Adapting ECG Data Stream Mining for Health Care Application

Nilima Motghare, Prof. Arvind Mewada

Abstract—Heart disease is the most leading cause of death. Electrocardiogram ECG data stream signal plays a vital role in the diagnosis of heart related problems. ECG is very sensitive in nature and even if small amount of noise interferes with it, the characteristics of the signal change. The ECG must be clearly represented and filtered to remove all noise and artifacts from the signal. The main objective of the processing of ECG signal is to provide us the accurate, fast and reliable information of clinically important parameters like duration of QRS complex, the R-R interval, occurrence, amplitude and duration of P, R and T waves. Recently, numerous research and techniques have been developed for analyzing the ECG signal. Schemes were mostly based on Fuzzy Logic Methods, Artificial Neural Networks (ANN), Genetic Algorithm (GA), Support Vector Machines (SVM), and other Signal Analysis techniques. The proposed paper uses FIR filter and Wavelet Decomposition to preprocess and filter the ECG signal to overcome from concept drift and fuzzy based ensemble classification for analyzing the ECG signal.

Index Terms— ECG, data stream mining, preprocessing.

1 INTRODUCTION

An atherosclerosis process in heart disease is a condition that develops when a substance called plaque builds up in the walls of the arteries. This process making it harder for blood to flow through the brain and other organs by buildup narrows and hardening of the arteries, The gradual clogging, hardening and damage done to the interior walls of our blood vessels are the primary cause of heart attacks and strokes.

Symptoms of cardiovascular disorder as follows:

- Chest pain
- Edema
- Limb pain
- Orthostatic
- Hypotension Palpitations
- Syncope

A data stream electrocardiogram ECG is used to measure the heart's electrical conduction system. It generates waveform by picking up electrical impulses generated by the polarization and depolarization of cardiac tissue. The waveform can give information regarding the rhythm of the heart whether that impulse is conducted normally throughout the heart, the size and position of the chambers, whether any part of the heart is contributing more or less than expected to the electrical activity of the heart as well as the presence of any damage to the heart and the effects of drugs or devices used to regulate the heart. ECG is very sensitive in nature and even if small amount of noise interferes with it, the characteristics of the signal change. FIR filter and Wavelet transform is used to preprocess and filter the ECG signal from noise and fuzzy based ensemble classification for analyzing the ECG signal.

The paper is organized as, ECG morphology approach discusses in section 2. Section 3 literature survey discusses with preliminaries and related work discussion in Section 4 and section 5 conclusions.

2 ECG MORPHOLOGY

2.1 ECG LEADS PLACING ON HUMAN BODY

ECG 'leads' are different viewpoints of heart electrical activity and display the instantaneous difference in potential between various electrodes.
The commonly used 12 lead ECG consist

- Three bipolar limb leads.
- Three augmented unipolar limb leads.
- Six unipolar precordial leads.

a) Bipolar Limb Leads:

The three standard bipolar limb lead selected by Einthoven detects variation in the electrical potential between two points in the frontal plane of body. Electrode are applied to the right arm (RA), left arm (LA) and left leg (LL) anywhere on the extremity, through usually they are applied just above the wrist and the ankles.

- Lead I is produced by potential difference between RA^- and LA^+
- Lead II is produced by potential difference between RA^- and LL^+

- Lead III is produced by potential difference between LA^- and LL^+

Relation Between the Bipolar Limb Leads: It the polarity of lead II was LL to RA instead of RA to LL, the three bipolar leads would have formed a closed circuit and lead I+II+III = 0. However, the polarity of lead II was reversed by Einthoven, probably to get an upright deflection in all the three leads.

Thus according to kirchoff's law

I+II+III = 0

b) Unipolar Augmented Limb Leads

In 'unipolar lead' cardiology two electrode are employed. By an electric arrangement potentials recorded by one electrode (indifferent electrode) are rendered negligible so that only ttivhe electrical activity of other electrode (Exploring eletrode) is recorded. The indifferent electrode terminal is RA+LA+LL and equal zero.

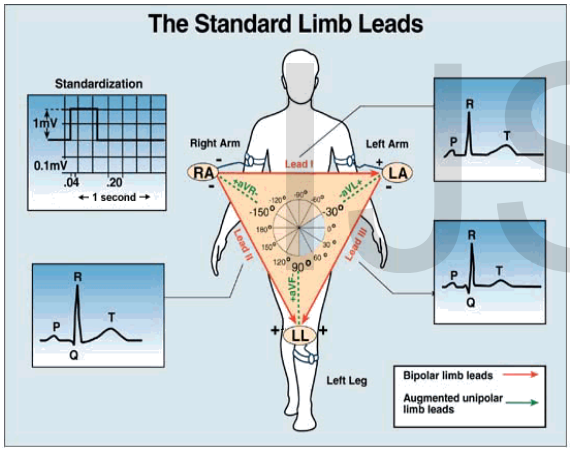

Fig. 1. Unipolar Precordial Leads (Chest Leads).

c) Six Uprecordial Leads.

The six precorcordial unipolar leads detect the electrical potential at specific points on the chest wall in the horizontal plane of the body. The common precordial positions used are as follows:

V1: Forth intercostal space at the right sternal border.
V2: Forth intercostal space at the left sternal border
V3: Equidistant between V2 and V4
V4: Fifth left intercostal space in the mid-clavicular line.
V5: Fifth left intercostal space in theanterior line.
V6: Fifth left intercostal space in the mid-auillary line.

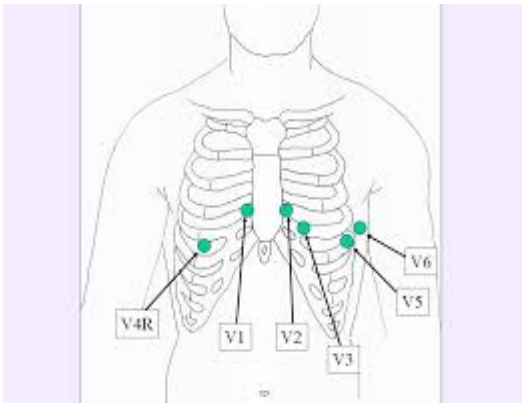

Fig. 1. Position of the Unipolar Chest Leads.

2.2 Electrocardiogram (ECG) Waveform

Each beat of the heart can be observed as a series of deflections away from the baseline on the ECG. These deflections reflect the time evolution of electrical activity in the heart which initiates muscle contraction. A single sinus (normal) cycle of the ECG, corresponding to one heartbeat, is traditionally labeled with the letters P, Q, R, S, and T on each of its turning points (Fig. 1). The ECG may be divided into the following sections.

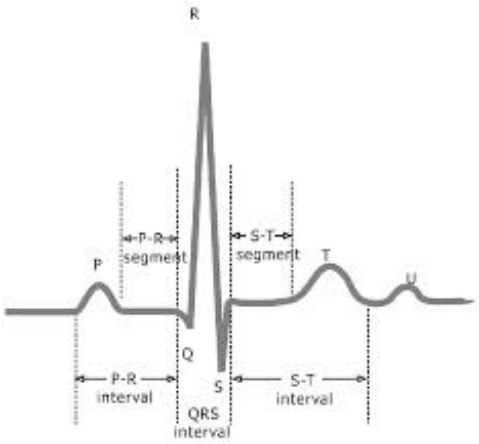

Fig. 6. Normalized ECG Signal Generated by Preprocessing

The meaning of various ECG parameter of standard ECG waveform as follows:

P wave = activation (depolarization) of atria.

PR interval = time interval between onset of atrial depolarization and onset of ventricular depolarization.

QRS complex = depolarization of ventricles, consisting of the Q, R, and S waves.

QT interval = time interval between onset of ventricular depolarization and end of ventricular repolarization.

R-R interval = time interval between QRS complexes.

T wave = ventricular repolarization.

ST segment plus T wave (ST-T) = ventricular repolarization.

U wave = probably after-depolarization (relaxation) of ventricles.

3 LITERATURE SURVEY

ECG is an important tool for the primary diagnosis of heart diseases. Abel-Rahman al-qawasmi et. al [5] proposed new approach using wavelet transform (WT) to remove all noise and artifacts from the signal from ECG signal and adapt the discrete wavelet transform (DWT) to enhance the ECG signal.

Rune fensli et. al [6] present concept for wireless and wearable electrocardiogram (ECG) sensor includes two electrical contact point applied directly to the patient's skin, transmitting signals to a diagnostic station at the hospital. The concept is intended for detecting rarely occurrences of cardiac arrhythmias and to follow up critical patient from their home while they are carrying out daily activities.

Dipati D. pati et. al [2] proposed introduced the methods for detecting concept drift in data stream. Wireless sensor network based Mobile Real-time Health care Monitoring (WMRHM) framework which has the capacity of giving health predictions online based on continuously monitored real time vital body signals.

Dipti D. Patil et. al [7] proposed data stream mining algorithm with system that integrated with wireless sensor network (WSN) to extract feature of ECG signal and clustering these feature to predict the type of arrhythmia consist of following steps:

1) Preprocessing ECG signal.
2) Feature extraction of ECG signal.
3) Data mining algorithm to learn from the collected data of ECG and form the rule which help to predict the type of arrhythmia.

Anpeng huang et. al [1] proposed WE-CARE system is useful and efficient mHealth (mobile health) tool includes QRS complex detection algorithm and T wave detection algorithm uses mobile 7-lead ECG devices for cardiovascular disease diagnosis and treatment in medical platform.

Vijanth S. Asirvadam et. al [4] proposed algorithm of translating 1D ECG in order for it to display in the form of 2D-Grayscale by using moving-average windowing technique using various moving setting of window sizes for better diagnostics of patient who are suffering from heart diseases.

Patrick E.McSharry et. al [8] proposed a new dynamical model which is capable of replicating many of the important features of the human ECG. Model parameters may be chosen to generate different morphologies for the PQRST-complex. The power spectrum of the RR-intervals can be selected *a priori* and used to drive the ECG generator. Both respiratory sinus arrhythmia at the high frequencies (HFs) and Mayer waves at the low frequencies (LFs) together with LF/HF ration are incorporated in the model

Dipati Patil et al. [3] proposed two main clustering algorithms k-mean algorithm D-mean algorithm. These Algorithm helps to determine fitness of person (fit or unfit) based on his/her historical and real time data, also In model building we perform feature extraction and cluster formation. This paper shows the experimental result of k-stream and D-stream accordingly we can say that D-stream gives more accurate prediction than k-mean when used for cluster formation of historical biomedical data.

4 PRELIMINARIES AND RELATED WORK

4.1 Pre-Processing of ECG Signal

The Pre-processing of an ECG signal is performed for the removal of noise associated with the ECG signal. While acquisition of ECG, it gets corrupted due to different types of artifacts and interferences such as Power line interference, Motion artifacts, Electrode contact noise, Muscle contraction, Base line drift, Electrosurgical noise, Instrumentation noise generated by electronic devices. Steps are carried out to filter out all these noise sources for geting meaningful and accurate detection.

A function of Mat-Lab smooth ( ) is used for noise removal.

Z = SMOOTH (Y, SPAN)

This function smooth data Y using SPAN as the number of points used to compute each element of Z. *smooth()* uses the Moving Average filter and FIR (Finite Impulse Response) Filter for smoothing the ECG signal.

4.2 FIR (Finite Impulse Response) Filter

FIR filters are widely used due to the powerful design algorithms that exist for them, their stability when implement in non-recursive form, ease with which one can attain linear phase, the ample hardware support and their simple extensibility to multi rate cases that exists for them among other reasons. It is necessary to specify the amount of passband ripple and stopband attenuation in order to determine a suitable filter order that will be tolerated. The width of the transition region around the ideal cut-off frequency is also necessary. The latter is done by setting the edge frequency of passband and stopband filter. The difference between the two determines the transition width.

4.3 ECG Wavelet Decomposition

The wavelet transform provides a decomposition of the signal over a set of basis functions, obtained by dilation and translation of a mother wavelet by a scale factor s. The Wavelet Transform provides a time- frequency representation of the signal and is well suited to the analysis of non-stationary signals such as ECG.

In this proposed paper, the ECG wavelet decomposition is performed by using Single-level discrete 1-D wavelet transform function. It performs a single-level 1-D wavelet decomposition up to specified scale factor s=8. The mother wavelet or basis function that is used in the decomposition is Dabachies6 (db6) wavelet procedure.

4.4 Fuzzy Based Ensemble Classification

a) Fuzzifier

Fuzzifier converts the crisp input to a linguistic variable using the membership functions stored in the fuzzy knowledge base.

- **Crisp input** consist of the set of linear objects, for example the crisp set of month may include the name of month only such as January, February, March, Monday, etc.
- **A linguistic variable** is a variable that, instead of numerical values, consists of linguistic terms. Consider the linguistic variable *speed* which may consist of the terms *slow, medium* and *fast.* These linguistic terms can each be described using a fuzzy set. As an example we might consider the speed *slow* to be below 40 mph (meter per hour), *medium* to be around 55 mph and fast to be above 70 mph.
- **Membership function** is a curve that defines the mapping between input values with the membership value/degree between 0 to 1.

b) Inference Engine

Inference Engine using "If-Then" type fuzzy rules converts the fuzzy input to the fuzzy output. The fuzzy output produced after applying fuzzy rules by an Inference Engine is called as fuzzy sets.

c) Defuzzifier

The job of defuzzifier is to converts the fuzzy output of the inference engine into crisp using membership functions analogous to the ones used by the fuzzifier.

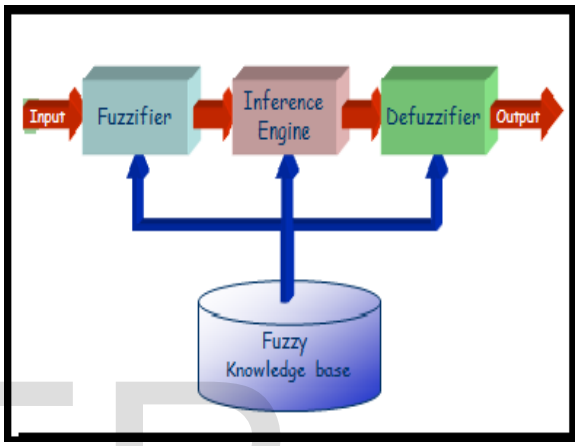

Fig.4. Architecture of Fuzzy Interence System

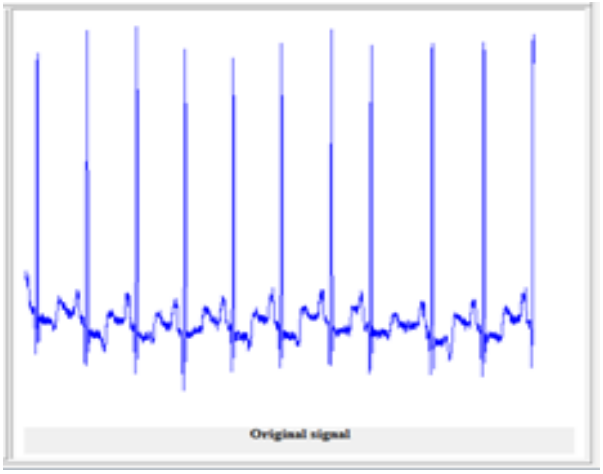

Fig. 5. Original ECG Signal.

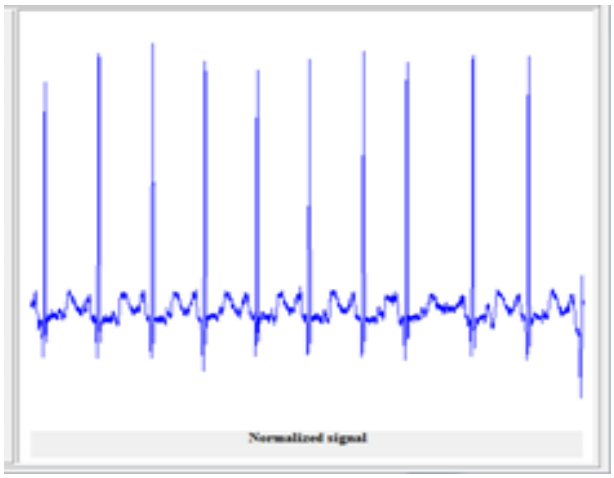

Fig. 6. Normalized ECG Signal Generated by Preprocessing

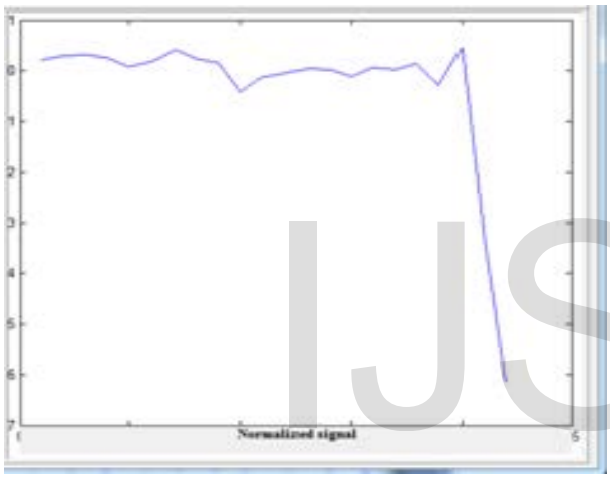

Fig. 1.Wavelet Decomposition.

5 CONCLUSION

We are focusing on with removal of noise and artifacts from data stream electrocardiogram (ECG) signal an accurate feature extraction like duration of QRS complex, the R-R interval, occurrence, amplitude and duration of P, R and T waves is most important for heart disease diagnosis. Fuzzy rule based has high capability of knowledge extraction and representation when modeling is complex, non linear classification problem, so the FIR filter and Wavelet decomposition is used to preprocess and filter the ECG signal to overcome from concept drift and fuzzy based ensemble classification for analyzing the ECG signal.

[1] Anpeng huang, chao chen, kaigui bian, xiaohui duan, Min chen, Hongqiao gao, chao meng, qian zheng, yingrui zhang, bingli jiao and linzhen xie, " we-care: an intelligent mobile telecardiology system to enable mHeath application"

[2] Dipati Durgesh pati, Vijay Wadhai, "Adaptive real time data mining methodology for wireless body area network based healthcare application" , Advanced Computing: An International Journal (ACIJ), Vol. 3, No.4 July 2012.

[3] Dipati Patil, Shehal Andhalkar, Mayur Gund, Bhagyashree Agrawal, Richa Biyani, V. M. Wadhai, "An adaptive parameter free mining approach for healthcare application", (IJACSA) International Journal of Advanced Computer Science and Application, Vol. 3, No. 1, 2012.

[4] Vijanth S. Asirvadam, Kauv Seth Pisal, Lila Iznita Izhar and Noor Anis A. M. Khuzi, "ECG Viewed using Grayscale Patterns", Proceedings of International Conference on Man-Machine System (ICoMMS) 11-13 October 2009.

[5] Abel-Rahman al-qawasmi, khaled daqrouq, "ECG signal enhancement using wavelet transform", WSEAS transactions on biology and biomedicine, issue 2, volume 7, april 2010.

[6] Rune fensli, Einar Gunnarson, Torstein Gundersen, "A wearable ECG-recording system for continuous arrhythmia monitoring in a wireless tele-home-care situation",18 IEEE international Symposium on computer-based medical systems, Dubline, June 23-24, 2005.

[7] Dipti D. Patil, Dhanashri Patil, Sharlin Pandharpatte, Ruta Dhekane, Trupti Mohal and Dr. V.M. Wadhai, "Intelligent Arrhythmia Diagnostics System", IJCSI International Journal of Computer Science Issues Vol.9, Issue 6, No 1, November 2012.

[8] Patrick E.McSharry, Gari D. Clifford, Lionel Tarassenko, and Leonard A. Smith, "A dynamical model for generating synthetic electrocardiogram signals ", IEEE transactions on biomedical engineering, vol. 50, No 3, March 2003.

[9] V.I ponomarenko, M. D. Prokhorov, A. B. Bespyatov, M.B. Bodrov, V.I. Gridnev, "Deriving main rhythms of the human cardiovascular system from the heartbeat time series and detecting their synchronization", 2004 Elsevier Ltd. All rights reserved.

[10] Dipati D. patil, Jyoti G. Mudkanna, Dnyaneshwar Rokade, Vijay M. Wadhai, "Concept Adapting Real-Time Data Stream Mining for Heath Care Application", advances in intelligent and soft computing volume 166, pp341-351, Springer 2012.