ANALYSIS AND CLASSIFICATION OF RISK ASSESSMENT IN PATIENTS SUFFERING FROM CONGESTIVE HEART FAILURE BY USING COMPUTER BASED STATISTICS

[1]M.Ramakrishnan , [2]C.Nalini
1PG Scholar, Department of Computer Science and Engineering, Bharath University
2Professor, Department of Computer Science and Engineering, Bharath University
[1]krishdeva_84@yahoo.co.in, [2]drnalinichadambaram@gmail.com

Abstract: A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at reasonable costs. Quality service suggests diagnosing patients correctly and administering treatments that are effective. Poor medical results can lead to disastrous consequences which are unacceptable. So Hospitals should reduce the cost of medical examinations. They can accomplish these results by employing proper computer-based statistics and/or decision support systems. Most of the hospitals nowadays service certain kind of hospital information to manage their healthcare or patient information. These systems are intended to maintain the patient billing, portfolio management and generation of simple data's. Particular clinics using decision support systems, but they are largely limited. Medical decisions are often made based on doctor's intuition and experience rather than on the knowledge rich data hidden in the record. This practice leads to unwanted errors and too much of medical allowances which affects the quality of service providing to patients. The main objective of this research is to develop a Disease Prediction Scheme using data mining modeling technique, namely, Naïve Bayes. It is implemented as web based questionnaire application .Based on the user answers, it can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease databank. It can answer difficult queries for analyzing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot. By providing perfect treatments, it also supports to reduce medical treatment costs effectively.

Keywords: CHF-Congestive Heart Failure, NYHA- New York Heart Association, ECG- Electro Cardio Graphic, ANS-Automatic Nervous System, HRV-Heart Rate Variability.

1. INTRODUCTION

Congestive heart failure (CHF) is a pathophysiological condition due to an abnormal cardiac function, which is responsible for the failure of the heart to pump blood as required by the body. CHF severity can be measured with the symptomatic classification scale of the New York Heart Association (NYHA). Classification via NYHA scale has been proved to be a risk factor for mortality. Heart rate variability (HRV) is the variation over time of the period between consecutive heartbeats (RR intervals) and is usually extracted from electro cardio graphic signal (ECG) recorded through a non-invasive technique.

HRV is commonly used to assess the influence of the autonomic nervous system (ANS) on the heart. HRV has been widely studied in patients suffering from CHF. Many studies demonstrated that HRV is an effective means for the risk assessment of mortality. A number of studies demonstrated the relationship of HRV measures and the NYHA classification scale. In our previous papers, we demonstrated that HRV might be used to detect CHF using short-term orlong-term measures.

Moreover, we proposed a classifier based on short-Term HRV measures to individuate strictness of CHF. Over the previous years, automatic classifiers, based on several clinical and instrumental parameters, have been planned to help CHF assessment. However, to the best of the author's knowledge, these classifiers are not based on HRV features, except for those proposed by Yang et al. who included HRV features but did not provide details about the related processing. In this study, we present a classifier, based on long-term HRV measures, for the individuation of high-risk conditions in CHF patients, estimated via NYHA scale.

Patients where considered at higher risk if suffering from severe CHF (NYHA III or IV) and at lower risk if suffering from mild CHF (NYHA I or II).The method we used to develop the classifier is classification and regression tree (CART). CART, developed by Breiman, has been used in several applications of pattern recognition especially for medical diagnosis. The CART algorithm iteratively separates the dataset, according to the criteria that exploits to splitting the information's, and constructing a tree-like decision structure. CART was applied to HRV measures for other investigations. We adopted CART in previous studies in which a larger dataset was available to train the CART and/or the final classification of the patients was based on a combination of trees. On the contrary, in this study, the selected dataset is small and unbalanced. A number of solutions to the class imbalance problem were previously proposed at data, feature selection, and algorithmic levels. At the data level, these solutions include many different forms of resampling.

At the algorithmic level, solutions include adjusting the costs of the various classes so as to counter the class inequality, correcting the probabilistic estimation at the tree leaf (when working with decision trees) and adjusting the decision threshold. As regards feature selection, Zheng et al. proposed a framework to deal with imbalanced dataset, showing the importance of feature selection methodology and performance measurement. In this study, we adopted CART algorithms with a feature selection algorithm in order to handle a small and unbalanced dataset. We compared the performance of the proposed method with a standard data-level-based method to deal with imbalance that is the oversampling technique.

We preferred a data-level solution as the benchmark since the algorithmic level solutions require the probability and the misclassification cost of the class that are difficult to estimate particularly in this case, as the rare class is also the milder one. Moreover, we compared the results of the proposed method with other classifiers based on decision trees, i.e., C4.5 and random forest (RF).We implemented decision trees as they provide a arrangement model, rules, which can be easy to read and interpret. This is crucial in medical applications in which the physician is personally responsible of the diagnosis. The HRV measures were extracted from two Holter monitor public databases by using only open source and validated HRV toolkit software in order to allow other scientists to reproduce our results.

2. EXISTING SYSTEM:

- Text mining based porter stemmer algorithm and clustering algorithm used.
- Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database.
- This practice points to unwanted biases, mistakes and excessive medical costs which affects the quality of service provided t

2.1 DISADVANTAGE

- There are many ways that a medical wrong diagnosis can present itself. Whether a doctor is at mistake, or clinical staff, faulty diagnosis of a serious disease can have very extreme and harmful effects.
- Less accuracy and overhead.
- Medical Misdiagnoses are a serious risk to our healthcare line of business. If they continue, then publics will fear going to the hospital for their treatment.
- We can give the conclusion to medical misdiagnosis by informing the public and filing claims and suits against the medical practitioners at fault.

3. PROPOSED SYSTEM

- This practice points to unwanted biases, mistakes and excessive medical costs which affects the quality of service provided to patients.
- Thus we suggested that the addition of medical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome.
- This suggestion is promising as data modeling and examination tools e.g., data mining, have the possibility to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.
- The main objective of this paper is to develop a prototype Intelligent Heart Disease Prediction

System (IHDPS) using three data mining modelling techniques, namely, Decision Trees, Naïve Bayes.

3.1 ADVANTAGE:

- Its providing perfect medical treatments, also it helps to decrease medical treatment costs. To improve visualization and ease of interpretation.
- Additional accuracy achieved comparing with other algorithms.

3.2 DATAFLOW DIAGRAMS

LEVEL 0:

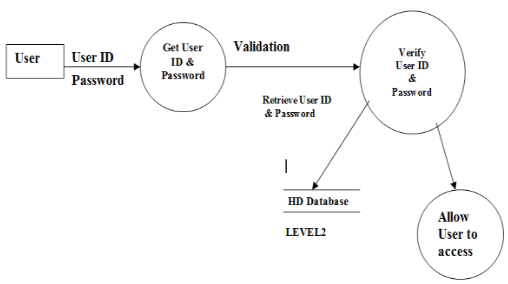

Figure: (1)

LEVEL 1:

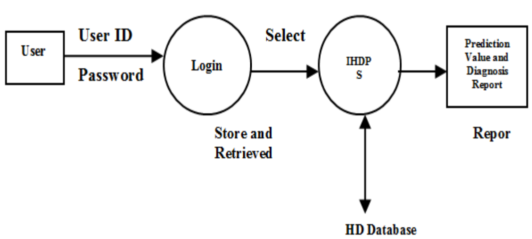

Figure :(2)

LEVEL 2:

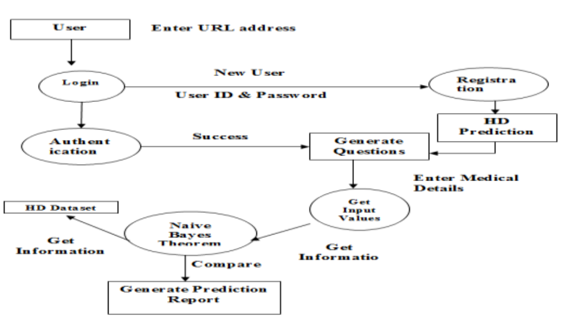

Figure: (3)

4. ARCHITECTURE

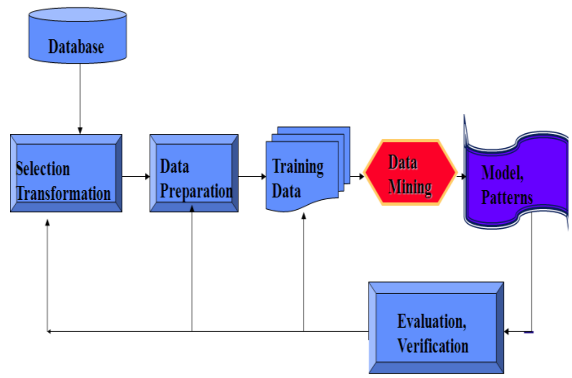

Figure :(4)

4.1 MODULE DESCRIPTION

Modules:

Data Set Generation

- Analyzing the Data set
- Bayes Classification
- Prediction

4.1.1 Data Set Generation.

Questionnaires have advantages over some other types of medical symptoms that they are very inexpensive, do not require as much effort from the interrogator as verbal or phone surveys, and often have standardized solutions that make it simple to compile data. However, such standardized solutions may irritate customers. Questionnaires are also sharply limited by the fact that respondents must be able to read the questions and respond to them.

Here our dataset is based on the following attribute, Input attributes

- Sex (value 1: Male; value 0 : Female)
- Chest Pain Type (value 1: typical type 1 angina value 2. usual type angina, value 3. non-angina pain; value 4. asymptomatic)
- Fasting Blood Sugar (value 1. > 120 mg/dl; value 0:< 120 mg/dl)
- Restecg – resting electrographic results (value 0: normal; value 1: 1 having ST-T wave abnormality; value 2: showing probable or definite left ventricular hypertrophy)

- Exang – exercise induced angina (value 1: yes; value 0: no)
- Slope – the slope of the peak exercise ST segment (value 1: unsloping; value 2: flat; value 3: downsloping)
- CA – number of major vessels colored by floursopy (value 0 – 3)
- Thal (value 3: normal; value 6: fixed defect; value 7:reversible defect)
- Trest Blood Pressure (mm Hg on admission to the hospital)
- Serum Cholesterol (mg/dl)
- Thalach – maximum heart rate achieved
- Oldpeak – ST depression induced by exercise relative to rest
- Age in Year
- Height in cms
- Weight in Kgs.

4.1.2Analyzing the Data Set

A data set is a collection of records, generally obtainable in tabular form. Each column denotes a specific variable. Each row matches to a given member of the data set in question. It lists principles for each of the variables, such as height and weight of thing or values of random information. Each value is called as a datum. The data set may comprise records for one or more members, equivalent to the number of rows.

The values may be numbers, such as real numbers or integers, for illustration demonstrating a person's height in centimeters, but may also be nominal data (i.e., not consisting of numerical values), for example representing a person's society. More values may be of any kinds of described as a level of scaling. For each variable, the values will usually all be of the same kind. However, there may also be "lost values", which want to be indicated in some technique. A total of 500 records with 15 medical characteristics (factors) were obtained from the Heart Disease database lists the characteristics. The records were split equally into two datasets: training dataset (455 registers) and testing dataset (454 registers). To avoid bias, the records for each set were selected randomly.

The attribute ―Diagnosis‖ was identified as the predictable characteristic with value ―1‖ for patients with heart disease and value ―0‖ for patients with no heart disease. The attribute ―Patient ID‖ was used as the key; the rest are input attributes. It is assumed that problems such as lost data, unpredictable data, and duplicate data have all been resolved. Here in our project we get a data set from .dat file as our file reader program will get the data from them for the input of Naïve Bayes based mining process.

4.1.3 Bayes Classification

Bayes' Theorem finds the probability of an event occurring given the probability of another event that has previously happened. If B represents the dependent happening and A represents the past event, Bayes' theorem can be indicated as follows.

Bayes' Theorem:

Prob (B given A) = Prob(A and B)/Prob(A)

To calculate the probability of B given A, the process calculates the number of cases where A and B occur together and divides it by the number of cases where A occurs alone. Applying Naïve Bayes to data with numerical attributes and using the Laplace correction (to be done at your own time, not in class) (data with some numerical attributes), predict the class of the following new example using Naïve Bayes classification: with some numerical attributes), predict the class of the following new example using Naïve Bayes classification.

4.1.4 Prediction

Medical data example, it may discovers that a set of indications or symptoms frequently occur together with another set of symptoms. The Bayes algorithm is an example of association techniques. Classification maps data items into one of several pre-defined classes. For example, classification rules about a disease can be extracted from previous known cases and then used to diagnose new patients of this disease based on their symptoms. Clustering recognizes the class or cluster for a set of unclassified objects according to their attributes. For example, the data set can be grouped into several clusters based on the similarities in their symptoms, and the common symptoms of the diseases in a cluster can be used to describe or predict that diseases

5. OUTPUT RESULTS

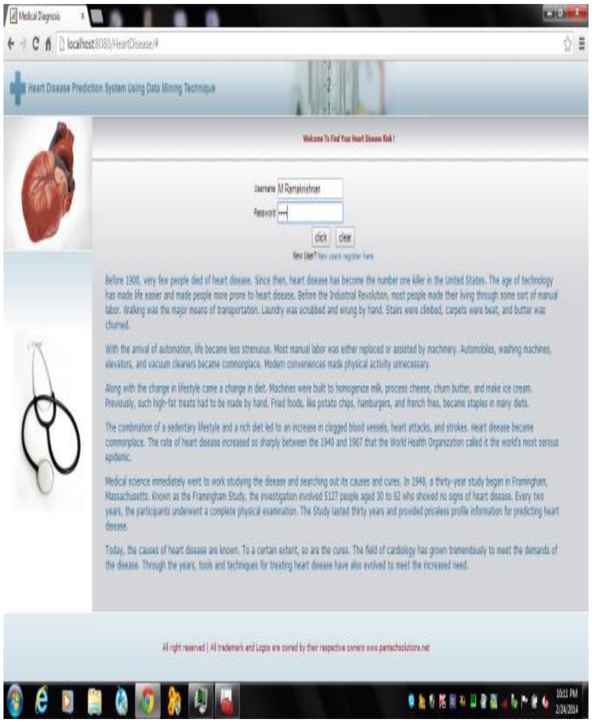

Figure: (5)

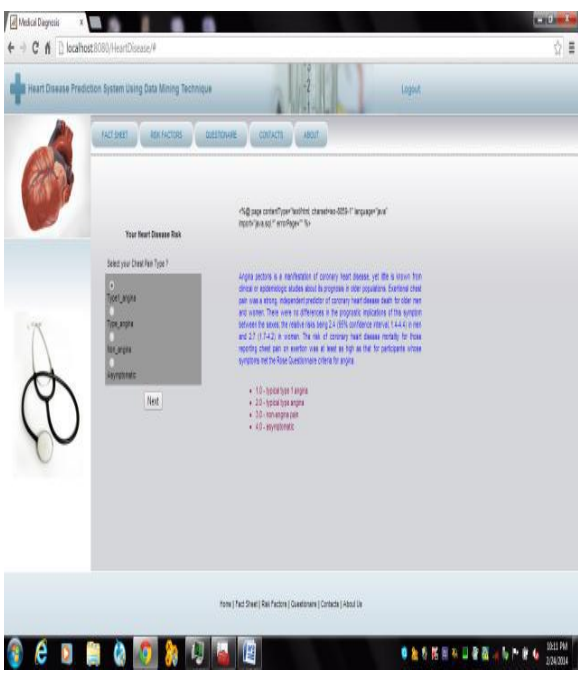

Figure: (6)

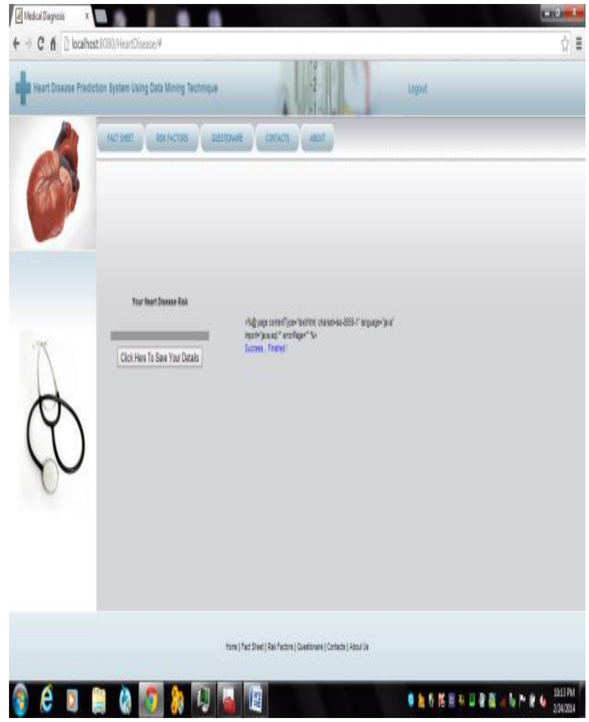

Figure: (7)

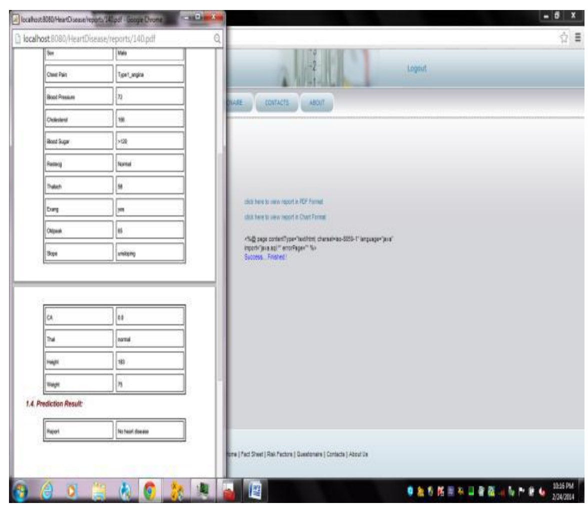

Figure: (8)

6. CONCLUSION

We conclude that the long-term HRV measures enable higher risk patients to be distinguished from lower risk ones. The classification trees developed achieved an accuracy rate of 85.4%, a sensitivity rate of 93.3% and a specificity rate of 63.6% (tenfold cross-validation estimates) using the combination of features TOTPWR, pNN10, pNN50, SDNNIDX and ULF, TOTPWR, PNN50. As a final point of our outputs are reliable with the consensus that depressed HRV values are associated with higher cardiovascular risk.

REFERENCE

[1] J. L. Fleg, I. L. Pina, G. J. Balady, B. R. Chairman, B. Fletcher, C. Lavie, M. C. Limacher, R. A. Stein, M. Williams, and T. Bazzarre, ―Assessment of functional capacity in clinical and research applications: An advisory from the committee on exercise, rehabilitation, and prevention, council on clinical cardiology, American heart association,‖ Circulation, vol. 102,no. 13, pp. 1591–1597, Sep. 26, 2000.

[2] M. M. Redfield, M. Senni, C. M. Tribouilloy, R. J. Rodeheffer, S. J. Jacobsen, J. M. Evans, and K. R. Bailey, ―Congestive heart failure in the community—A study of all incident cases in Olmsted County, Minnesota, in 1991,‖ Circulation, vol. 98, no. 21, pp. 2282–2289, Nov. 24, 1998

[3] M. Gheorghiade, L. Klein, C. M. O'Connor, J. D. Leimberger, W. Gattis- Stough, I. L. Pina, M. Felker, K. F. Adams, R. M. Califf, and O.-C. Investigators, ―Lower serum sodium is associated with increased short-term mortality in hospitalized patients with worsening heart failure—Results from the Outcomes of a Prospective Trial of Intravenous Milrinone for Exacerbations of ChronicHeart Failure (OPTIME-CHF) study,‖ Circulation, vol. 111, no. 19, pp. 2454–2460, May 17, 2005.

[4] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, ―Heart rate variability: a review,‖ Med. Biol. Eng. Comput., vol. 44, no. 12, pp. 1031–1051, Dec. 2006.

[5] M.Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A.Malliani, A. J. Moss, and P. J. Schwartz, ―Heart rate variability: Standards of measurement, physiological interpretation, and clinical use,‖ Eur. Heart J., vol. 17, no. 3, pp. 354–381, Mar. 1, 1996.

[6] G. Panina, U. N. Khot, E. Nunziata, R. J. Cody, and P. F. Binkley, ―Role of spectral measures of heart rate variability as markers of disease progression in patients with chronic congestive heart failure not treated with angiotensin-converting enzyme inhibitors,‖ Amer.Heart J., vol. 131, no. 1, pp. 153–157, Jan. 1996.

[7] J. E. Mietus, C. K. Peng, I. Henry, R. L. Goldsmith, and A. L. Goldberger, ―The pNNx files: Re-examining a widely used heart rate variability measure,‖ Heart, vol. 88, no. 4, pp. 378–380, Oct. 2002.

[8] T. D. J. Smilde, D. J. van Veldhuisen, and. P. van den Berg, ―Prognostic value of heart rate variability and ventricular arrhythmias during 13-year follow-up in patients with mild to moderate heart failure,‖ Clin. Res. Cardiol., vol. 98, no. 4, pp. 233–239, Apr. 2009.

[9] J. T. Bigger, J. L. Fleiss, R. C. Steinman et al., ―RR variability in healthy, middle-aged persons compared with patients with chronic coronary heartdisease or recent acute myocardial-infarction,‖ Circulation, vol. 91, no. 7, pp. 1936–1943, Apr. 1, 1995.

[10] S. Guzzetti, R. Magatelli, E. Borroni, and S. Mezzetti, ―Heart rate variability in chronic heart failure,‖ Auton. Neurosci. Basic Clin., vol. 90, no. 1–2, pp. 102–105, Jul. 20, 2001.