International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 04 | Apr-2016 www.irjet.net p-ISSN: 2395-0072

MACHINE LEARNING USING FOR CLASSIFICATION OF HEART FAILURE

***Dr.V.Selvi*[1] and *Ms.M.Juliet vinitha*[2]**
[1]Assistant Professor and [2]M.Phil Scholar
Department of Computer Science
Mother Teresa Women's University
Kodaikanal.

Abstract - *Physicians classify patients into those with or without a specific disease. Classification trees are frequently used to classify patients according to the presence or absence of a disease. In the data-mining and machine learning, alternate classification schemes have been developed. These include Regression Tree, Random Forest, Boosting and Support Vector Machines (SVM). To analyze the heart failure, Regression Tree and SVM methods are compared in this paper. In Regression Tree method, it takes 30 sec to evaluate the result accurately and in SVM Learning Optimization, the result is evaluated in less than 30 sec. This paper shows that out of these two classification models SVM predicts heart disease with highest accuracy.*

***Key Words*: Data mining, Heart disease, Regression tree, Support vector machines.**

1. INTRODUCTION

There is an increasing interest in using classification methods in clinical research. Accurate classification of disease states (disease present/absent) or subtype allows subsequent investigations, treatments, and interventions to be delivered in an efficient and targeted manner. In the data mining and machine learning literature, alternatives to and extensions of classical classification trees have been developed in recent years. This methods include classification trees, random forests, and boosted trees. Alternate classification methods include support vector machines. In patients with acute heart failure (HF) there are two distinct subtypes: HF with preserved ejection fraction (HFPEF) vs. HF with reduced ejection fraction (HFREF). The distinction between HFPEF and HFREF is particularly relevant in the clinical setting. While the treatment of HFREF is based on a multitude of large randomized clinical trials, the evidence-base for the treatment of HFPEF is much smaller, and more focused on related comorbid conditions.

2. HEART DISEASE

The heart is important organ of human body part. A number of factors have been shown that increases the risk of Heart disease.

[1] Family history
[2] Smoking
[3] Poor diet
[4] Smoking
[5] Poor diet
[6] High blood pressure
[7] High blood cholesterol
[8] Obesity
[9] Physical inactivity
[10] Hyperb tension

Factors like these are used to analyze the Heart disease.

3. LITERATURE SURVEY

An Intelligent Heart Disease Prediction System (IHDPS) is developed by using data mining techniques Support vector machines, Regression tree was proposed by Sellappan Palaniappan et al. [3]. Each method has its own strength to get appropriate results. To build this system hidden patterns and relationship between them is used. It is web-based, user friendly & expandable.

To develop the multi-parametric feature with linear and nonlinear characteristics of HRV (Heart Rate Variability) a novel technique was proposed by Heon Gyu Lee et al. [5]. To achieve this, they have used several classifiers e.g. Bayesian Classifiers, CMAR (Classification based on Multiple Association Rules), C4.5 (Decision Tree) and SVM (Support Vector Machine).

The prediction of Heart disease, Blood Pressure and Sugar with the aid of neural networks was proposed by Niti Guru et al. [4]. The dataset contains records with 13 attributes in each record. The supervised classification i.e. Support vector machines and Regression tree with back

propagation algorithm is used for training and testing of data.

The problem of identifying constrained association rules for heart disease prediction was studied by Carlos Ordonez [7]. The resultant dataset contains records of patients having heart disease. Three constraints were introduced to decrease the number of patterns [6]. They are as follows: The attributes have to appear on only one side of the rule. Separate the attributes into groups.

i.e. uninteresting groups. Franck Le Duff et al. [9] builds a Regression tree with database of patient for a medical problem.

Latha Parthian et al. [10] HF with preserved ejection fraction (HFPEF) vs. HF with reduced ejection fraction (HFREF) model used Support vector machines and Regression tree.

4. PROPOSED METHOD

The comparison of predictive ability of different regression method on accuracy was assessed using two different metrics. First, we calculated the area under the receiver operating characteristic (ROC) curve (abbreviated as the AUC), which is equivalent to the c-statistic [28,30]. Second, we calculated the Brier Score [28] (mean squared prediction error), which is defined as $\frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i - Y_i)^2$ where N denotes the sample size, *Pi* is the predicted probability of the outcome and *Yi* is the observed outcome (1/0). We used the value problem function from the *Design* package to estimate these two measures of predictive accuracy. The SVM Accuracy of classification was assessed using sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

5. DATA SOURCE

The Enhanced Feedback for Effective Cardiac Treatment (EFFECT) Study was an initiative to improve the quality of care for patients with cardiovascular disease in Ontario [26,27]. The EFFECT study consisted of two phases. During the first phase, detailed clinical data on patients hospitalized with HF between April 1, 1999 and March 31, 2001 at 103 acute care hospitals in Ontario, Canada were obtained by retrospective chart review. During the second phase, data were abstracted on patients hospitalized with HF between April 1, 2004 and March 31, 2005 at 96 Ontario hospitals. Data on patient demographics, vital signs and physical examination at presentation, medical history, and results of laboratory tests were collected for this sample.

We examined the predictive accuracy of each method using the EFFECT-1 sample as the model derivation sample and the EFFECT-2 sample as the model validation sample. Using each prediction method, a model was developed for predicting the probability of HFPEF using the subjects in the EFFECT-1 sample. We then applied the developed model to each subject in the EFFECT-2 sample to estimate that subject's predicted probability of having HFPEF. Note that the derivation and validation samples consist of patients from the same jurisdiction (Ontario). Furthermore, most acute hospitals that cared for HF patients were included in both of these two datasets. However, the derivation and validation samples are separated temporally (1999/2000 and 2000/2001 vs. 2004/2005). The study design ensured that there was very little overlap in patients between the two study periods.

For each subject in the validation sample, a true HF sub-type was observed (HFPEF vs. HFREF) and a classification was obtained (HFPEF vs. HFREF) for each classification method developed in the EFFECT-1 sample.

Comparison of patients with HFPEF and HFREF in EFFECT-1 and EFFECT-2 samples

Variable	EFFECT-1 sample		
	HFREF	HFPEF	P-Value
Age (years)	75.0 (66.0-81.0)	77.0 (70.0-83.0)	<.001
Male	1,547 (61.2%)	423 (36.2%)	<.001
Heart rate	96.0 (78.0-113.0)	90.0 (74.0-110.0)	<.001
Current Smoker	415 (16.4%)	130 (11.1%)	<.001
Cancer	282 (11.2%)	132 (11.3%)	0.893
Hemoglobin	12.7 (11.3-14.1)	12.3 (10.7-13.5)	<.001

Comparison of patients with HFPEF and HFREF in EFFECT-1 and EFFECT-2 samples

Variable	EFFECT-2 sample		
	HFREF	HFPEF	P-Value
Age (years)	75.0 (66.0-81.0)	77.0 (70.0-83.0)	<.001
Male	1,686 (60.7%)	643 (37.0%)	<.001
Heart rate	94.0 (76.0-112.0)	86.0 (70.0-105.0)	<.001
Current Smoker	382 (13.8%)	157 (9.0%)	<.001
Cancer	292 (10.5%)	186 (10.7%)	0.851
Hemoglobin	12.7 (11.2-14.0)	12.1 (10.7-13.4)	<.001

6. RESULT AND DISCUSSION

The sensitivity, specificity, PPV, and NPV of the different classification methods are reported in Table1. The sensitivity in the different models developed in the EFFECT-1 sample ranged from a low of 0.378 for the Regression tree to a high of 0.500 for the Support vector machine.

We compared the performance of modern classification and SVM methods with classification and regression trees to classify patients with HF into one of two mutually exclusive categories HFPEF (HF with preserved ejection fraction) vs. HFREF (HF with reduced ejection fraction), or to predict the probability of the presence of HFPEF. We found that modern classification methods offered improved performance over conventional classification trees for classifying HF patients according to disease subtype.

Table 1- Accuracy of prediction in EFFECT-1 sample

Prediction method	AUC	Brier Score
Regression tree	0.683	0.2152
Support vector machine	0.401	0.2079

Table 2-Sensitivity and specificity of classification in EFFECT-2 sample

Classification Method	Sensitivity	Specificity	PPV	NPV
Regression tree	0.683	0.897	0.696	0.709
Support vector machine	0.401	0.887	0.616	0.697

7. CONCLUSION

The overall objective of our work is to predict more accurately the presence of heart disease. In this paper, two more input attributes obesity and smoking are used to get more accurate results. Two data mining classification were applied namely Regression Tree & support vector machines. From results it has been seen that support vector machines provides accurate results as compare to Regression Tree.

8. REFERENCE

[1] Breiman, L.; Friedman, JH.; Olshen, RA.; Stone, CJ. Classification and Regression Trees. Chapman & Hall/CRC; Boca Raton: 1998.

[2] Austin PC. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. Statistics in Medicine. 2007; 26(15):2937–2957. [PubMed: 17186501]

[3] Clark, LA.; Pregibon, D. Tree-Based Methods. In: Chambers, JM.; Hastie, TJ., editors. Statistical Models in S. Chapman & Hall; New York, NY: 1993. p. 377-419.

[4] Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning Data Mining, Inference, and Prediction. Springer-Verlag; New York, NY: 2001.

[5] Gansky SA. Dental data mining: potential pitfalls and practical issues. Advances in Dental Research. 2003; 17:109–114. [PubMed: 15126220]

[6] Harrell, FE, Jr. Regression modeling strategies. Springer-Verlag; New York, NY: 2001.

[7] Ho JE, Gona P, Pencina MJ, Tu JV, Austin PC, Vasan RS, Kannel WB, D'Agostino RB, Lee DS, Levy D. Discriminating clinical features of heart failure with preserved vs. reduced ejection fraction in the community. European Heart Journal. 2012; 33(14):1734–1741. [PubMed: 22507977]

[8] Steyerberg, EW. Clinical prediction models: a practical approach to development, validation, and updating. Springer; New York, NY: 2009.

[9] Zhou, X.; Obuchowski, N.; McClish, D. Statistical Methods in diagnostic medicine. Wiley-Interscience; New York: 2002.

[10] Austin PC, Lee DS, Steyerberg EW, Tu JV. Regression trees for predicting mortality in patients with cardiovascular disease: what improvement is achieved by using ensemble-based methods? Biometrical Journal. 2012;54(5):657–673.10.1002/bimj.201100251 [PubMed:22777999]