ISSN 2320 - 2602
Volume 3, No.3, March 2014
International Journal of Advances in Computer Science and Technology
Available Online at http://warse.org/pdfs/2014/ijacst08332014.pdf

Risk Assessment Analysis of Patients Suffering From CHF Using improved Boosting Classification tree

S.Iniya Raghavi [1], Mrs.P.Lalitha [2]
[1]Research Scholar , Department of Computer Science ,Hinduthan College of Arts And Science , Nava India,Coimbatore-641028, Tamilnadu, India, iniyaraghavi@gmail.com.
[2]Assistant Professor, Department of Computer Science , Hindusthan College Of Arts and Science , Nava India , Coimbatore-641028, Tamilnadu, India, lalithavalliyangiri@ gmail.com

ABSTRACT
Due to the over past cast of years HRV have been well thoroughly is studied in numerous years, but it have less prediction result in case of work pressure throughout sleep time .In order to overcome these problem, proposed system accessibility to record dependable information designed for short epoch HRV has move up the concentration to discover the CHV signal. It separately classifies low level of HRV patients into high level of HRV patients through earlier measured HRV. Patient who has less HRV related to RR and it is classified as (NN/RR), it is used to measure other HRV that include a acceptable pointer superiority. In this paper C5.0 with boosted learning algorithm was proposed to classify HRV rates. It performs well efficient manner and it is well suited for each and every attributes in the dataset .This investigation efficiently gives unnecessary information in substance of nonlinear key and well-built differentiation in quantification of HRV afford self-governing information in possibility of CHF patients.

Keywords: Congestive Heart Failure (CHF), mining, decision tree, factorization boosting classification tree, heart rate variability (HRV).

1. 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. Heart failure cans origin a numeral of symptoms together with shortness of breath, leg enlargement, and work out small-mindedness. The heart failure problem is assessed by physical activates of user and established through echocardiography. Several numbers of blood tests also taken during this process, it helps to analysis the results of etiology diagnosis. Treatment depends beginning rigorousness and etiology of congestive heart failure (CHF).

Age of the populace and enhanced continued existence after discriminating myocardial infarction determination accelerates this movement, most important to an amplified concentration in demonstration for predictive stratification. Several indices measures are used to analysis of heart rate variability (HRV) have constantly provide considerable analytical information that related those subjects in efficient manner, self-regulating of experimental and well-designed consideration [1-3] .In earlier work CHF strictness can be accessed through the result of suggestive analysis for New York Heart Association (NYHA) [4] and simultaneously classification also performed ,it have been demonstrate to be a danger reason for humanity [5]. HRV is generally defined as the difference among normal heartbeats (RR intervals) [6] & heart attack rate difference, it is normally extorted from electrocardiographic signal (ECG) evidence all the way through a noninvasive procedure.

Previous studies have been proven that automatic assessment of HRV rate through non-invasive methods that utilization of HRV results in efficient manner and it have been developed in earlier methods [7-8] . The HRV is connected through sympathovagal equilibrium and it be capable of be a convenient and perfect technique to evaluate the property of discriminating do exercises and preparation on the automatic inflection of HR [9]. HRV is resulting beginning examination of successive beat to beat oscillations of occurrence province, which are essentially act as a go-between through the activities of autonomic nervous parts in the human body. Conversely, further neural, and metabolic factors force in addition stimulate alteration on together for HR and HRV parameters. Similarly these study assessment the estimated results of HRV reaction, throughout a progressive sequence ergometer investigation.

Heart rate variability (HRV) analysis result is normally used to estimate the results authority of the autonomic nervous system (ANS) on the heart. In this work present an efficient classification and feature selection method for long terms HRV analysis for ANS from individual person being HRV values at different conditions in CHF patients, predictable through New York Heart Association (NYHA) scale. So the major objective of this work is to analysis the effects of large levels of risks estimation in CHF patients and on HRV-work speed curvature throughout progressive implement. During the examination of HRV work status it would be change to the increasing and to the accurate commands, by means of associated diminution in heart rate in subordinate maximal phase.

In this paper proposed an Ada boost based C5.0 decision tree algorithm that accurately pruning the irrelevant features from HRV and generates different rules for gathering classification results assessment of patient results from CHF .It shows that it achieve more accurate rules generated from Ada C5.0 decision tress algorithm ,improved speed and lesser error rate and well efficient suited method for best feature selection result to CHF analysis results in the way to estimation of large and small dataset for CHF patient records . Then we compared the performance accuracy of CHF patient's analysis results from HRV beat rates through a normal classification algorithm such as CART and other classification methods that can be carried out exercise to HRV records of each patients. Then preferred an Ada Boost C5.0 classification result reduce the results of misclassified HRV rate with less error and as the rare cases of HRV also exactly classified as accurate manner.

2. BACKGROUND STUDY

Number of studies [10-11] recognized the correlation of HRV measures as well as the NYHA categorization scale. In our preceding papers, recognize the established to HRV may be used in the direction of distinguish CHF by way of short-term [12] or long-term measures [13]. Additionally, planned a classifier base on short-term HRV procedures in the direction of individuate rigorousness of CHF [14].

In earlier work uses a short-term detrended fluctuation analysis (DFA) has formerly be experienced in a number of populations of CHF as well as situation myocardial infarction patients, as well as was a authoritative self-governing predictor of mortality [15,16].

The physiological situation of DFA has been lately examine through [17] demonstrate so as to considerate as well as vagal concern draw out through moreover pharmacological involvement in the key of heart rate dynamics.

Additional contentious is the analysis of the LF aspect which is calculated through several [18] as a marker of kindly modulation as well as through others with the purpose of includes together considerate and vagal control. This inconsistency is appropriate to the reality that in a number of condition, related through considerate excitation, a reduce in the entire control of the LF part is experimental. Actually, it has been lately recognized that in patients following discriminating myocardial indices of HRV maintain their predictive power in the beta-blocking era. In the same way, [19] assessing the aptitude of short-term results of HRV in forecasting unexpected death in CHF patients.

3. ADABOOST C5.0 TREE ALGORITHM

In this work present an efficient feature selection and classification algorithm for identification of inequalities, although usually ignored, component of biological system, these proposed technique be not all right recognized through clinicians, probably since present is no clear physiological accepting of their fundamental mechanisms, most important to various notion of statistical wizardry as well as increase doubts programmed their significance toward medical practice. This ambiguity is more compounded through the great number of dissimilar nonlinear method projected awake to at the present. Still the cross-validated approximation may possibly give restricted information during little as well as excessive dataset, feature selection (FS) develop classification for the reason that it make possible to think about collection of n feature through n smaller than the range of presented samples.

Enhanced decision tree C5.0 tools a stepwise FS; it could probably take place with the purpose of one feature be disqualified for the reason that further variables masked its result. This might probably be mostly serious in little as well as unbalanced dataset. In classify to compact through masking as well as to be assured so as to the tree incorporated greatest subset of features, implement the theoretical comprehensive search method [20], examine every one the practicable grouping of k away of N features. Still the cross-validated approximation may possibly give restricted information during little as well as excessive dataset, feature selection (FS) develop classification for the reason that it make possible to think about collection of n feature through n smaller than the range of presented samples.

The examiner work available currently considers C5.0. The improvement of C5.0, which demonstrate improved performance because evaluate in the direction of other existing decision tree methods. C5.0 algorithm in the direction of constructs any a decision tree otherwise a ruling set. A C5.0 representation mechanism through split example based programmed the field consequently as to make available maximum information gain. Every subsample separate through unique split be consequently split yet again, commonly base lying on a dissimilar field, also procedure do again in anticipation of the subsamples cannot be there divide each added. Lastly, the lowest-level divide are estimated for HRV data, as well as individuals to do not provide considerably significance of the illustration be there pruned. C5.0 conserve build two types of models. A decision tree is simple details of divide establish through the algorithm. Every one terminal node describe a specified subset of features for HRV data, as well as every case during the training data be in the right place to accurately individual life-threatening node in the tree.

Decisions trees are constructing as a result of desire the import which decreases the training error. But those items do not required diminish the examination inaccuracy as well. Therefore, it formulates intelligence to disregard a number of items which show the way to smaller training error however we are uncertain concerning their examination error. The results illustrate with the intention of items which are very accepted pose lower indecision regarding their examination error, notion the motivation is not understandable for us.

A decision tree is a necessary justification of opening predictable through the algorithm. Each one life-threatening node in the hierarchy indicate a particular set of preparation data samples all the way through encrypted layout as well as every case in the training data in the tree. It generates a ruling set to determine individual information all the way through confidentiality confined. Rule set are fashioned by means of with the attribute that are connected to user records and, in a procedure, communicate to a make straightforward explanation of the information well-known in the C5.0 decision tree. Since of the approach rule sets attempt, they do not enclose the related property as decision trees.

The essentially considerable distinction is by means of the purpose of rule set, supplementary than particular rule capacity be applicable designed for each challenging record. If frequent rules be appropriate, each instruction attains values that are weight parameters for each and every rule, and the finishing computation is strong-minded through arrangement of special weight values end result intended for each and every one rules by means of heart rate variability (HRV). If no other rules are matched to heart rate variability (HRV) test than we perform random selection process with largest IG values as best features and important feature for HRV classification of many attribute values that corresponds to similar HRV results it is normally represented in the following manner

$$\text{“}if\ A\ and\ B\ and\ C\ and\ \ldots then\ class\ X\text{”},$$

In the specified rules set heart rate variability (HRV) if the corresponding data HRV is related to set A and it is classified as HRV rate for class A or else it is classified as class B rule set ,If it is matched to class C it is classified as none of the data that corresponds to heart rate variability (HRV), assets of all case consequently as to capacity be appropriate in the direction of its course group are accessible, still although a amount of belongings capacity enclose unfamiliar principles. C5.0 is proficient to understanding with a number of numbers of attributes. Rule set are fashioned by means of with the attribute that are connected to user records and, in a procedure, communicate to a make straightforward explanation of the information well-known in the C5.0 decision tree. Since of the approach rule sets attempt, they do not enclose the related property as decision trees. Some of the decision tree algorithm that can be represented and applicable to number of applications through C4.5 information entropy measure to analysis results [21] .

The usual description of decisions trees are constructing as a result of desire the import which decreases the training error. But those items do not required diminish the examination inaccuracy as well. The training heart rate variability (HRV) if the corresponding data HRV is related to set A and it is classified as HRV rate for class A or else it is classified as class B rule set ,until all HRV dataset that corresponds to same class . Subsequently every one pathway beginning the origin to a leaf node corresponds in the direction of specified rules set heart rate variability (HRV), and the complete decision tree communicate in the direction of a group of specified rules set heart rate variability (HRV).

In this work HRV based classification is rely on general and improved boosting based classification decision tree algorithm and it is suitable to both discrete and continuous attributes that corresponds to HRV dataset, it is somewhat different from normal C5.0 algorithm in the direction of added rule pruning to accomplish more precise creation rules and poorer error rate, it is supplementary appropriate for categorization of huge data sets [22].

This function is rely on linear discriminated analysis refer to [23] ,it divides HRV data samples into two sets namely training and testing Z_1 &Z_2. Based on the beyond methods, in this manuscript, we make use of stepwise Fisher's discriminated examination, and representation of expression is shown below:

$$Z = C_1X_1 + C_2X_2 + \cdots + C_mX_m \quad (1)$$

In equation (1), C_i is defined as the coefficient threshold value for HRV rate measurement, X_i is the original heart rate variability (HRV) measures data samples and it is categorized into two dataset which is training and testing data .IF the classes result of HRV values Z are larger that the class value of ZC for HRV classification, it is classified HRV samples as class one; or else it is classified as type 2; and if is same to threshold of HRV classified as 1 or 2. The equation of ZC is listed as follows,

$$ZC = \frac{\overline{Z_1} + \overline{Z_2}}{2} \qquad (2)$$

The improved decision tree C5.0 based increasing and abbreviates process is frequent heart rate variability (HRV) measures upto highest HRV measures results are estimated at every time of process and measures results are examined through a difficult set. The misclassified HRV rate measures results are estimated as percentage of wrongly classified data samples from HRV subsets. This HRV is analyzed until more number of HRV is correctly classified in the tree.

Algorithm 1:Heart rate variability (HRV) analysis using improved boosting classification tree

Input node n, HRV dataset D, partitioning method of $HRPF_n()$;

Output Decision tree on node n as its root node, HRV dataset D and partitioning method of $HRPF_n()$;

1) Procedure Build Tree
2) Initialization root node n
3) Find decision feature n according to $HRPF_n()$; on Dataset D
4) If node n meets with conditions of partition
5) According to decision feature, Partition Dataset D into D_1 and D_2, and creates two subnodes of n, namely n_1 and n_2
6) Build Tree $(n_1, D_1, HRPF_n()\ )$;
7) Build Tree $(n_2, D_2, HRPF_n()\ )$;
8) End if
9) End procedure
10)

4. EXPERIMENTAL RESULTS

In this section measure the performance accuracy of proposed classification result of CHF class with HRV related sample dataset through a normal CART. The uncommon class (mild CHF) was oversampled through generate original artificial uncommon class. To measure performance accuracy of classification result through various parameters such as $F1$ measure, Accuracy, Precision, Sensitivity and specificity was selected.

Table 1 : Binary Classification Performance Measures

Measure	Abbreviation	Formula
F_1	F_1	$\frac{2TP}{2TP + FP + FN}$
Accuracy	ACC	$\frac{TP + TN}{TP + TN + FP + FN}$
Precision	PRE	$\frac{TP}{TP + FP}$
Sensitivity	SEN	$\frac{TP}{TP + FN}$
Specificity	SPE	$\frac{TN}{FP + TN}$
TP=Number of HRPs correctly classified TN=Number of LRPs correctly classified FP=Number of LRPs incorrectly classified as HRPs FN=Number of HRPs incorrectly classified as LRPs		

In comparative to the method, the thorough examine designed for FS enhanced classification methods and it can be compared to CART with efficient features to Boosted C5.0 representation acquire by means of each and every one the features. The improved adaboost decision tree algorithm achieves highest performance accuracy in terms of Classification Accuracy, Precision, Sensitivity, F1 measure and less specificity, which is one of the most important specified parameters to estimate HRP class results shown in Figure 1 and Figure 2.

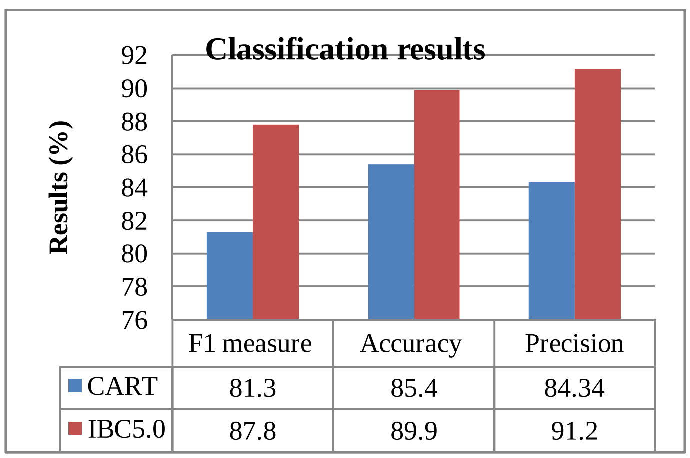

Figure 1: Classification Results with F1, Accuracy and Precision

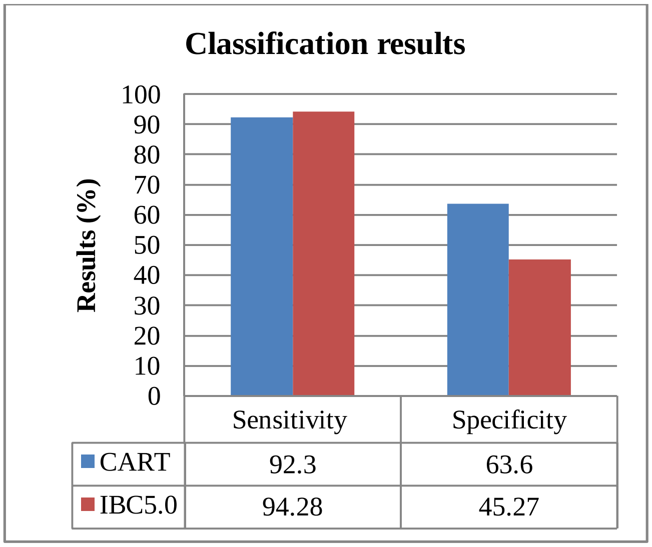

Figure 2: Classification results with Sensitivity and Specificity

5. CONCLUSION AND FUTURE WORK

In conclusion, this investigation says that proposed well efficient C5.0 gives unnecessary information in substance of nonlinear key and well-built differentiation in quantification of HRV afford self-governing information in possibility of CHF patients. It is supposed to be stressed out, though, that our examination be not considered as a predictive learning although, moderately, it's aspire was to formulate available helpful information designed for a dangerous and further well-organized make use of the huge amount of HRV manifestation projected consequently future in medicinal study. Finally we conclude and compare that long term HRV evaluation results permit superior risk patients to be there present well-known beginning lesser hazard ones. The Adaboost C5.0 tree learning and classification tree is developed and it accomplish an accuracy results of 89.9% , Precision rate of 91.2 % a sensitivity rate of 94.28% and a specificity rate of 45.27% obtained from various combination of features .

References

1. Galinier M, Pathak A, Fourcade J, Androdias C, Curnier D, Varnous S, Boveda S, Massabuau P, Fauvel M, Senard JM, Bounhoure JP, **" Depressed low frequency power of heart rate variability as an independent predictor of sudden death in chronic heart failure",** *Eur Heart J* 2000;21:475-482.

2. La Rovere MT, Pinna GD, Maestri R, Mortara A, Capomolla S, Febo O, Ferrari R, Franchini M, Gnemmi M, Opasich C, Riccardi PG,Traversi E, Cobelli F**," Short-term heart rate variability strongly predicts sudden cardiac death in chronic heart failure patients**", *Circulation* 2003;107:565- 570.

3. Guzzetti S, La Rovere MT, Pinna GD, Maestri R, Borroni E, Porta A, Mortara A, Malliani A**," Different spectral components of 24 h heart rate variability are related to different modes of death in chronic heart failure",** *Eur Heart J* 2005;26:357-362.

4. 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.

5. 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 Chronic Heart Failure (OPTIME-CHF) study Circulation*, vol. 111, no. 19, pp. 2454–2460, May 17, 2005.

6. 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.

7. H. M. Stauss, **"Heart rate variability,"** *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 285, pp. R927–R931, 2003.

8. A. Hautala, **"Effect of physical exercise on autonomic regulation of heart rate,"** *Academic Dissertation (Faculty of Medicine) – University of Oulu*, Finland, 2004, pp. 76.

9. L. Mourot, M. Bouhaddi, S. Perrey, J. D. Rouillon, and J. Regnard, **"Quantitative Poincaré plot analysis of heart rate variability: effect of**

endurance training," *Eur J Appl Physiol,* vol. 91, pp. 79–87, 2004

10. 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.

11. G. N. Arbolishvili, V. Y. Mareev, Y. A. Orlova, and Y. N. Belenkov, **"Heart rate variability in chronic heart failure and its role in prognosis of the disease,"** *Kardiologiya*, vol. 46, no. 12, pp. 4–11, 2006.

12. L. Pecchia, P. Melillo,M. Sansone, and M. Bracale, **"Discrimination power of short-term heart rate variability measures for CHF assessment,"** *IEEE Trans. Inf. Technol. Biomed.,* vol. 15, no. 1, pp. 40–46, Jan. 2011.

13. P. Melillo,R. Fusco,M. Sansone,M.Bracale, and L. Pecchia, **"Discrimination power of long-term heart rate variability measures for chronic heart failure detection,"** *Med. Biol. Eng. Comput*., vol. 49, no. 1, pp. 67–74, 2011.

14. L. Pecchia, P. Melillo, and M. Bracale, **"Remote health monitoring of heart failure with data mining via CART method on HRV features,"** *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 800–804, Mar. 2011.

15. Huikuri HV, Makikallio TH, Peng CK, Goldberger AL, Hintze U, Moller M **," Fractal correlation properties of R-R interval dynamics and mortality in patients with depressed left ventricular function after an acute myocardial infarction",** Circulation 2000;101:47-53.

16. Tapanainen JM, Thomsen PE, Kober L, Torp-Pedersen C, Makikallio TH, Still AM, Lindgren KS, Huikuri HV**," Fractal analysis of heart rate variability and mortality after an acute myocardial infarction",** *Am J Cardiol* 2002; 90:347-352.

17. Tulppo MP, Kiviniemi AM, Hautala AJ, Kallio M, Seppanen T, Makikallio TH, Huikuri **," Physiological background of the loss of fractal heart rate dynamics"**,Circulation 2005;112:314-319.

18. Tsutamoto T, Wada A, Maeda K et al.,"**Plasma brain natriuretic peptide level as a biochemical marker of morbidity and mortality in patients with asymptomatic or minimally symptomatic left ventricular dysfunction",** *Comparison with plasma angiotensin II and endothelin-1*. Eur Heart J 1999; 20: 1799–807.

19. Jokinen V,Tapanainen JM, Seppanen T, Huikuri HV,"**Temporal changes and prognostic significance of measures of heart rate dynamics after acute myocardial infarction in the beta-blocking era",** Am J Cardiol 2003;92:907-912.

20. A. K. Jain, R. P. W. Duin, and M. Jianchang, **"Statistical pattern recognition: a review,"** *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000

21. Xindong Wu, Vipin Kumar,J.Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey, J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael steinbach, David J. Hand, and Dan Steinberg, 2007, **"A survey paper on Top 10 algorithm in data mining",** published by Springer ,2007.

22. P. Xiong, **"Data Mining Algorithms and Practice of Clementine",** *Tsinghua University Press*, Beijing, China, 2011.

23. J. L. Xia and H. W. Jiang,**" Medical Statistics",** *People′s Medical Publishing House*, Beijing, China, 2005.