ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 12, December 2019

Enhancing the Performance of Locality Sensitive Hashing in Predicting Heart Disease

Mr. Michael Raj [1], Dr.K.B.Manikandan[2]

Research Scholar, Department of Computer Science, Providence College for Women, Coonoor, The Nilgiris, Tamilnadu, india[1].

Assistant Professor, Department of Computer Science, Providence College for Women, Coonoor, The Nilgiris, Tamilnadu, india[2].

ABSTRACT: Clinical decision making is a frequent and a ubiquitous task by every physician in their daily medical practice. Most of them use their experience and knowledge, sometimes from the history of patients, research, literature etc., for predicting or diagnosing the diseases. However, with the abundance of health data and the complex characteristics of diseases, clinical prediction approaches should be more effective to address these challenges. In this paper, the supervised machine learning concept is utilized for making the predictions of heart disease, using the classification model, Locality Sensitive Hashing. This paper emphasis the prediction of heart disease accuracy using the LSH technique. Experiments are conducted for the proposed work and the final output shows that the proposed work outperforms well compare to the Decision tree, Neural Network and Naïve Bayes algorithms.

KEY WORDS: Data mining, Heart Disease, Classification, LSH.

I. INTRODUCTION

Data Mining is an interesting field of research whose significant goal is to discover interesting and useful patterns from large volume of data sets. The change in life style and hereditary reasons, nowadays, health diseases are increasing drastically. Especially, a heart disease has become widespread recently. About 17.3 million deaths per year are due to heart disease and it is ranked high as the major cause of death in the world. Based on a person's age, blood pressure, cholesterol, pulse, gender etc., heart disease can be predicted using data mining classification technique.

In this research work, the supervised machine learning concept is utilized for making the predictions. A comparative analysis of the three data mining classification algorithms namely neural network, Decision tree, Naïve Bayes and LSH are used to make predictions. The analysis is done at several levels of Model building time, correctly classified instances, incorrectly classified instance and accuracy %. The Cleaveland dataset from UCI machine learning repository is utilized for making heart disease predictions in this research work. The predictions are made using the classification model, Locality Sensitive Hashing that is built from the classification algorithms when the heart disease dataset is used for training. This final model can be used for prediction of any types of heart diseases and is compared with the following algorithms.

II. LITERATURE SURVEY

The main goal of this work is to predict the silent heart attacks by identifying the key patterns and features collected from the medical data. For this, classification algorithms are used to select the proper attributes. Some of the popular data mining techniques in predicting heart diseases proposed by the researchers are Naïve Bayes, Neural network, Decision Tree etc.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 12, December 2019

Some of the proposed classification algorithms by various authors are summarized in the following table 1.1

Sl.No	Author	Method	Merits	Demerits
1	Dangare et al (2012) [1]	Multilayer Perception Neural Network (MLPNN) with Back Propagation (BP) Algorithm	• Easily Handles Missing values • Multiple tasks can be executed in parallel.	Identified Obesity and smoking are the cause for Heart disease.
2	Jabbar et al. (2013) [2]	Associative classification algorithm using genetic approach	• Rules are highly comprehensible	• Includes more feature • Execution takes much time
3	Krishnaiah (2014) [3]	Fuzzy Logic	• Used uncertainty of unstructured data	Not suitable for all datasets
4	Kumar (2013)[4]	Fuzzy Logic	• Calculations based on membership functions of the fuzzy variable	Handles only 6 attributes
5	Masethe HD (2014) [5]	Decision Tree and compared he result with J48, Naive Bayes, Bayes Net, EPTREE algorithm and Simple Cart, and R	• Used reliable parameters for each method	Variations in parameters leads to complexity in problem definition
6	Cemil et al.(2015) [6]	Heart stroke prediction model using ANN and SVM	• Handle Multiclass functionalities	Complex implementation More execution time.

III. PROPOSED METHODOLOGY

One of the major drawbacks of the works specified in the literature review is that the main focus was on applying heart disease prediction classification techniques rather than researching various techniques of data cleaning and pruning to plan and render a dataset appropriate for mining. A correctly cleaned and pruned dataset has been found to provide much better accuracy than an unclean data set with missing values. Selection of suitable data cleaning techniques along with correct classification algorithms can lead to the development of prediction systems that provide enhanced accuracy.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 12, December 2019

3.1 FEATURE SELECTION

The Cleveland Dataset contains 76 attributes. Wrapper Subset Eval method along with Greedy stepwise search method is used for attributes selection and selected the following attributes.

Predictable attribute

1. Diagnosis (value 0: < 50% if the diameter is narrowing means no heart disease);
If the diameters' valueis 1: > 50% diameter is narrowing (has heart disease))

Key attribute
1. PatientID – Patient's identification number
Input attributes
1. Sex (value 1: Male; value 0 : Female)
2. Chest Pain Type (value 1: typical type 1 angina, value 2: typical type angina, value 3: non angina pain; value 4: asymptomatic)
3. Fasting Blood Sugar (value 1: > 120 mg/dl; value 0: < 120 mg/dl)
4. Restecg – resting electrographic results (value 0: normal; value 1: 1 having ST-T wave
abnormality; value 2: showing probable or definite left ventricular hypertrophy)
5. Exang – exercise induced angina (value 1: yes; value 0: no)
6. Slope – the slope of the peak exercise ST segment (value 1: unsloping; value 2: flat; value 3:
down sloping)
7. CA – number of major vessels colored by floursopy (value 0 – 3)
8. Thal (value 3: normal; value 6: fixed defect; value 7: reversible defect)
9. Trest Blood Pressure (mm Hg on admission to the hospital)
10. Serum Cholesterol (mg/dl)
11. Thalach – maximum heart rate achieved
12. Oldpeak – ST depression induced by exercise relative to rest
13. Age in Year

IV. EXPERIMENTAL RESULTS

In our work, we used Locality Sensitive Hashing classifier for predicting Heart disease. The main goal of this paper is to compare our results with different classification model. For that, we have compared our result obtained from LSH with Decision Tree, Neural Network, and Naïve Bayes.

4.1 EVALUATION OF MINING GOALS

Three mining objectives are calculated based on the data and goals of the cardiovascular study. The trained models were evaluated. The results show that all three models have achieved the stated objectives, which imply that they can be used to help decision making for clinical assessment and the identification of cardiovascular factors.The goals are explained below.

Goal 1: Identify the clinical impact and relation of the predictable disorder — cardiac disease

The proposed system is most likely (99.08%) to indicate in the relationship between attributes that patients with heart disease are present:' Shest Pain Type=4 and CA=0 and Exang=0, and Trest Blood Pressure >= 142 and <159.' Experiments are carried out to evaluate the usefulness and the performance of different classification algorithm for predicting heart disease. Out of 909 records, 416 records contain patients with heart disease whereas 493 records contain patients with no heart disease.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 12, December 2019

Table 4.1 reveals the performance analysis of the LSH compared with the Decision tree, Neural Networks and Naïve Bayes models in terms of Model building time, correctly classified instances, incorrectly classified instances and Accuracy %.

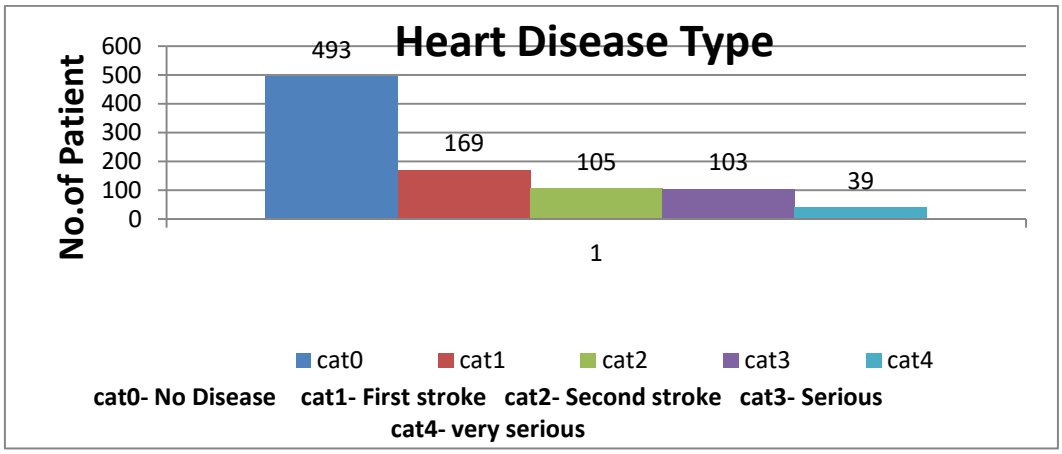

Fig. 4.1 Heart Disease - Category

V. CONCLUSION AND FUTURE WORK

The results are mainly discussed on three goals, predict patients likely to be diagnosed with heart disease, despite their health profiles, identify the important influences and connections in the predictable state of medical input and identify the clinical impact and relation of the predictable disorder. The final output shows that the proposed algorithm outperforms in terms of accuracy and execution time.

This study can be improvised by improving in terms of feature reduction and using optimization techniques. The ultimate focus is on future research directions for customized predictive models. Although the results shown in this paper have demonstrated some approaches to improving the quality of predictive models, much work is still needed. Some of the possible research to be investigated in the future include: Biological spectrum study of a more comprehensive set of biomarkers (e.g. proteomic and genomic data). This would allow highly relevant and predictive biomarkers to be identified, which could better predict the progression of a disease.

REFERENCES

[1] DangareA.Data mining approach for prediction of heart disease using neural network .IJCET 2012; 3: 30-40.
[2] Jabbar MA, DeekshatuluBL,Priti C. Prediction of risk score for heart disease using associative classification and hybrid feature Selection. IEEE ISDA 2012; 628-634.
[3] Krishnaiah V. Diagnosis of heart disease patients using fuzzy classification techniques. ICCCT 2014; 1-7.
[4] Kumar. Detection of heart disease using fuzzy logic.IJETT 2013; 4.
[5] Masethe HD, Masathe MA. Prediction of heart disease using classification algorithm.Wcess 2014.
[6] Cemil. Colak, E. Karaman, and M. G. Turtay, "Application of knowledge discovery process on the prediction of stroke," Comput. Methods Programs Biomed., vol. 119, no. 3, pp. 181–185, 2015.