

A Review On: Heart Disease Prediction using Data Mining

| | | |
|----------------------------------|----------------------------------|----------------------------------|
| Shukla Yash Chetankumar | Dheeraj Kumar Singh | Sumitra Menaria |
| Student | Assistant Professor | Head of Department |
| I.T. Department | I.T. Department | I.T. Department |
| Parul Institute of Engg and Tech | Parul Institute of Engg and Tech | Parul Institute of Engg and Tech |
| Vadodara, Gujarat, India | Vadodara, Gujarat, India | Vadodara, Gujarat, India |

Abstract— Data mining has an important aspect of predicting possibilities of any event. now a days health related problem can be predicted based on various classification, clustering algorithms which uses machine learning technologies and predict the possibilities of accordance. Health care specially heart related problem are now more often that we can get more data as a training purpose this will lead to better accurate prediction. Our main objective is to predict the possibilities of heart disease so that there will be less medical tests and we can save patients who doesn't know about their disease.

Keywords—Heart disease, algorithms, ID3 Naïve bayes;

I. INTRODUCTION

Information knowledge base mining is widely used domain which has significant role in every filed and its use to predict the possibilities based on specific predefine data. It has vital role in clinical area which has huge number of patients are admitted for various treatments. That patients records can be used to predict any disease. In Information retrieval based on mining use hidden information from data set of medical fourms. There are several types of approaches are used for prediction such as classification, clustering based on stastical values.

The heart is vital organ of shape half. it's nothing quite a pump, that pumps blood through the body. If circulation of blood in body is inefficient the organs like brain suffer and if heart stops operating altogether, death happens inside minutes. Life is totally smitten by economical operating of the guts. The term malady (cardiovascular disease) refers to disease of heart system inside it.

A number of things are shown that will increase the chance of Heart disease:

- Case history
- Smoking
- Poor diet
- High pressure
- High blood sterol
- Obesity
- Physical inactivity
- Hyper tension

Factors like these area unit accustomed analyze the center illness. In several cases, identification is usually supported patient's current check results & doctor's expertise. therefore the identification could be a advanced task that needs abundant expertise & high ability.

In this paper, our main focus on Heart Disease, specifically the University of California (UCI) heart disease dataset. Numerous researches have scrutinized this dataset for better prediction procedures. Our aim of these paper is to convey a comparative study of different algorithms in estimating the heart disease accurately.

Inappropriate assessments will cause ruinous consequences that are unacceptable. Health care correct prediction additionally curtail the value of medical tests. they'll accomplish these results by exploitation acceptable machine-driven machine learning algorithmic program and correct call.

Each Dataset purpose is related to real patient. information just like the patient name, Blood teams, Heart Beats, pressure, Chest issue in age wise cluster. The conclusion of datasets assortment show the modification of person heart failure illness and symptoms of the center attack. This method of predicting heart diseases detection additionally helps doctors to predict patient health and their medical background.[10]

Technically knowledgeset is dividing in field parts and attributes in order that data will simply regenerate into Naïve Bayes algorithms for prediction. once assumption of independence holds, a Naive Bayes classifier performs higher compare to alternative models like logistical regression and you would like less coaching knowledge. It performs well just in case of categorical input variables compared to numerical variable, Gaussian distribution is assumed (bell curve, that may be a robust assumption).[12].

II. Literature Review

Automated diagnosing helps the doctors to calculate the proper malady with less time. Table one highlights the foremost objectives of the authors operating within the field of predicting medical disease(s) victimisation data processing methodology. data gained by elbow grease of aim(s) of knowledge mining are often accustomed create booming choices which will improve success of care organization and health of the patients.

Table-1: Authors and their objective

| Author | Objective |
|---------------------|---|
| Asha Rajkumar et al | To achieve high accuracy by classifying algorithms |
| Jyoti Soni | Predictive data mining for medical diagnosis: An overview of heart disease |
| Akhil jabbar et al | Proposed a system for heart disease prediction using data mining techniques |
| Mai Shouman | Applying K-nearest neighbor in diagnosing heart disease patients |
| Abhishek Taneja | To design a predictive model for heart disease detection to enhance their liability of heart disease diagnosis. |

As above mentioned in table there are various work has been completed to achieve successful prediction. There are various methods that was proposed by different authors.

Purusothaman G et al has compared various algorithm and their accuracy and concluded that best model is hybrid model. Single data mining models:

Decision Tree 76%
 Associative Rules 55%
 K-NN 58%
 Artificial Neural Networks 85%
 Support VectorMachine 86%
 Naïve Bayes 69%
 Hybrid models 96%

Gnanasoundhari SJ et al and John Peter T has proposed different methodologies and came with output that Weighted Associative Classifier gave best results with 11 attributes.

Naive Bayes 52.33%
 Naïve Bayes 83.70%
 Decision Tree 76.66%
 K-NN 75.18%
 Neural Network 78.485

Halaudi DM has proposed comparison of five methodologies with 11 attributes and prove that J48, REPTREE, and SIMPLE CART algorithm have best results.

J48 99.0741%
 Bayes Net 98.148%
 Naive Bayes 97.222%
 Simple Cart 99.0741%
 REPTREE 99.0741%

Srinivas K et al has given comparative analysis on 15 attributes and Neural networks (MLP) gave best results.

Decision Tress (C4.5) 82.5%
 Neural networks (MLP) 89.75
 Naïve Bayes 82%
 SVM 82.5%

Chaitrali S gave the best results with 13 attributes and 15 attributes. which are as follows:

Decision Trees 96.66% 99.62%

Naive Bayes 94.44% 90.74%
 Neural Networks 99.25% 100%

We have done some literature survey of different papers

Approach[2] which summarize that This method refer to the methods implementing the naïve bayes algo in heart disease and it has limitations that There is an complexity in implementing the method. So it might affect accuracy

[5,6] deals with the data in the classification of the information mentioned in the databases Here the accuracy factor is very much important. Usage of weka tool is important. In that method there is limitation is that it fives more accurate rate can be classified and sometime the same method will not give the desired outcome in the other models.

[10,12] summarize that Usage of the different algorithms is done which include naïve and decision tree. There will be an comparison of the methods and then it will be decided which method can be used for the heart disease it has limitation that Sometimes there can be difficulty in applying the desired method and for the decision tree proper information gain should be calculated.

After this we have sturdy various papers and create comparative analysis based on different methodologies.

Paper-1

| Methods Used | Remark | Future Work |
|------------------------------|---|--|
| Decision Tree Naïve Bayes | Decision trees provides accurate results as compare to Naive Bayes. | Clustering, Time series, Association rules can be used |

Paper-2

| Methods Used | Remark | Future Work |
|------------------------------|--|---|
| Decision Tree Naïve Bayes | Decision tree was found to be the best. It gave the most accurate result whether the patient had the possibility of the heart disease. | detect the specific type of heart disease in particular |

Paper-3

| Methods Used | Remark | Future Work |
|--------------|--|--|
| Naïve Bayes | It may also incorporate different data processing techniques, e.g., time series, clustering and association Rules. | -Text Mining to mine the vast quantity of unstructured information offered in healthcare databases |

Paper-4

| Methods Used | Remark | Future Work |
|---|--|--|
| Nine voting Equal Frequency Discretization with Gini Index Decision Tree | tested decision tree type and voting to identify a more robust, more accurate method.. | larger volumes of data from different medical databases -same can be extended in different horizontal and vertical domains of medical science |

Paper-5

| Methods Used | Remark | Future Work |
|------------------------------------|---|---|
| Decision trees, Naive Bayes & KNN. | Nearest neighbour is very handy and useful for all kind of datasets even for large size datasets. | increase the speed of algorithm using proximity graphs. |

Paper-6

| Methods Used | Remark | Future Work |
|--------------|--|--|
| KNN and ID3 | risk rate of heart disease was detected and accuracy level also provided for different number of ttributes | numbers of attributes could be reduced and accuracy would be |

Paper-7

| Methods Used | Remark | Future Work |
|--|---|--|
| Generalized Linear Model,SVM, Bagging algorithm, Boosting, Hybrid Classifier with Weighted Voting - HCWV | achieved higher accuracy by using hybrid data | Generalized Linear Model,SVM, Bagging algorithm, Boosting, Hybrid Classifier with Weighted Voting - HCWV |

Paper-8

| Methods Used | Remark | Future Work |
|---|--|--|
| SVM Classifier and GA optimization , clustering | K-mean clustering and MAFIA algorithm for Heart disease prediction system and achieved the accuracy of 89% . | improved algorithm of clustering which achieve the accuracy more than the present algorithm. |

Paper-9

| Methods Used | Remark | Future Work |
|-------------------------|--------|-------------------------|
| hidden naïve bayes(HNB) | | hidden naïve bayes(HNB) |

III. CONCLUSION

We have discussed various importance of data mining in medical science how data mining techniques can be used to predict the possibilities of heart disease. we have discussed approaches of different authors and compares the results of each algorithm Our main focus was on analysis of various algorithms and study their work and summarise the limitations that will help us to propose our work

. In future, the records of attributes could be reduced and accuracy would be increased using some other algorithms.

REFERENCES

- Karthiga, A. Sankari, M. Safish Mary, and M. Yogasini. "Early Prediction of Heart Disease Using Decision Tree Algorithm."
- Priyanka, N., and Pushpa RaviKumar. "Usage of data mining techniques in predicting the heart diseases—Naïve Bayes & decision tree." *Circuit, Power and Computing Technologies (ICCPCT), 2017 International Conference on.* IEEE, 2017.
- Singh, Garima, et al. "Heart disease prediction using Naïve Bayes." *International Research Journal of Engineering and Technology (IRJET) e-ISSN (2017):* 2395-0056.
- Kirmani, Mudasir Manzoor, and Syed Immamul Ansarullah. "Prediction of Heart Disease using Decision Tree a Data Mining Technique." *IJCSN International Journal of Computer Science and Network* 5.6 (2016): 885-892.
- Rairikar, Abhishek, et al. "Heart disease prediction using data mining techniques." *Intelligent Computing and Control (I2C2), 2017 International Conference on.* IEEE, 2017.
- Thomas, J., and R. Theresa Princy. "Human heart disease prediction system using data mining techniques." *Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on.* IEEE, 2016.
- Saini, Meenal, Niyati Baliyan, and Vineeta Bassi. "Prediction of heart disease severity with hybrid data mining." *Telecommunication and Networks (TEL-NET), 2017 2nd International Conference on.* IEEE, 2017.
- Chadha, Ritika, and Shubhankar Mayank. "Prediction of heart disease using data mining techniques." *CSI transactions on ICT* 4.2-4 (2016): 193-198.
- Jabbar, M. A., and Shirina Samreen. "Heart disease prediction system based on hidden naïve bayes classifier." *International Conference on Circuits, Controls, Communications and Computing (I4C).* Vol. 3. No. 62. 2016.
- Rairikar, Abhishek, et al. "Heart disease prediction using data mining techniques." *Intelligent Computing and Control (I2C2), 2017 International Conference on.* IEEE, 2017.
- Shirwalkar, Nikita, et al. "Human Heart Disease Prediction System Using Data Mining Techniques." (2018).
- Alizadehsani, Roohallah, et al. "Non-invasive detection of coronary artery disease in high-risk patients based on the stenosis prediction of separate coronary arteries." *Computer methods and programs in biomedicine* 162 (2018): 119-127.
- Kinge, Durga, and S. K. Gaikwad. "Survey on data mining techniques for disease prediction." (2018).
- Singh, Navdeep, Punjab Ferozpur, and Sonika Jindal. "Heart disease prediction using classification and feature selection techniques." (2018).
- Kaur, Amandeep, and Jyoti Arora. "HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES: A SURVEY." *International Journal of Advanced Research in Computer Science* 9.2 (2018).
- Shylaja, S., and R. Muralidharan. "Comparative Analysis of Various Classification and Clustering Algorithms for Heart Disease Prediction System." *Biometrics and Bioinformatics* 10.4 (2018): 74-77.
- Saini, Meenal, Niyati Baliyan, and Vineeta Bassi. "Prediction of heart disease severity with hybrid data mining." *Telecommunication and Networks (TEL-NET), 2017 2nd International Conference on.* IEEE, 2017.
- Patel, Ajud, et al. "Heart Disease Prediction Using Data Mining." *International Research Journal of Engineering and Technology* 4.1 (2017): 1705-1707.
- Priyanka, N., and Pushpa RaviKumar. "Usage of data mining techniques in predicting the heart diseases—Naïve Bayes & decision tree." *Circuit, Power and Computing Technologies (ICCPCT), 2017 International Conference on.* IEEE, 2017.
- Glymour, C., D. Madigan, D. Pregidon and P. Smyth, 1996. *Statistical inference and data mining.* Communication of the ACM, pp: 35-41.
- Shams, K. and M. Frashita, 2001. *Data Warehousing Toward Knowledge Management.* Topics in Health Information Management, 21: 3.
- Frawley and Piatetsky-Shapiro, 1996. *Knowledge Discovery in Databases:An Overview.* The AAAI/MIT Press, Menlo Park, C.A.

23. Miller, A., B. Blott and T. Hames, 1992. Review of neural network applications in medical imaging and signal processing. *Med. Biol. Engg. Comp.*, 30: 449-464.
24. Glymour, C., D. Madigan, D. Pregidon and P.Smyth, 1996. Statistical inference and data mining. *Communication of the ACM*, pp: 35-41.
25. Chen, J., Greiner, R.: Comparing Bayesian Network Classifiers. In *Proc. of UAI-99*, pp.101–108 ,1999. 6. "Hospitalization for Heart Attack, Stroke, or Congestive Heart Failure among Persons with Diabetes", Special report: 2001 – 2003, New Mexico.
26. 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), pp.642-656 © EuroJournals Publishing, Inc. 2009.
27. Sellappan Palaniappan, Rafiah Awang, Intelligent Heart Disease Prediction System Using Data Mining sTechniques, 978-1-4244-1968- 5/08/\$25.00 ©2008 IEEE.
28. Richard N. Fogoros, M.D, The 9 Factors that Predict Heart Attack 90% of heart attacks are determined by these modifiable risk factors, About.com Guide.
29. Harleen Kaur and Siri Krishan Wasan, Empirical Study on Applications of Data Mining Techniques in Healthcare, *Journal of Computer Science* 2 (2): 194-200, 2006 ISSN 1549-3636 © 2006 Science Publications.
30. <https://www.youtube.com/watch?v=YziLtwSNKr0>
31. <https://www.youtube.com/watch?v=D3DHNDD-hbY>
32. Prediction of heart disease using the decision tree.
33. Usage of data mining in predicting heart disease using Naïve Bayes

