A Survey on Usage of Data Mining Techniques for Prediction of Heart Disease

Ms.Mehdi Khundmir, Prof. Vikas Maral

[1, 2]K.J College of Engineering and Management Research Pune, India

Abstract:In India at every 33 seconds one person is dying because of heart attack and it has become very serious concern. One of the solutions to the problem is usage of "Data Mining". Data mining is field of 21st century, many fields where historical data plays important role data mining is employed to remove dependency on person. However, historical data varies person to person mostly in medical field, data remains with individual medical practitioner and develops his own medical expertise. To overcome this disadvantage data mining has started playing a major role in the field of medical science. Many new researches are collaborating with medical practitioner to find out best, effective and efficient solution. In this article best few papers are summaries to guide upcoming researchers in this field. Author also has done extensive study while writing this article and effective choice of data mining techniques from available techniques is also discussed.

Keywords: Data mining, *Heart* diseases, KNN, naïve bayes, etc

I. Introduction

Cardiovascular system is prepared by using two main parts, heart and blood vessels. Cardiovascular disease (CVD) is defined as any serious, abnormal condition of the heart or blood vessels(arteries, veins). Cardiovascular disease includes coronary heart disease (CHD), stroke, peripheral vascular disease, congenital heart disease, endocarditis, and many other conditions. Many cardiovascular diseases are preventable.Heart diseases also termed as cardiovascular diseases (CVD's) have become most severe and leading problem of mortality in India. In India, early age heart attack has become a major concern amongst medical practitioner the mortality because of CVD is increased from 23 million (1990) to 37 million (2010) and now it is around 50 million (at the end of 2016). The major reasons and causes are shown in Fig. No.1.
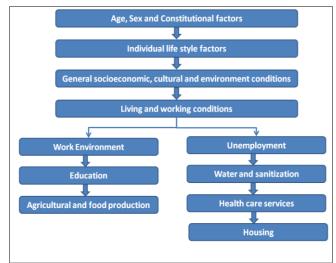

Fig.No.1. Different factors responsible for heart diseases

Data Mining

Data mining is very important techniques is 21st century and inspiring area for the researchers. Data mining is found in late 90's and it appealed to many researchers as it is very effective and powerful tools for finding an previously unknown patterns and vital information from big data set. A motive behind this technique is find an vital information from the large data sets. In current century medical field has gaining lot of importance from the various researchers as it highly unorganized field and many times it is observed that, effectiveness depends on individual medical practitioner. Many researcher assumes that data mining and knowledge discovery in database are same terms and they use it interchangeably, but many practitioners assumes that this are different terms and data mining is one stage for knowledge discovery in database(KDD).
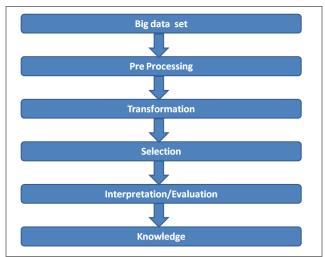
Fig. No.2. Steps for knowledge discovery (Data Mining)

Author	Publication Year	Approaches	Accuracy
Yan et al.	2003	Multilayer Perceptron	63.6%
Andreeva, P	2006	Naïve Bayes	78.56%
		Decision Tree	75.73%
		Neural Network	82.77%
		Kernel Density	84.44%
Hara et al.	2008	Automatically Defined Groups	67.8%
		Immune Multi-agent Neural Network	82.3%
Sitar-Taut et al.	2009	Naïve Bayes	62.03%
		Decision Tree	60.40%
Chang et al.	2009	Decision Tree	90.89%
		Artificial Neural Network	92.62%
		Decision Tree combined with ANN	86.89%
		Decision Tree with sensitivity Analysis	80.33%
		ANN with sensitivity Analysis	83.61%
Rajkumar et al.	2010	Naïve Bayes	52.33%
		Decision tree	52%
		KNN	45.67%
Srinivas et al.	2010	Naïve Bayes	84.14%
		One Dependency Augmented Naïve Bayes classifier	80.46%
Kangwanariyaku l, et al.	2010	Back-Propagation Neural Network	78.43%
		Bayesian Neural Network	78.43%
		Probabilistic Neural Network	70.59%
		Linear Support Vector Machine	74.51%
		Polynomial Support Vector Machine	70.59%
		RBF- kernel Support Vector Machine	60.78%
Anbarasi, et al.	2010	Genetic with Decision tree	99.2%
		Genetic with Naïve Bayes	96.5%
		Genetic with Classification via Clustering	88.3%
Fan et al.	2010	CHAID	69.75%

Table 1: Summary of Data Mining Approaches Used in Healthcare

Application of Data Mining in Health science:
1. Effective management of Hospital resource
2. Hospital Ranking
3. Better patient relation
4. Hospital Infection Control
5. Smarter Treatment Techniques
6. Improved Patient care
7. Decrease Insurance Fraud
8. Recognize High-Risk Patients
9. Health Policy Planning

II. Literature Review:
Ashish Kumar Sen, Shamsher Bahadur Patel, Dr. D. P. Shukla:

Heart diseases is biggest reason for mortality worldwide, this calls the needs of prediction of heart disease at the early stage is became vital. As large amount of data is been gathered in various hospital in medical centers but this is not recorded properly. There is a wealth of hidden information present on individual

expertise. There is so many unused information can be converted useful information. For conversion of unused information to useful information different data mining techniques can be used. While implementing this paper author has defined new and novel approach using two different layers for identification possible diseases. The critical factors due to which heart attack arises are taken at first level and all other factors at second level. Author has stated the developed two level approach increases the performance of our work as it helps in dictation and prediction of heart diseases at early stage. In this paper neural and fuzzy network are used jointly. Neural network training is accomplished by using the data set from the neural network and once neural network is trained fuzzy rules are applied for prediction of the chances of heart diseases. After the algorithm applied it can segregate the heart diseases as low, medium or critical.

Syed Umar Amin, Kavita Agarwal, Rizwan Beg:

Data mining is mostly used technique in clinical research, and widely adopted by the many medical practitioner as it helps in taking decision faster and it also helps in diagnosigning the various diseases with good accuracy. Everything single method used here is for prediction of heart diseases uses a clinical data having large number of parameters and concluded results from various lab experiments. But no fool proof system is found for prediction of heart diseases accurately as it depends of various factors such as age, family history, diabetic, cholesterol, hypertension, smoking, obesity, alcohol intake or it may be also because of the physical inactivity. According to author mentioned above system which monitors above risk factors may or may not help to medical practitioner but in turn it will help to patients a warning of presence of heart diseases before he meets any medical practitioner or visits hospitals to carry out highly sophisticated and costly treatment. To solve this problem author has presented and demonstrated one technical using MATLAB simulink, which will predict heart disease using major risk factors. In this technique two most versatile and accurate technique is been used. Neural network and generic algorithm and used combinely in this paper. The hybrid system will have advantages of genetic as well as neural network and will produce more accurate results. Genetic algorithm is used for producing initial weights in neural network. The learning of the neural network is fast as compared to back propagation algorithm. The author has produced accuracy of 89% by using demonstrated model.

Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni:

In this paper author has summarized different constraints of various methodologies. Different algorithms based on data mining are used for prediction of heart diseases are discussed. In this paper author has focused on different algorithms and various combinations for different target attributes for data mining are discussed in great detail including their accuracy and time for execution of the algorithm. For envisage heart diseases, 15 different attributes were listed are shown in table no. 1 and compared with basic data mining technique e.g. ANN, time series, association rules, clustering and soft computing approaches etc .

PARAMETERS	WEIGHTAGE	
MALE/FEMALE	AGE <30	0.1
	AGE> 30 TO <50	0.3
	AGE>50 AND AGE<70	0.7
	AGE>70	0.8
SMOKING	NEVER	0.1
	PAST	0.3
	CURRENT	0.6
OVERWEIGHT	YES	0.8
	NO	0.1
ALCOHOL INTAKE	NEVER	0.1
	PAST	0.3
	CURRENT	0.6
HIGH SALT DIET	YES	0.9
	NO	0.1
HIGH SATURATED FAT DIET	YES	0.9
	NO	0.1
EXERCISE	NEVER	0.6
	REGULAR	0.1
	HIGH IF AGE>30	0.1
	HIGH IF AGE >50	0.6
SEDENATRAY LIFESTYLE	YES	0.7
	NO	0.1
HEREDITARY	YES	0.7
	NO	0.1
BAD CHOLESTROL	VERY HIGH>200	0.9
	HIGH 160 TO 200	0.8
	LOW <160	0.1
BLOOD PRESSURE	NORMAL (130/89)	0.1
	LOW (<119/79)	0.8
	HIGH (>200/160)	0.9
BLOOD SUGAR	HIGH (>120 & <400)	0.5
	NORMAL (>90 & <120)	0.1
	LOW (<90)	0.4
HEART RATE	LOW (<60BPM)	0.9
	NORMAL (60 TO 100 BPM)	0.1
	HIGH (>100 BPM)	0.9

Table No.1. Heart attack parameters with corresponding values and their weightages

ALGORITHM USED	ACCURACY	TIME TAKEN
NAÏVE BAYES	52.33%	609ms
DECISION LIST	52%	719ms
KNN	45.67%	1000ms

Table No.2 Performance study of data mining approaches.

The outcome of the predictive technique used for data mining shows that decision tree outperforms Bayesian classification possesses similar accuracy when compared to decision tree but other predictive methods like KNN, NN and classification based on clustering are giving good result. Author has also concluded that the accuracy of the Bayesian classification and decision tree improves further after application of genetic algorithm for reduction of actual data size for getting an optimal subset of attribute.

K.Srinivas, B.Kavihta Rani, Dr. A.Govrdhan:

In this paper author has studies the problem constraining and summarizing of various algorithms relating to the data mining. Author has made focused on the several attributes for prediction of combinations of different attributes. Finally author has developed on e algorithms for prediction of heart attack using data mining technique. In first stage, very effective approach for extraction of significant patterns from the heart care institutes is been collected, this helps in efficient prediction of heart attack. For this purpose significant weight is been assigned depending upon the severity of the heart attack. Assigning the significant weightage is based on the frequent patterns and predefined value of predicted heart attack. Five mining goals are predefined which is solely based on the intelligence and data exploration. The goal extraction is solely based on the evaluation of the trained models. The trained model defined in this paper can readily answer in complex queries like predicting heart attack.

Sellappan Palaniappan, Rafiah Awang:

The healthcare industry possesses the large amount of data, but the bad thing about of this are not mined for discory of hidden pattern and information which can be very useful for the effective decision making. The other advantage of using data mining technique is, it will help fresher's medical practitioner for decision making. There are various data mining technique which can be used for development prototype intelligent heart disease prediction system (IHDPS) uby using various advanced data mining techniques, e.g. Naïve Bayes, Neural Network and decision trees. The results obtained in this paper shows that each technique mentioned above has unique different advantage, so author has developed hybrid approach to achieve defined data mining goals and named it as IHDPS. Author has collected different medical profiles for heart care institute and then segregated data in different groups based on, Medical profiles, age, sex, blood pressure and blood sugar which will helps in predicting the detection of heart disease. Developed IHDPS is web-based, scalable, reliable user friendly and expandable and author has implemented this using .NET platform.

III. Concluson

In India many people's are dying because of heart diseases, there are various reason for that e.g. unhealthy eating, high salt eating, diabetic ratio and blood pressure is also high in India. Hence it calls for algorithm which will predict heart attack in early stage. While going through literature review its been observed that Naïve Bayes will produce more accurate result as compared to other methods. But based on certain observations and history, it's better to combine two different methods and develop hybrid approach for increasing effectiveness and accuracy.

References

[1]. Ashish Kumar Sen, Shamsher Bahadur Patel, Dr. D. P. Shukla, "A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated" Approach Two Level", International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 2 Issue 9 Sept., 2013 Page No. 2663-2671.
[2]. Syed Umar Amin, Kavita Agarwal, Rizwan Beg,, " Genetic neural network based data mining in prediction of heart disease using risk factors", Information and communication technologies (ICT), 2013 conference on information and communication technologies.
[3]. 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 (0975 – 8887), Volume 17– No.8, March 2011.
[4]. Divya Tomar and Sonali Agarwal, " A survey on Data Mining approaches for Healthcare", International Journal of Bio-Science and Bio-Technology, Vol.5, No.5 (2013), pp. 241-266.
[5]. Asha Rajkumar and B. Sophia Reena, "Diagnosis Of Heart Disease Using Data mining Algorithm" , Global Journal of Computer Science and Technology, Vol. 10, No. 10, pp. 38 - 43, 2010
[6]. Latha Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic ISSN: 2229-6956(ONLINE) ICTACT JOURNAL ON SOFT COMPUTING, JULY 2013, VOLUME: 03, ISSUE: 04
[7]. M. Anbarasi, E. Anupriya and N.CH.S.N. Iyenga, "Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm", International Journal of Engineering Science and Technology, Vol. 2, No. 10, pp. 5370 - 5376, 2010

[8].   Chaitrali S. Dangare and Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications, Vol. 47, No. 10, pp. 0975 – 888, 2012

[9].   D. Shanthi, G. Sahoo and Dr. N. Saravanan, "Designing an Artificial Neural Network Model for the Prediction of Thrombo-embolic Stroke", International Journal of Biometric and Bioinformatics, Vol. 3, No. 1, pp. 250 - 255, 2008.

[10]. P. K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules", Journal of King Saud University Computer and Information Sciences, Vol. 11, pp. 309 - 314, 2011.

[11]. J. C. Obi and A. A. Imainvan, "Decision Support System for the Intelligent Identification of Alzheimer using Neuro Fuzzy logic", International Journal on Soft Computing , Vol. 2, No. 2, pp. 25 - 38, 2011.

[12]. D. Shanthi, G. Sahoo and Dr. N. Saravanan, "Evolving Connection Weights of Artificial Neural Network Using Genetic Algorithm With Application to the Prediction Stroke Diseases", International Journal of Soft Computing, Vol. 2, pp. 95 - 101, 2009.

[13]. A. Sudha, P. Gayathiri and N. Jaisankar, "Effective Analysis and Predictive Model of Stroke Disease using Classification Methods", International Journal of Computer Applications, Vol. 43, No. 14, pp. 0975 – 8887, 2012.

[14]. Mai Shouman, Tim Turner and Rob Stocker, "Integrating Decision Tree and K-Means Clustering with Different Initial Centroid Selection Methods in the Diagnosis of Heart Disease Patients", Proceedings of the International Conference on Data Mining, 2012.

[15]. Olatubosun Olabode and Bola Titilayo Olabode, "Cerebrovascular Accident Attack Classification Using Multilayer Feed Forward Artificial Neural Network with Back Propagation Error", Journal of Computer Science, Vol. 8, No. 1, pp.18 - 25, 2012.

[16]. K.Srinivas, B.Kavihta Rani, Dr. A.Govrdhan, "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks", International Journal on Computer Science and Engineering, Vol. 02, No. 02, 2010, 250-255

[17]. Shellappan, Palaniappan, Rafian Awang, "intellingent heart disease prediction system using data mining techniques", 2008 IEEE/ACS international conference on computer systems and applications.