INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 2)
Available online at: www.ijariit.com

Heart Disease Prediction System using Hybrid Technique of Data Mining Algorithms

Navdeep Singh
er.deep@hotmail.com
Shaheed Bhagat Singh State Technical Campus, Firozpur, Punjab

Sonika Jindal
sonikamanoj@gmail.com
Shaheed Bhagat Singh State Technical Campus, Firozpur, Punjab

ABSTRACT

Most countries confront high and growing rates of heart illnesses or Cardiovascular Disease. Regardless of the way that, best in class pharmaceutical is making the giant measure of data reliably, little has been done to use this open data to understand the challenges that face a viable illustration of echocardiography examination comes about. To design a perceptive model for heart illnesses acknowledgment using data mining strategies that are fit for enhancing the constancy of heart infections conclusion. Learning Discovery in Database strategy including nine iterative and instinctive advances was grasped to think basic cases from a dataset containing a couple of echocardiography examination reports of heart patients over the globe. Thereafter, we divide this data into Training and Testing Data Sets and employ SVM technique to obtain relatively higher prediction accuracy. The primary goal of this research paper is to devise out a model that gives a highly accurate prediction of Heart Disease. As we have done a combination of Genetic and Naïve Bayes Technique, the Investigation developed a Hybrid model of both these techniques and called it Hybrid Genetic Naïve Bayes Model for predicting high accuracy in results.

Keywords: *Heart Disease, Data Mining, Classification, Linear SVM, GA, Python.*

1. INTRODUCTION

Because of a wide accessibility of superlative measure of information and a need to change over this accessible huge measure of information to helpful data requires the utilization of information mining strategies. Information Mining and KDD (learning disclosure in the database) have turned out to be prominent as of late. The popularity of information mining and KDD (information revelation in database) shouldn't be an amazement since the measure of the information increases that are accessible are extremely extensive to be analyzed physically and even the techniques for programmed information investigation in view of established insights and machine adapting frequently threaten issues when preparing large, dynamic information increases comprising of complex items [11].

Information Mining is the centerpiece of Knowledge Discovery Database (KDD). Numerous individuals regard Data Mining as an equivalent word for KDD since it's a key piece of KDD process. There are sure stages of information mining that you will need to get comfortable with, and these are exploration, pattern identification, and deployment. Information mining is an iterative procedure that commonly includes the accompanying stage [17].

1.1 HEART DISEASE

A key challenge confronting healthcare organizations (hospitals, medical centers) is the facility of quality services at reasonable prices. Quality amenities suggest diagnosing patients accurately and regulating medications that are effective. Poor clinical choices can prompt deplorable results, which are in this manner unsatisfactory. Hospitals should limit the cost of clinical tests. They can accomplish these outcomes by utilizing fitting PC based data and additionally choice emotionally supportive networks [4][6].

The heart is the essential piece of our body. Life is itself reliant on effective working of the heart. In the event that task of the heart isn't legitimate, it will influence the other body parts of human, for example, cerebrum, kidney and so on. Coronary illness is a sickness that effects on the activity of the heart. There is a number of elements which builds danger of Heart ailment [13][17].

Some of them are listed below:
- The family history of heart disease
- Smoking

- Cholesterol
- High blood pressure
- Obesity
- Lack of physical exercise

2. LITERATURE SURVEY

Heart disease could be a term that assigns to an oversized range of medical conditions related to the heart. These medical conditions describe the abnormal health conditions that directly influence the heart and all its elements. A heart condition could be a major ill health in today's time. The foremost studies done by victimization neural networks with fifteen attributes has outperformed over all different data processing techniques.

Kiran Jyoti(October - 2012) [24] used the artificial neural network (ANN), usually referred to as a "neural network" (NN), is a mathematical model or machine model supported biological neural network. In alternative words, it's an emulation of the biological neural system. Cardiovascular disease prediction system has been developed using fifteen attributes. Earlier thirteen attributes were used for prediction, however, this analysis work incorporated a pair of additional attributes, i.e. fatness and smoking for economical designation of a heart condition. The data mining tool weka 3.6.6 is employed for the experiment. Initially, missing values were known within the dataset and that they were replaced with acceptable values using Replace Missing Values filter from 3.6.6. Further, numerous data processing techniques are analyzed on heart disease information. Confusion matrix is obtained for every classifier.

Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin(2012) [22] presented prototype model for the breast cancer in addition to cardiovascular disease prediction using data processing techniques. Two decision tree algorithms C4.5 and also the C5.0 is used on these datasets for prediction and performance of each algorithm are compared. Pruning algorithmic rule is used to scale back an error and avoiding the overfitting. Pruning a tree is the action to interchange an entire subtree by a leaf. The replacement takes place if the expected error rate within the subtree is greater than in the single leaf. During this study, they started by generating the entire (generally over fitted) classification tree and change it using pruning simply once.

MA.JABBAR, Dr. PRITI CHANDRA (October 2011) [14] implemented CBARBSN Cluster based Association Rule Mining supported Sequence number during which they projected a new rule which combines the construct of sequence numbers and cluster. .The entire information base is split into partitions of equal size, every partition is referred to as a cluster. Every cluster is taken into account one at a time by loading the primary cluster into memory and hard frequent item sets. Then the second cluster is taken into account equally and hard frequent item sets. This approach reduces main memory demand since it considers solely a little cluster at a time and it's scalable and efficient.

Ms. Ishtake S.H (April 2013) [32] was implemented a model heart disease prediction system is developed using 3 data processing classification modeling techniques specifically, Decision. Trees, Naïve Bayes and Neural Network The system extract hidden information from a historical heart disease information. DMX command language and functions are accustomed build and access the models. five mining goals are defined supported business intelligence and information exploration. The goals are evaluated against the trained models. All three models might answer complicated queries, each with its own strength with reference to ease of model interpretation, access to elaborate data and accuracy.

Dr. K. Usha Rani (September 2011) [18] used Neural Network approach for the analysis of cardiovascular disease. Neural Networks have emerged as a crucial tool for classification.The advantages of Neural Networks helps for efficient classification of given information. to extend the potency of the classification method parallel approach is additionally adopted within the training part. The experiment is conducted with cardiovascular disease dataset by considering the one and multilayer neural network modes. Back propagation algorithmic program with momentum and variable learning rate is employed to train the networks. The experimental results verified that neural networks technique provides satisfactory results for the classification task.

Shantakumar B.Patil (February 2009) [10] presented an efficient approach for extracting vital patterns from the heart sickness data warehouses for the efficient prediction of attack. The preprocessed cardiovascular disease information warehouse was clustered to extract information most relevant to attack mistreatment K-means clump algorithmic rule. The frequent things are mined effectively mistreatment MAFIA algorithmic rule. based on the calculated important weightage, the frequent patterns having worth larger than a predefined threshold were chosen for the precious prediction of attack. In our future work, we've planned to style and develop an economical attack prediction system with the help of those chosen vital patterns mistreatment artificial intelligence techniques.

Dilip Roy Chowdhury (2011) [16] represents the utilization of artificial neural networks in predicting neonatal illness identification. The projected technique involves training a Multi-Layer Perceptron with a BP learning algorithmic program to acknowledge a pattern for the designation and prediction of neonatal diseases. the back propagation algorithmic program was accustomed train the ANN design and also the same has been tested for the various classes of neonatal illness. Concerning ninety-four cases of various sign and symptoms, parameter is tested in this model. This study exhibits ANN based prediction of neonatal illness and improves the identification accuracy of seventy-five with higher stability.

Milan Kumari (June 2011) [17] proposed study covers information processing classification techniques such as ripper classifier, decision Tree, Artificial neural networks (ANNs), and Support Vector Machine (SVM) are analyzed on heart disease dataset. Performance of those techniques is compared through sensitivity, specificity, accuracy, error rate, True Positive Rate, and False Positive Rate. 10-fold cross validation methodology was used to measure the unbiased estimate of those prediction models. As per

our results error rates for ripper, decision Tree, ANN, and SVM are a pair of.756, 0.2755, 0.2248 and 0.1588 respectively. The accuracy of the ripper, decision Tree, ANN, and SVM are eighty-one .08%, 79.05%, 80.06% and 84.12% respectively. The analysis shows that out of those four classification models SVM predicts heart disease with least error rate and highest accuracy.

Niti Guru (January-June 2007) [4] proposed a system that uses a neural network for prediction of cardiovascular disease, blood pressure, and sugar. a collection of seventy-eight records with thirteen attributes are used for training and testing. He urged supervised network for diagnosing of cardiovascular disease and trained it using back propagation formula. On the idea of unknown information is entered by a doctor the system can notice that unknown information from training data and generate a list of possible illness from that patient will suffer.

3. METHODOLOGY

For the implementation of the project, we used different algorithms i.e. Naïve Bayes and genetic algorithm. The platform used is Python 3.6.

Data set

We take the data set in this paper with 303 records and 14 attributes collected from the online dataset repository of archive.ics.edu/ml/datasets. The dataset parameters are listed in Table 3.1

Table 3.1: The dataset attributes

Sr. No.	Attributes Name	Description
1	(age)	Patient Age
2	(sex)	Male/Female
3	(cp)	Chest pain type
4	(trestbps)	Resting blood pressure (in mm Hg on admission to the hospital)
5	(chol)	Serum cholestoral (mg/dl)
6	(fbs)	Fasting blood sugar
7	(restecg)	Resting ECG results
8	(thalach)	Maximum heart rate achieved
9	(expand)	Exercise induced angina
10	(old peak)	ST depression included by exercise relative to rest
11	(slope)	The slope of the peak exercise ST segment
12	(ca)	Number of major vessels (0-3) colored by flouroscopy
13	(thal)	3 = normal, 6 = fixed defect, 7 = reversible defect.
14	(num)	Angiographic disease status (Diagnosis of heart disease)

Algorithm 1: GA for Feature Selection

STEP 1: Input dataset

STEP 2: Authenticate the data in respect to make in the appropriate format

STEP 3: Algorithm discovers the best solution for given values and returns the personal best and global best

$$m_H(i+1) = F_H(i)m_H(i)\left[1 - pc\frac{l_H}{l-1}\right][(1 - P_m)^H] \quad (1)$$

STEP 4: Calculate the pBest and gBest on each component of the dataset for every block

STEP 5: Apply fitness function using Z statistics Z=S(X)-Minimum support/SQRT (min sup*(1-minsup))/N Where S(X) is provision of pattern and min sup is user defined threshold

$$f(x) = \sum_{i=1}^{n-1}[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2] \quad (2)$$

STEP 6: Cut the features based on Z statistics. After evaluation the rules having maximum fitness are filtered.

FUZZY RULES

These are the rules that categorize a set of values into user defined output following some Inference Mechanisms. The Inference mechanisms follow such a way, which helps in laying and processing the Fuzzy Rules. While designing the rules, every parameter value must be mapped to an output (should not be an empty set) in order to generate a useful output from a set of values given in

the Input. Inference Mechanisms Inference mechanism uses the principle of fuzzy logic for synchronizing an input to output using fuzzy logic. It uses membership functions, logical operations and if-then rules. Madani method This method works like below:
Consider:
P1: if m is S1 and n is T1 then q is U1
P2: if m is S2 and n is T2 then q is U2
Result: q is U, where m equals m0 and n equals n0. Sugeno method
This method works like below:
Consider:
P1: if m is S1 and n is T1 then q is q1 = a1m1+b1n1
P2: if m is S2 and n is T2 then q is q2 = a2m2+b2n2
Result: q0, where m equals m0 and n equals n0.

Algorithm2: Cardiovascular Disease Classification Using Naive Bayes

Compute identification="yes", identification="no" Pyes, Pno from training input.

$p(C_k|\ x_1, \ldots\ldots, x_n)$ (1)

For Every Input Record

Determine Category of Attribute

$p(C_k|x) = \frac{p(C_k)\ p(x|C_k)}{p(x)}$ (2)

Compute Probabilities Of Identification="Yes", Identification="No"

$p(C_k, x_1, \ldots\ldots, x_n)$ (3)

For Each Feature

Compute Ryes= Ryes* P(Attr,Yes),Rno= Rno*P(Attr,No); $p(C_k|x_1, \ldots\ldots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^{n} p(x_i|C_k)$ (4)

If(Ryes > Rno) Then Identify="Yes";

Else Identify ="No";

4. RESULTS/COMPARATIVE ANALYSIS

As the goal of this study is to observe heart condition using data processing techniques a classification technique was adopted to develop a predictive model. The models were designed with two totally different supervised machine learning algorithms i.e. genetic rule and Naive Bayes. Parameters results have given below. The results are obtained in the type of the various performance parameters as precision, recall and accuracy percentage shown in Table 1. Figure 1 displays the graphical representation of results.

Table 1: Parameters Results of Proposed Model

Parameter	Accuracy	Precision	Recall
Value	97.14%	98%	97.14%

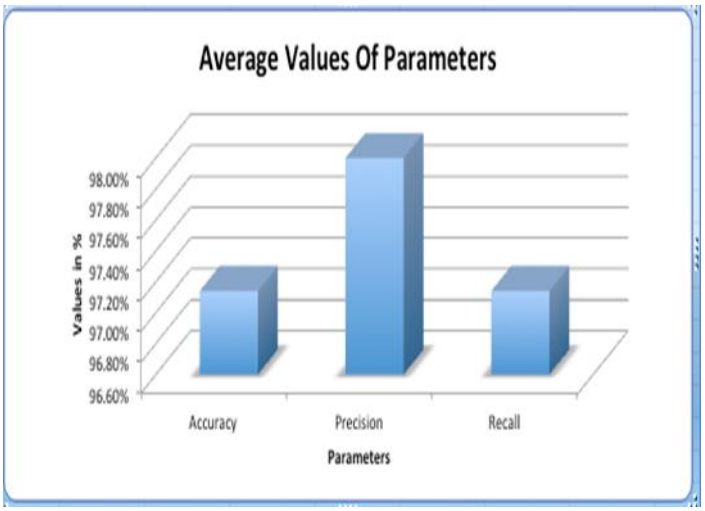

Figure 1: Graph of Proposed Model Results

Table 2: Comparison of Accuracy with Existing Model

Sr. No.	Model	Accuracy
1	Weighted Fuzzy Rules	57.85%
2	Logistics Regression	77%
3	Naïve Bayes	81.48%
4	Attribute Weighted Artificial Immune System	82.59%
5	Artificial Immune System	84.5%
6	Modified Artificial Immune System	87.43%
7	Neural Networks ensemble	89.01%
8	ANN_Fuzzy_HP	91.10%
9	GA_Fuzzy_Naive (Proposed)	97.14%

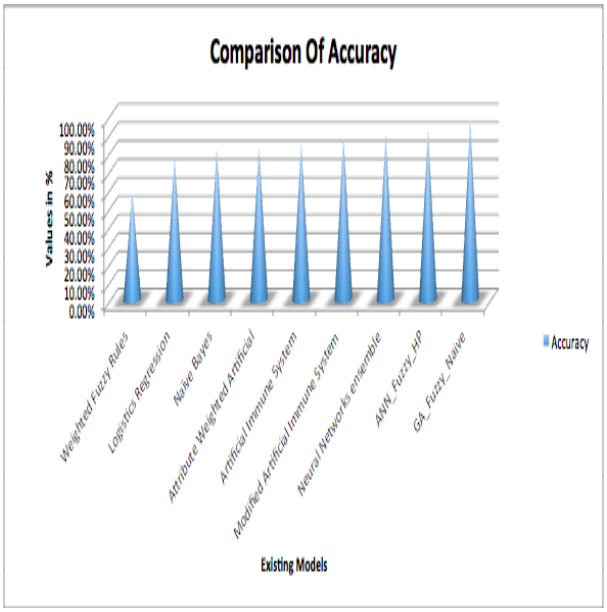

Figure 2: Comparison of Accuracy with Existing Models

5. CONCLUSION

In this study, our aim was to design a heart disease prediction system using various data mining techniques and to perform the analysis of the results obtained for all implemented techniques. So for the completion of the heart disease prediction model survey, we have evaluated the popular and effective heart disease prediction methods from the literature survey and finally select the most effective algorithms of Naïve Bayes and Genetic Algorithm for their performance analysis on the heart disease prediction. The performances of the models were evaluated using the genetic algorithm and naïve Bayes. In this paper, total 14 attributes are used to get more accurate results. From results, it's been seen that projected model provides correct results as compare to existing models. This technique may be additionally expanded. Alternative data processing techniques also can be used for predication e.g. clustering, statistic, and Association rules. The text mining may be used to mine an immense quantity of unstructured information out there in healthcare business info.

6. FUTURE SCOPE

As future work, considering the more attributes as input data can expand this study. Furthermore, the work could be done on early detection of heart disease by processing family's historical data.

7. REFERENCES

[1] Nada Lavrac, Selected techniques for data mining in medicine, Elsevier, 1999.
[2] Carlos Ordonez, Edward Omiecinski, Mining Constrained Association Rules to Predict Heart Disease, IEEE. Published in International Conference on Data Mining (ICDM), p. 433-440, 2001.
[3] Marcel A.J. van Gerven, Predicting carcinoid heart disease with the noisy-threshold classifier, Elsevier, 2007.
[4] Niti Guru, Anil Dahiya, Navin Rajpal, Decision Support System for Heart Disease Diagnosis Using Neural Network, Delhi Business Review, Vol. 8, No. 1, January-June 2007
[5] Humar Kahramanli, Novruz Allahverdi, Design of a hybrid system for diabetes and heart diseases, Elsevier, 2008.

[6] Latha Parthiban and R.Subramanian, Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm, International Journal of Biological and Medical Sciences, 2008.
[7] Tanawut Tantimongcolwat, Thanakorn Naenna, Identification of ischemic heart disease via machine learning analysis on Magnetocardiograms, Elsevier, 2008.
[8] Resul Das, Ibrahim Turkoglu, Abdulkadir Sengur, Effective diagnosis of heart disease through neural networks ensembles, Elsevier, 2009.
[9] Resul Das, Ibrahim Turkoglu, Abdulkadir Sengur Diagnosis of valvular heart disease through neural networks ensembles, Elsevier, 2009.
[10] Shantakumar B.Patil, Y.S.Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network‖"; European Journal of Scientific Research, ISSN 1450-216X Vol.31 No.4, 2009.
[11] Rupa G. Mehta, Dipti P. Rana, Mukesh A. Zaveri, ―A Novel Fuzzy Based Classification for Data Mining using Fuzzy Discretization‖; World Congress on Computer Science and Information Engineering, 2009.
[12] M. Anbarasi, E. Anupriya, N.ch.s.n.Iyengar, Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm, International Journal of Engineering Science and Technology,2010.
[13] Asha Rajkumar, G. Sophia Reena, Diagnosis of Heart Disease Using Datamining Algorithm‖; Global Journal of Computer Science and Technology, Page 38 Vol. 10 Issue 10 Ver. 1.0 September, 2010.
[14] Ma.jabbar, Dr.prirti Chandra, B.L. Deekshatulu, cluster based association rule mining for heart attack prediction, Journal of Theoretical and Applied Information Technology, 2011.
[15] Jyoti Soni, Sunita Soni et al. "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
[16] Dilip Roy Chowdhury, Mridula Chatterjee & R. K. Samanta, An Artificial Neural Network Model for Neonatal Disease Diagnosis, International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume (2): Issue (3), 2011.
[17] Milan Kumari, Sunila Godara, Comparative Study of Data Mining Classification Methods in Cardiovascular Disease Prediction, IJCST Vol. 2, Iss ue 2, June 2011
[18] Dr. K. Usha Rani,analysis of heart diseases dataset using neural network approach,International Journal of Data Mining & Knowledge Management Process, 2011.
[19] Venkatadri.M, Dr. Lokanatha C. Reddy a review on data mining from past to the future. International Journal of Computer Applications, 2011.
[20] Nan-Chen Hsieh & Lun-Ping Hung & Chun-Che Shih, Intelligent Postoperative Morbidity Prediction of Heart Disease Using Artificial Intelligence Techniques, J Med Syst, 2012.
[21] Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, A data mining approach for prediction of heart disease using neural networks, international journal of computer engineering and technology, 2012.
[22] Mohammad Taha Khan, Dr. Shamimul Qamar and Laurent F. Massin, A Prototype of Cancer/Heart Disease Prediction Model Using Data Mining, International Journal of Applied Engineering Research, 2012.
[23] N. Aditya Sundar, P. Pushpa Latha, M. Rama Chandra,performance analysis of classification data mining techniques over heart diseases data base, international journal of engineering science and advanced technology, 2012.
[24] Nidhi Bhatla Kiran Jyoti, An Analysis of Heart Disease Prediction using Different Data Mining Techniques, International Journal of Engineering Research & Technology (IJERT), 2012.
[25] T. John Peter, K. Somasundaram, study and development of novel feature selection framwork for Heart disease prediction, International Journal of Scientific and Research Publications, 2012.
[26] Shadab Adam Pattekari and Asma Parveen, prediction system for heart disease using naïve bayes, International Journal of Advanced Computer and Mathematical Sciences, 2012.
[27] Chaitrali S. Dangare,Sulabha S. Apte, Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques, International Journal of Computer Applications, 2012.
[28] Oleg Yu. Atkov, Coronary heart disease diagnosis by artificial neural networks including genetic polymorphisms and clinical parameters, Elsevier, 2012.
[29] M.Akhil jabbar, Dr.Priti Chandra, Dr.B.L Deekshatulu, Heart Disease Prediction System using Associative Classification and Genetic Algorithm, International Conference on Emerging Trends in Electrical, Electronics and Communication Technologies, 2012.
[30] Jesmin Nahar, Tasadduq Imama, Kevin S. Tickle, Yi-Ping Phoebe Chen, Association rule mining to detect factors which contribute to heart disease in males and females, Elsevier, 2013.
[31] Jesmin Nahar, Tasadduq Imam, Computational intelligence for heart disease diagnosis: A medical knowledge driven approach, Elsevier, 2013.
[32] Ishtake S.H ,Prof. Sanap S.A., "Intelligent Heart Disease Prediction System Using Data Mining Techniques", International J. of Healthcare & Biomedical Research,2013.
[33] Indira S. Fal Dessai, Intelligent Heart Disease Prediction System Using Probabilistic Neural Network, International Journal on Advanced Computer Theory and Engineering, 2013.
[34] Rashedur M. Rahman, Farhana Afroz, Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis, Journal of Software Engineering and Applications, 2013.
[35] Adebayo Peter Idowu, Data Mining Techniques for Predicting Immunize-able Diseases: Nigeria as a Case Study, International Journal of Applied Information Systems, 2013.
[36] Abhishek taneja, Heart Disease Prediction System Using Data Mining Techniques, Oriental Scientific Publishing Co., India, 2013.
[37] Monali Dey, Siddharth Swarup Rautaray, Study and Analysis of Data mining Algorithms for Healthcare Decision Support System, International Journal of Computer Science and Information Technologies, 2014.

