An Overview of Data Mining Classification Methods in Aortic Stenosis Prediction

T. Revathi, P. Sumathi

Abstract- There is a huge amount of data in medical science industry. But most of this data is not mined to find out the hidden information. To discover those hidden information, advanced data mining techniques are used. Models developed from these techniques are seemed to be very useful for medical practitioners to take effective decision. In this research paper data mining classification techniques Decision Tree and Support Vector Machine (SVM) are analyzed on Aortic Stenosis disease dataset. Performance of these techniques is compared by sensitivity, specificity, accuracy, error rate, True Positive Rate and False Positive Rate. As per our results error rates for Decision Tree and SVM are 0.2755 and 0.1488 respectively. Accuracy of Decision Tree and SVM are 79.05% and 85.12% respectively. Our analysis shows that among these two classification models SVM predicts Aortic Stenosis disease with least error rate and highest accuracy.

Key words: ***Heart disease, Aortic Stenosis, Data Mining techniques, decision tree and support vector machine.***

I. INTRODUCTION

A) Heart Disease

The term "heart diseases" refers to diseases of the heart and the blood vessel system within it. Congestive heart failure, cardiomyopathy, and valvular heart disease are among the heart disease types [1]. There are more than 50 types, each one affecting different areas of the heart and/or the blood vessel system within it. Some people are born with heart disease known as congenital heart disease; others develop during a person's lifetime [2]. As there are so many heart disease types, it may be helpful to separate them into problems in the:

- Heart chambers
- Heart muscle itself
- Heart valves
- Coronary arteries and coronary veins
- Electrical system
- Heart lining.

B) Heart Disease Types Affecting the Valves

Different kinds of heart disease that can affect the heart valves (valvular heart disease) include [4]:

- Mitral valve regurgitation
- Mitral valve prolapsed
- Aortic stenosis
- Aortic regurgitation
- Tricuspid stenosis
- Tricuspid regurgitation
- Rheumatic heart disease

In this paper, the predictive process is carried out for Aortic Stenosis.

Manuscript Received on August 2014.

T. Revathi, Research Scholar, Manonmaniam Sundaranar University, Tirunelveli, India.

Dr. P. Sumathi, Asst. Prof., PG & Research Department, Department of Computer Science, Govt. Arts College, Coimbatore, India.

C) Aortic Stenosis

Aortic stenosis is a narrowing of the aortic valve in the heart. This forces the heart to work harder to pump blood out of the left ventricle to the rest of the body. In many cases, symptoms may include chest pain, fainting, heart failure, and heart arrhythmias.

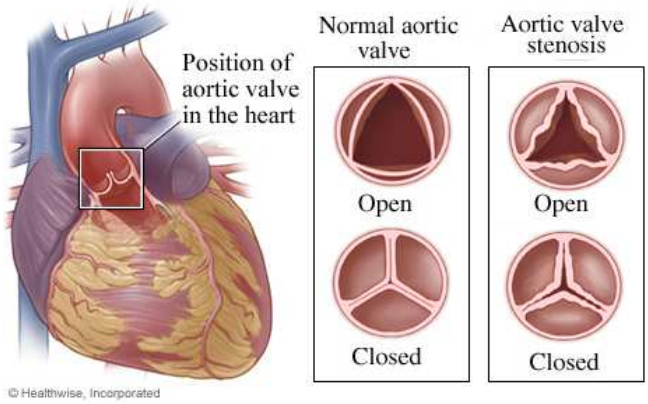

Fig. 1 Aortic Stenosis

The treatment for this condition may include options such as lifestyle changes, medications, or even surgery. The aortic valve is on the left side of the heart. It separates the left ventricle and the entrance to the aorta, the artery that carries blood to the body. Valves are like doors that open and close. They open to allow blood to flow through to the next chamber or to one of the arteries, and then they shut to keep blood from flowing backwards. The aortic valve opens when the left ventricle contracts. This allows blood to flow into the aorta. The aortic valve closes to prevent blood from coming back into the left ventricle. In aortic stenosis, the aortic valve does not open all the way. This forces the heart to work harder to pump blood out of the left ventricle to the rest of the body. Over time, the left ventricle thickens and becomes less effective in pumping blood. Aortic stenosis can be caused by several health issues, including:

- Conditions a person is born with (congenital defects)
- A bicuspid aortic valve (a valve that only has two leaflets instead of three)
- Calcium buildup on the valve
- Inflammation from rheumatic fever.

Most of these conditions cause a thickening of the aortic valve, making it unable to fully open. Aortic stenosis is three times more common in males than females. Symptoms usually begin after the age of 60, but can occur in children and young adults due to congenital conditions, such as a bicuspid valve. Symptoms of aortic stenosis include:

- Breathlessness with activity
- Chest pain, angina-type
- Fainting, weakness, or dizziness with activity

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

- Sensation of feeling the heart beat (palpitations)

In infants and children, symptoms include:

- Becoming easily tired with exertion (in mild cases)
- Failure to gain weight
- Poor feeding
- Serious breathing problems that develop
- within days or weeks of birth (in severe cases)

Children with mild or moderate aortic stenosis may get worse as they get older. They also run the risk of developing a heart infection (bacterial endocarditic).

D) Diagnosing Aortic Stenosis

Doctors diagnose aortic stenosis based on the patient's symptoms, a physical exam, and certain tests and procedures. The tests used to diagnose aortic stenosis includes Chest x-ray, Doppler echocardiography, Exercise stress testing, Left cardiac catheterization, MRI of the heart, Transesophageal echocardiogram (TEE), Electrocardiogram and Cardiac catheterization. Complications caused by aortic stenosis are Angina (chest pain), Arrhythmias, Endocarditic, Fainting (syncope), Left-sided heart failure and Left ventricular hypertrophy (heart wall thickening) caused by the extra work of pushing blood through the narrowed valve[3][5].

II. RELATED WORK

In [6] four different supervised machine learning algorithms i.e. RIPPER, decision tree, ANN, SVM have been used for analyzing the dataset. The experiment is performed using training data set consists of 303 records with 14 different attributes. On 296 records all the above techniques were carried out. The dataset is divided into two parts that is 70% of the data are used for training and 30% are used for testing. Based on the experimental results, it is clear that the classification accuracy of SVM is better compared to other algorithms. In [7], the overview of predictive data mining technique for heart disease has been given.

III. AORTIC STENOSIS PREDICTION MODELS

A. Decision Trees

Decision trees are powerful classification algorithms. Popular decision tree algorithms include Quinlan's ID3, C4.5, C5, and Breiman et al.'s CART. As the name implies, this technique recursively separates observations in branches to construct a tree for the purpose of improving the prediction accuracy [8][9]. In doing so, they use mathematical algorithms to identify a variable and corresponding threshold for the variable that splits the input observation into two or more subgroups. This step is repeated at each leaf node until the complete tree is constructed [10][11]. The most commonly used mathematical algorithm for splitting includes Entropy based information gain (used in ID3, C4.5, C5), Gini index (used in CART), and Chi-squared test (used in CHAID). Fig 1 shows the example of decision tree on patient diagnosis.

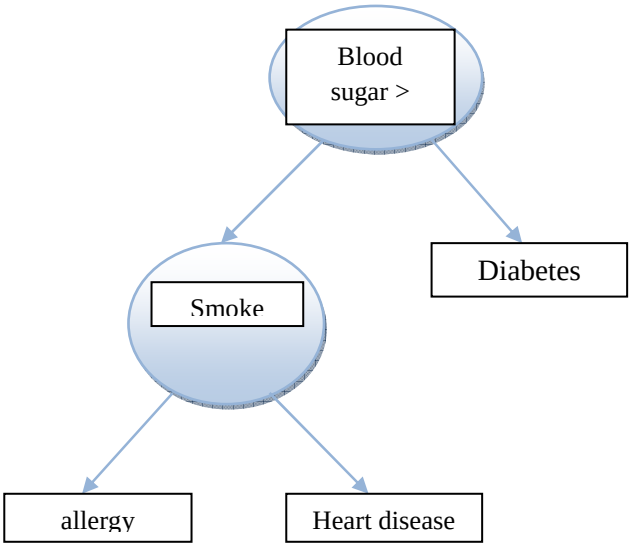

Fig. 2 Decision Tree for Patient Diagnosis

B. Support Vector Machine

The SVM is a state-of-the-art maximum margin classification algorithm rooted in statistical learning theory. SVM is method for classification of both linear and non-linear data. It uses a non-linear mapping to transform the original training data into a higher dimension. Within this new dimension it searches for linear optimal separating hyper plane. With an appropriate nonlinear mapping to a sufficiently high dimension, data from two classes can always be separated by a hyper plane. The SVM find this hyper plane using support vectors and margins [12]. SVM performs classification tasks by maximizing the margin separating both classes while minimizing the classification errors. Fig 3 shows SVM topology in hyperspace:

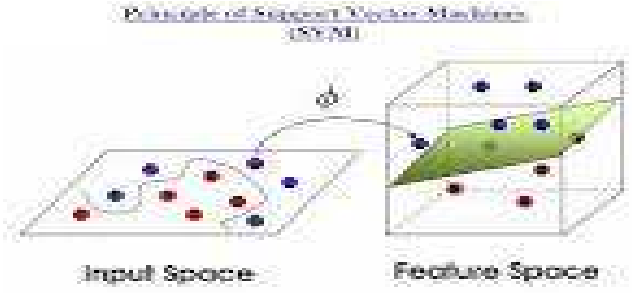

Fig. 3 SVM Topology

IV. DATA SET

To compare these data mining classification techniques dataset from UCI repository is used [13]. This dataset has 11 attributes and are listed below:

Table 1 Attributes for Aortic Stenosis

No	Attribute name	Medical meaning
1	AGE	Patient Age
2	SEX	Gender
3	CP	Chest pain type
4	CHOL	Cholesterol
5	SMOKE	Smoking habit
6	FBS	Fasting blood sugar > 120 mg/dl
7	DIAB	Diabetes
8	HTA	Hypertension
9	PSTROKE	Prior stroke
10	PSUR	Prior surgery
11	NUM	Class

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

Description

The attribute name AGE is calculated in years. If the gender of the patient is male, its denoted as 1 and if its female, its denoted as 0. Chest pain type is identified as 1= typical angina, 2= atypical angina, 3= non-anginal pain, 4= asymptotic. Cholesterol is measured in mg/dl. The attributes fasting blood sugar, diabetes, hypertension, prior stroke and prior surgery are given as 1 if exists and given as 0 if it does not exists. If the final class is 1, its concluded that the patient has heart disease and if its 0, its concluded that the patient is healthy.

V. EXPERIMENTAL SETUP AND RESULT ANALYSIS

Experiments were conducted with Weka 3.6.0 tool and Data set of 11 attributes. All attributes are categorized and made simple by removing the redundancy. Observations say that the Support vector machine technique outperforms decision tree technique after incorporating feature subset selection. Table 2shows the accuracy of the SVM obtained from the experiment.

Table 2 Results Obtained from Experiment

Data mining techniques	Accuracy	Error rates
Decision trees	79.05%	0.2755
Support vector machine	85.12%	0.1488

VI. ISSUES AND CHALLENGES

Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely and efficiently. The automation of the same would be highly beneficial[14]. Clinical decisions are often made based on doctor's intuition and experience rather than on the knowledge rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Data mining have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

VII. CONCLUSION AND FUTURE WORK

In this paper the different models of data mining used in the prediction of Aortic stenosis are discussed. The focus is on using different models for intelligent and effective Aortic stenosis prediction using data mining. For predicting Aortic stenosis, significantly 11 attributes are listed and with basic data mining technique other approaches e.g. ANN, Time Series, Clustering and soft computing approaches etc. can also be incorporated. The outcome of predictive data mining technique on the same dataset reveals that SVM outperforms but other predictive methods like KNN, Neural Networks, Classification based on clustering are not performing well. The proposed work can be further enhanced and expanded for the automation of Aortic stenosis prediction. Real data from Health care organizations and agencies needs to be collected and all the available techniques will be compared for the optimum accuracy.

REFERENCES

[1] Palaniappan, S.; Awang, R., "Intelligent Heart Disease Prediction System Using Data Mining Techniques". Proceedings of IEEE/ACS International Conference on Computer Systems and Applications 2008, pp. 108-115.
[2] K.Srinivas, B.Kavihta Rani , A.Govrdhan , Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks, (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010, 250-255.
[3] Harleen Kaur , Siri Krishan Wasan and Vasudha Bhatnagar, The impact of data mining techniques on medical diagnostics, Data Science Journal, Volume 5, 19 October 2006 pp119-126.
[4] Avci, E.; Turkoglu, I., "An intelligent diagnosis system based on principle component analysis and ANFIS for the heart valve diseases". Journal of Expert Systems with Application, Vol. 2, No. 1, 2009, pp. 2873-2878.
[5] Asha Rajkumar, G.Sophia Reena, Diagnosis Of Heart Disease Using Datamining Algorithm, Global Journal of Computer Science and Technology 38 Vol. 10 Issue 10 Ver. 1.0 September 2010.
[6] Milan Kumari, Sunila Godara, "Comparitive study of data mining classification methods in cardiovascular disease prediction", International Journal of Computer Science and Technology (IJCST) vol.2, Issue 2, june 2011, pp. 304-308.
[7] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni, "Predictive Data Mining for medical diagnosis: An overview of heart disease prediction", International journal of Computer Applications, vol. 17-no. 8, march 2011.
[8] Chau, M.; Shin, D., "A Comparative Study of Medical Data Classification Methods Based on Decision Tree and Bagging Algorithms". Proceedings of IEEE International Conference on Dependable, Autonomic and Secure Computing 2009, pp. 183-187.
[9] Khemphila, A.; Boonjing, V., "Comparing performance of logistic regression, decision trees and neural networks for classifying heart disease patients". Proceedings of International Conference on Computer Information System and Industrial Management Applications 2010, pp. 193 – 198.
[10] Yao, Z.; Lei, L.; Yin, J., "R-C4.5 Decision tree model and its applications to health care dataset". Proceedings of International Conference on Services Systems and Services Management 2005, pp. 1099-1103.
[11] Magnus Stensmo, Terrence J. Sejnowski Automated Medical Diagnosis based on Decision Theory and Learning from Cases, World Congress on Neural Networks 1996 International Neural Network society pp. 1227-1 231.
[12] Han. J, Kamber, M., "Data Mining Concepts and Techniques". 2nd Edition, Morgan Kaufmann, San Francisco.
[13] N. Aditya Sundar, P.Pushpa Latha, M. Rama Chandra," Performance Analysis of classification data mining techniques over heart disease data base", IJESAT, vol. 2, Issue 3, pp 470-478.
[14] Ruben D. Canlas Jr., Data mining in healthcare: current applications and issues, August 2009.

T. Revathi, is B.Sc (Applied Science –Computer Technology) MCA, M.Phil., working as Associate Professor ,Department of Computer Science in PSG College of Arts & Science, Coimbatore. She is presently pursuing her Ph.D in Manonmaniam Sundaranar University, Tirunelveli. From the year 1997 to 2001 August she was working in Sri Ramakrishna Polytechnic College, Coimbatore. Since 2001 August she has been in PSG College of Arts & Science, Coimbatore. Her area of research is Data Mining.

Dr. P. Sumathi, is MCA, M.Phil, Ph.D, working as Assistant Professor in PG & Research Department, Department of Computer Science, Government Arts College, Coimbatore. She was working as Lecturer in Sengunthar Arts and Science College, Tiruchengode from June 1997 to May 1998. From June 1998 to February 2011 she was working as Assistant Professor & Head, Department of Computer Science, PSG College of Arts & Science, Coimbatore. Worked as Assistant Professor in Chikkanna Government Arts College, Tirupur and presently she is working in Government Arts College, Coimbatore. Her area of research includes Data Mining and Grid Computing.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.