ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 6, June 2014

Computer Based Risk Analysis and Classification in Patients Suffering From Congestive Heart Failure

Reddy Santosh Kumar[1], B. Sreepathi[2]

M.Tech pursuing, Department of Computer Science and Engineering, RYMEC, Bellary, Karnataka, India [1]

Asst. professor & HOD, Department of Information Science and Engineering, RYMEC, Bellary, Karnataka, India [2]

Abstract: Medical diagnosis is considered an art regardless of all standardization efforts made, which is greatly due to the fact that medical diagnosis necessitates an expertise in coping with uncertainty simply not found in today's computing machinery. The researchers are encouraged by the advancement in computer technology to develop software to assist doctors in making decision without necessitating the direct consultation with the specialists. Diagnosis of Heart disease is a important and critical task which can provide prediction about the heart disease so that treatment made easy. The main objective of this research is to develop a Disease Prediction Scheme using data mining modelling technique, namely, Naïve Bayes. Here, data mining played a vital role in diagnosis of heart disease with improved value. So, analyzing those diagnosis techniques may lead to new improvement in area of research. This is implemented as web based questionnaire .Based on the 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 analysing 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, Data Mining, ECG –Electro Cardio Graphic, HRV –Heart Rate Variability**.**

I. INTRODUCTION

Due to the development of usage of unwanted things to eat each and every time, several major diseases occurs at a any time for human beings .One of the most important disease occurs at now a days are heart problem this is known as a heart failure (HF).It is also known as congestive heart failure (CHF), happen what time the heart is not capable to afford adequate propel accomplishment to preserve blood flow to congregate requirements of body. The term "heart failure" does not mean that the heart stops beating completely, but that the heart is not working as efficiently. 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 or long-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 were 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.

II. EXISTING SYSTEM

In our previous study, we demonstrated that Text mining based porter stemmer algorithm and clustering algorithm might be used to diagnose risk in patients. Moreover, Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. This might points to unwanted biases, mistakes and excessive medical costs which affects the quality of service provided to patients.

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.

III. PROPOSED SYSTEM

This paper aims to develop a Disease Prediction Scheme using data mining modelling technique, namely, Naïve Bayes. It is implemented as web based questionnaire. It can discover and extract hidden knowledge 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.

It also supports to reduce medical treatment costs effectively. This practice points to ward off unwanted biases, mistakes and excessive medical costs which affects the quality of service provided to patients.

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 data mining modelling techniques, namely, Decision Trees, Naïve Bayes.

Advantage

- Provides 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

IV. METHODS

A. Data Sets

A dataset (or data set) is a collection of data usually presented in tabular form. Each column represents a particular variable. Each row corresponds to a given member of the dataset in question. It lists values for each of the variables, such as height and weight of an object. Each value is known as a datum. The dataset may comprise data for one or more members, corresponding to the number of rows. 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 attribute, Age in Year, Sex, Chest Pain Type, Fasting Blood Sugar, Electrocardiography (ECG), Exercise, Slope, Thallium Scan, Blood Pressure, Cholesterol, Oldpeak, Thalach, Height in cms, Weight in Kgs.

1. *Age*: This input consists of four data subsets i.e. Linguistic variable (Young, Mid, Old, Very old). The range of the data subsets for age is shown in Table I.

Table I. Age

input field	range	linguistic representation
age	<38	young
	33-45	mid
	40- 58	old
	>52	very old

2. *Chest Pain*: This input field has four Chest Pain types: Typical Angina, Atypical Angina, NonAngina, and Asymptomatic. One Patient can have only one type of Chest Pain at a time. To represent Chest Pain, 1= Typical Angina, 2 = Atypical Angina, 3= Non Angina and 4 = Asymptomatic.

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 6, June 2014

3. *Cholesterol*: This input field influences the result much more comparing to other input fields. Cholesterol can be Low Density Lipoprotein (LDL) and High density Lipoprotein (HDL). In our system, we only consider LDL. However, it is possible to consider HDL instead of LDL. We use only one type at a time. This field has four subsets. The range of the sets for Cholesterol is given in Table II.

Table II cholesterol

input field	range	linguistic representation
cholesterol	< 197 198-240 250-290 >291	low medium high very high

4. *Gender:* This input Field has two representations (Male and Female). 1 represents male and 0 indicates female.

5. *Blood Pressure*: Another important risk factor is Blood Pressure. It can be Systolic, Diastolic and Mean types. In our system, we consider Systolic Blood Pressure. It is possible to choose any type of Blood Pressure. This field has four sets. The ranges for the Linguistic variable representation are given in Table III.

Table III Blood Pressure

input field	range	linguistic representation
blood pressure	< 120 121- 139 140-159 >160	low medium high very high

6. *Blood Sugar*: This field
plays an important role in changing the results. It has two linguistic representations. Each fuzz variable is associated with prediction based on the range. The ranges of sets are given in Table IV.

Table IV Blood sugar

input field	range	linguistic representation
blood sugar	>=120 <120	yes(1) no(0)

7. Electrocardiography (ECG): ECG is measured by several waves in a graph paper such as T wave, Q wave, P wave, S wave of electric in pulse of Heart muscle. Normally the waves stay upper bound of the graph. If any of the wave goes down then it is thought as abnormality. In order to develop our system, we assume that S wave and T wave go down to represent the abnormality and this abnormality is named as ST_T abnormal in our designed system. This input field has three sets: Normal, ST_Tabnormal and Hypertrophy.

Table V Electrocardiography

input field	range	linguistic representation
Electrocardiography (ecg)	<0.4 0.4 - 1.8 >1.8	normal st_t abnormal hypertrophy

8. *Exercise:* This field indicates whether the patient need exercise test. This input field has two fuzzy sets representations. If the Patient requires an Exercise test then Value 1 is entered and if the patient does not need Exercise test value zero is entered. The linguistic representations are "Yes" for 1 and "No" for 0.

9. *Old Peak:* This field means ST depression induced by exercise relative to test. The meaning of ST depression is related to the ECG field. It means previously the patient's T wave and S wave in the ECG graph paper were down. Old Peak is necessary to assure the present condition of T wave and S wave of the ECG.

Table VI old peak

input field	range	linguistic representation
old peak	<2.0 1.5 - 4.2 >2.5	low risk terrible

10. *Thallium Scan*: Thallium scan is the redistribution of heart image. This input field has three linguistic representations: Normal, Reversible Defect and Fixed Defect. It depends on the hours that a heart image appears on the screen of the Gamma camera. This Gamma camera is able to detect radioactive dye in the body. To develop our system we assume that the linguistic representation of thallium scan in the Normal, the heart image appears within 3 hours, in fixed Defect heart image appears within 6 hours and in the Reversible Defect the heart image appears within 7 hours.

Table VII Thallium Scan

input field	range	linguistic representation
thal	3 6 7	normal fixed defect reversible defect

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 6, June 2014

B. Feature selection.

Feature selection identification of features that are relevant and not redundant for the prediction task . A common method of identifying relevant features is to compute the class-feature correlations for all the features present in the data and then select only those features with class-feature correlation values that are above a specified threshold. It is common practice, for Naïve Bayes classification, to discretise all numeric features so that all features for NB classification are categorical. Feature selection serves two main purposes. First, it makes training and applying a classifier more efficient by decreasing the size of the effective vocabulary. This is of particular importance for classifiers that, unlike NB, are expensive to train. Second, feature selection often increases classification accuracy by eliminating noise features.

C. Bayes Classification

Naïve Bayes is a statistical classifier which assumes no dependency between attributes. It attempts to maximize the posterior probability in determining the class. By theory, this classifier has minimum error rate but it may not be case always. However, inaccuracies are caused by assumptions due to class conditional independence and the lack of available probability data. Observations show that Naïve Bayes performs consistently after reduction of number of attributes.

According to Bayesian theorem

P (A|B) =P (A)*P (B/A)/P(B), Where P (B|A)=P(A∩B)/P(A)

Based on above formula, Bayesian classifier calculates conditional probability of an instance belonging to each class, and based on such conditional probability data, the instance is classified as the class with the highest conditional probability. In knowledge expression it has the excellent interpretability same as decision tree and is able to use previous data to build analysis model for future prediction or classification.

V. RESULTS

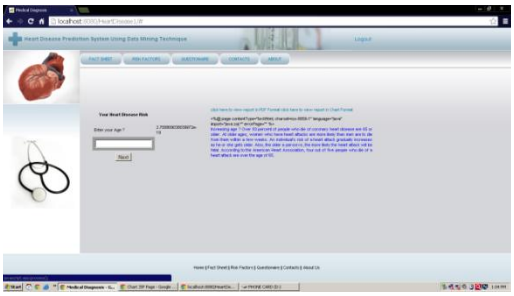

Fig 1. Questionnaire

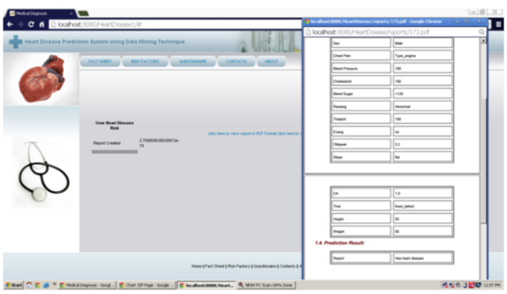

Fig 2. Report in PDF form.

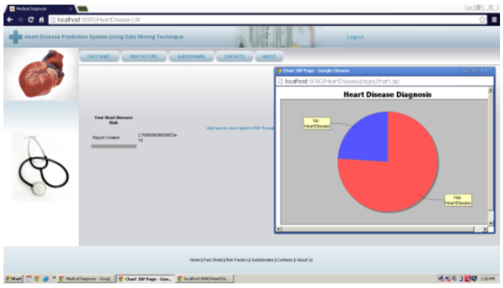

Fig 3. Pie chart for heart disease diagnosis.

VI. CONCLUSION

In this study, we have a tendency to develop a classification tree supported normal semi permanent HRV for risk assessment in patients full of CHF. The planned classifier separates lower risk patients from higher risk ones, victimisation normal semi permanent pulse variability (HRV) measures. The planned technique achieved the best performance in terms of accuracy rate and sensitivity. Compared to our previous studies, supported short measures, the classifier planned within the current studies achieved associate higher accuracy and sensitivity (85.4% versus 79.3% and 93.3% versus 82.4%, respectively), albeit with a lower specificity (75.0% versus 63.6%). This result leaded us to think about semi permanent HRV measures simpler for the individuation of upper risk patients than short ones.

REFERENCES

[1] J. L. Fleg, I. L. Pina, G. J. Balady, B. R. Chaitman, 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, andM. P. van den Berg, ―Prognostic value of heart rate variability and ventricular arrhythmias during 13-year follow-up in patients with mild to

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 6, June 2014

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.

BIOGRAPHY

Mr. Reddy Santosh Kumar is an M.Tech student of Rao Bahadur Y Mahabaleswarappa Engineering College, Bellary, India. Presently he is pursuing his M.tech(CSE) from this college and he received his B.tech from Ballari Institute of Technology & Management, Bellary. His area of interest includes Data Mining, Web Technology and Object Oriented Programming languages, all current trends and techniques in Computer Science.