International Journal of Engineering, Science and Mathematics
Vol. 6 Issue 3, July 2017,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com

HYBRID APPROACH OF DATA MINING TECHNIQUES TO PREDICT HEART DISEASE

Anmoldeep Kaur[a]*, Yogesh Kumar[b],

a Computer Engineering, Gurukul Vidyapeeth Group Of Institutions, Banur
b Computer Engg, Gurukul Vidyapeeth Group Of Institutions, Banur

ABSTRACT

Most nations face high and expanding rates of heart diseases or Cardiovascular Disease. Despite the fact that, advanced pharmaceutical is creating colossal measure of information consistently, little has been done to utilize this accessible information to comprehend the difficulties that face an effective elucidation of echocardiography examination results. The present paper based on the study of models utilizing the different heart diseases forecast calculations like: Decision Tree, Greedy Algorithm and Neural Network.
In this paper different authors paper reviewed for prophetic model for heart diseases prediction using data mining methods for improving the dependability of heart diseases.

Keywords: Data Mining, Heart Disease Prediction, Decision Tree, Classification, And Neural Network.

1. Introduction

Electrocardiogram

ECG (Electrocardiogram) is an interpretation tool that tells the electrical activity of heart recorded electrode by skin. The morphology and heart rate emulate the cardiac health of human heartbeat [1]. It is a not invasive technique that is used for the signal is measured on the surface of a human body, which is used in the identification of the heart diseases [2]. The amplitude and duration of the P-QRS-T wave contain useful information about the nature of disease afflicting the heart. The electrical wave is due to depolarization and repolarization of Na+ and k-ions in the blood [2].The ECG signal provides the following information about a human heart [3]

- Heart position and its relative chamber size impulse origin and propagation
- Heart rhythm and conduction disturbances extent and location of myocardial ischemia changes in electrolyte concentrations drug effects on the heart.

Data mining

Popular use the code to greater trade production, much-computerized business and government activity, leading to the collection of data tools has furnished too with the enormous quantity of data. Recently, our efficiency of the pair developing as well as the cluster has enhanced expeditiously. Trillion databases have been worn into thead ministration of government, experimental, management of business and management the engineering data, rife another application.

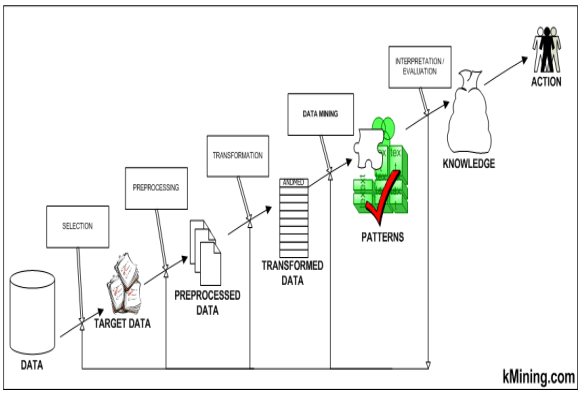

Figure 1.1 data mining

Classification is design or progress at discovery the sample in which find the unknown objects in the class. This is the common map for data objects in one of the predefined classes. It is Crate generates the rules for feature data in training dataset. Classification is the future unknown data items is used more these rules. It is important ways to doing things. Medical identification of disease or problem or its cause is an important application of classification.

Importance of mining the data

In the extensive reason is attract a great deal of through in the instruction business in the latter year is as the vast opportunity of large amounts data. Data instruction and knowledge is imminent needs for turning. It is quickly adapt for more number of domains. Great areas use for a deal with a large number of data such as science, business, and other environments like education and medical field, administration of government, experimental experimental, administration of government, administration of government, experimental. The gain the large instruction from the various applications like education, medical, market study, science exploration etc.

2. Literature Review

Yahyaoui et al. [2016] has developed the feature-based trust sequence classification algorithm. The authors have proposed a novel feature-based approach to assessing the trusting behaviours of a service. By analysing the possible trust behaviours of services, trust patterns are outlined to describe trust sequences supported 3 criteria: its overall behaviour, the starting behaviours and ending behaviour [7].

Bo Liu and Qiang Chen et al. [2015] Analysing statistic information will reveal the temporal behaviour of the underlying mechanism manufacturing the info. Statistic motifs, that area unit similar subsequence's or often times occurring patterns, have important meanings for researchers, particularly in the medical domain. This work proposes an economical Motif discovery methodology for Large-scale statistics (MDLats). By computing the motifs, MDLats eliminates a majority of redundant computation within the connected arts and reuses existing data to the most. All the motif sorts and subsequence's area unit generated for subsequent analysis [1].

Harvey et al. [2015] has worked on automated feature style for numeric classification by genetic programming. In this method, a genetic programming variant evolves a population of candidate features designed from a library of sequence-handling functions. Numerical optimization strategies, In this, hybrid approach ensure that the fitness of candidate algorithms is measured victimization optimum parameter values. Auto feed represents the first automatic feature style system for numeric sequences to leverage the ability and potency of each numerical optimization and commonplace pattern recognition algorithms [2].

Aleksandra et al. [2015] has worked on the uncertainties in the measurement of nonlinear dynamics in heart rate variability. Different strategies for the mathematical analysis of heart-rate variability (HRV) ar thought-about. Analysis and comparison of the noise stability of HRV indices, determined on the basis of nonlinear dynamics, geometrical and statistical strategies were carried out [3].

Dustin et al. [2014] have performed the robust analysis of time series classification algorithms for structural health observation. The supervised learning methodology for data-driven SHM involves computation of low-dimensional, damage-sensitive features from raw measuring information that are then used in conjunction with machine learning algorithms to discover, classify, and quantify damage states.Probabilistic approaches to robust SHM system style suffer from the incomplete information of all conditions a system can expertise over its period of time. Info-gap decision theory allows non-probabilistic analysis of the strength of competitor models and systems in a sort of deciding applications. Previous work employed info-gap models to handle feature uncertainty once choosing numerous elements of a supervised learning system, namely options from a pre-selected family and classifiers. In this work, the info-gap framework is extended to robust feature style And classifier choice for general time series classification through an economical, interval arithmetic implementation of an info-gap information model [4].

Wang et al. [2013] has worked towards the bag-of-words representation for medical speciality time series classification. In this work, a simple however effective bag-of-words illustration that's originally developed for text document analysis is extended for medical specialty statistic illustration. The proposed technique treats a statistic as a text document and extracts native segments from the time series as

words. The biomedical time series is then described as a bar chart of code words, each entry of that is the count of a codeword appeared within the statistic. This is able to capture high-level structural data as a result of each native and world structural information are well used [5].

Maneesha V. Ramesh et al. [2012] proposed a system that is low cost, low power wearable wireless sensor helps in early detection and prevention of cardiac arrest.System works on models which are based upon the state of the patient.Discrete wavelet transformation (DWT) is applied to compress the data for power optimization and communicate it Bluetooth medium is used so as to deliver it to the doctor on move or to the caretakers. The advantage of proposed system is that, for compression, DWT is used that will take the discrete samples that in turn reduces the data to be sent but a disadvantage is a medium chosen for sending data as it is not reliable and more power consuming with the shorter range of communication.

Youssouf Zatout [2012] did a comparative study of the various wireless technologies that can be used for health monitoring in the heterogeneous wireless sensor network (WSN). He has presented an analysis on how to choose the best wireless technology for your application in WSN based upon factors like delay, Quality of Service (QoS), topology, energy consumption and other factors.He laid down the pros and cons of each wireless technology that can be embedded in WSN for data communication. Technologies which he studied like Bluetooth, Zigbee, UWB, Z-wave, Wi-Fi and others, Zigbee shows remarkable results and suits the WSN technology for data transmission.

S. Muthulakshmi et al. [2012] have proposed a method for the ECG classification based upon feature extraction.Proposed method reduces the computation complexity in selecting the best feature. Computational search turns to be complex with the increase in a number of features to be extracted. Particle Swarm optimization (PSO) method is preferred for the feature selection as it works with best feature combination detection method in the search space.Multi-layer Perceptron (MLP)is used for the classification of the selected features, another method Support Vector Machine (SVM) is used for the classification. The advantage is proposed method reduces the complexity in feature selection but the disadvantage is classification method induces complexity.

Mohammed Abo-Zahid et al. [2012] have proposed an efficient Electrocardiogram (ECG) signals compression technique that is based on QRS-complex detection and estimation,2-D ECG conversion and Wavelet Transformation (WT). Original ECG signal is pre-processed for QRS-complex detection and then the difference between the original signal and QRS-complex is estimated. Estimated difference is represented in a 2-D array as there are some repeated beats and samples. 1-D data turned into 2-D data with reference to R-signal and of Length L. Residual or the difference is assembled in a row and then segmented into 32 by 32 matrices. The wavelet transform is applied and resulting coefficients are segmented into groups and threshold. The threshold value is declared specific for the specific group based on the entropy of coefficient and resulted threshold DWT coefficients are coded using coding scheme.The advantage is proposed method achieves high compression ratio with relatively low distortion and low computational complexity.Disadvantage works only for 2-D data matrix.

AgusDwi Suarjaya [2012] has proposed a new compression method for the lossless data. Proposed method is j-bit encoding (JBE), this will manipulate each bit of data present in the file that contains lots of data without changing its actual meaning. Proposed algorithm is clubbed with other losses algorithm to get better results. The advantage of proposed algorithm is it saves space for the storage of the intended data but a disadvantage is cannot work alone and time-consuming as works on bit level.

3. Conclusions

Following points can be concluded from the literature review.

1. The system extracts hidden knowledge from a historical heart disease database.

2. The most effective model to predict patients with heart disease appears to be Neural Network followed by Decision Trees and Greedy Algorithm.

3. The relationship between attributes produced by Neural Network is more difficult to understand.

4. Research Scope

It can incorporate other medical attributes. It can also incorporate other data mining techniques, e.g., Time Series, Clustering and Association Rules.

REFERENCES

[1] Bo Liu and Qiang Chen. "Efficient Motif Discovery for Large-Scale Time Series in healthcare." *IEEE Transactions.Indusrtial. informatics*, 11(3):583-590, (2015).

[2] Dustin Y. Harvey and Michael D. Todd. "Automated feature design for numeric sequence classification by genetic programming." *Evolutionary Computation, IEEE Transactions,* 19(4): 474-489, (2015).

[3] Fedotov, Aleksandra A., Anna S. Akulova, and Sergey A. Akulov. "Uncertainties in measurement of nonlinear dynamics in heart rate variability." In *6th European Conference of the International Federation for Medical and Biological Engineering*, pp. 102-105. Springer International Publishing, (2015).

[4] Harvey, Dustin Y., Keith Worden, and Michael D. Todd. "Robust evaluation of time series classification algorithms for structural health monitoring."In*SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*, pp. 90640K-90640K.International Society for Optics and Photonics, (2014).

[5] Wang, Jin, Ping Liu, Mary FH She, Saeid Nahavandi, and Abbas Kouzani. "Bag-of-words representation for biomedical time series classification."*Biomedical Signal Processing and Control* 8(6):634-644, (2013).