.