IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 10, Issue 7, July 2022

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

9940 572 462 | 6381 907 438 | ijircce@gmail.com | www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007100 |

Smart Healthcare System for Automatic Diagnosis of Heart Diseases Using Blockchain

SHIYAM V, Dr.M.Kannan, M.E.,M.S., Ph.D.,

Master of Engineering, Department of Computer Science and Engineering, Mahendra Engineering College (Autonomous), Namakkal, Tamilnadu, India

Professor & HoD, Department of Computer Science and Engineering, Mahendra Engineering College (Autonomous), Namakkal, Tamilnadu, India.

ABSTRACT : Health care is an inevitable task to be done in human life. Health concern business has become a notable field in the wide spread area of medical science. Health care industry contains large amount of data and hidden information. Effective decisions are made with this hidden information by applying data mining techniques**:** As large amount of data is generated in medical organization's (hospitals, medical centers) but as this data is not properly used. There is a wealth of hidden information present in the datasets. This unused data can be converted into useful data. For this purpose we can use different data mining techniques. This paper presents a classifier approach for detection of heart disease and shows how Naive Bayes can be used for classification purpose. In our system, we will categories medical data into five categories namely no, low, average, high and very high. Also, if unknown sample comes then the system will predict the class label of that sample. Hence two basic functions namely classification (training) and prediction (testing) will be performed. Accuracy of the system is depends on algorithm and database used.

KEYWORDS*: healthcare system, Automatic Diagnosis, heart diseases, naïve bayes algorithm, heart diseases prediction system, blockchain.*

I. INTRODUCTION

Health care system is the preservation of mental and physical heath by preventing or treating illness through services offered by the provision. In day today life heart diseases is the major cause of deaths in the world. The world health organization (who) has estimated that 12 million death occurred worldwide, each year due to heart diseases. Over 80% of deaths in world are because of heart diseases. Who estimated in future, almost 23.6 million people will die due to heart diseases. The euro heart survey on heart diseases was conducted by 25 countries it included adults with moderate severe native heart diseases, infective endocarditic are previous valve intervention. Heart diseases was native in 71.9% of patients and 28.1% had had a previous intervention. Mean age was 64 to 14 years. Here of rheumatic origin. Data mining has been played an important role in the intelligent medical health care systems. Medical data mining in health care is regarded as an important yet complicated task that needs to be executed accurately and efficiently. Health care data mining attempts to solve real world health problems in diagnosis and treatment of disease. The relationship of disorders and real cause of disorders and the effects of symptoms that are spontaneously seen in patients can be evaluated by using the heart disease prediction system, is a computerized method for diagnosing heart diseases based on prior data and information.

II. RELATED WORK

Heart disease is a term that assigns to a large number of medical conditions related to heart. These medical conditions describe the abnormal health conditions that directly influence the heart and all its parts. Heart disease is a major health problem in today's time. This paper aims at analyzing the various data mining techniques introduced in recent years for heart disease prediction. Table 1 shows different data mining techniques used in the diagnosis of Heart disease over different Heart disease datasets.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007100 |

2.1C4.5

C4.5 algorithm is a greedy algorithm and was developed by Ross Quinlan and its used for the induction of decision trees.C4.5 is often referred to as a statistical classifier. The decision tree algorithm C4.5 is developed from ID3 in the following ways

2.2K-NN

It is the nearest neighbor algorithm. The k-nearest neighbor's algorithm is a technique for classifying objects based on the next training data in the feature space. The algorithm operates on a set of d-dimensional vectors, D = {xi | i = 1. . . N}, where xi $\in$ k d denotes the i^{th} data point. Selection k points in k d as the initial k cluster representatives. Algorithm iterates between two steps till junction.

FigTable 1: Table shows different data mining techniques used in the diagnosis of Heart disease over different Heart Disease datasets

Author	Year	Technique used	Attributes
Dr.K.Usharani	2011	Clasification/Neural Networks	13
Jesminahar,et al	2013	Apriori/Predictive Apriori/Tertius	14
Latha,et al	2008	Genetic Algorithm/CANFIS	14
Majabber,et al	2011	Clustring/Association Rule	14
		Mining/Sequence number	
Ms.Lshtake et al.	2013	DecisionTree/Neural Network/Naive bayes	15
Nan-Chen et al.	2012	(EVAR)/Machine Learning/Markov blanket	
Oleg et al.	2012	ANN/Genetic Poliymorphisms	
Shadab et al	2012	Naive bayes	15
Shantakumar et al.	2009	MAFIA/Clustring/KMeans	13
Carlos et al	2001	Association Rule	25

III. METHODOLOGY

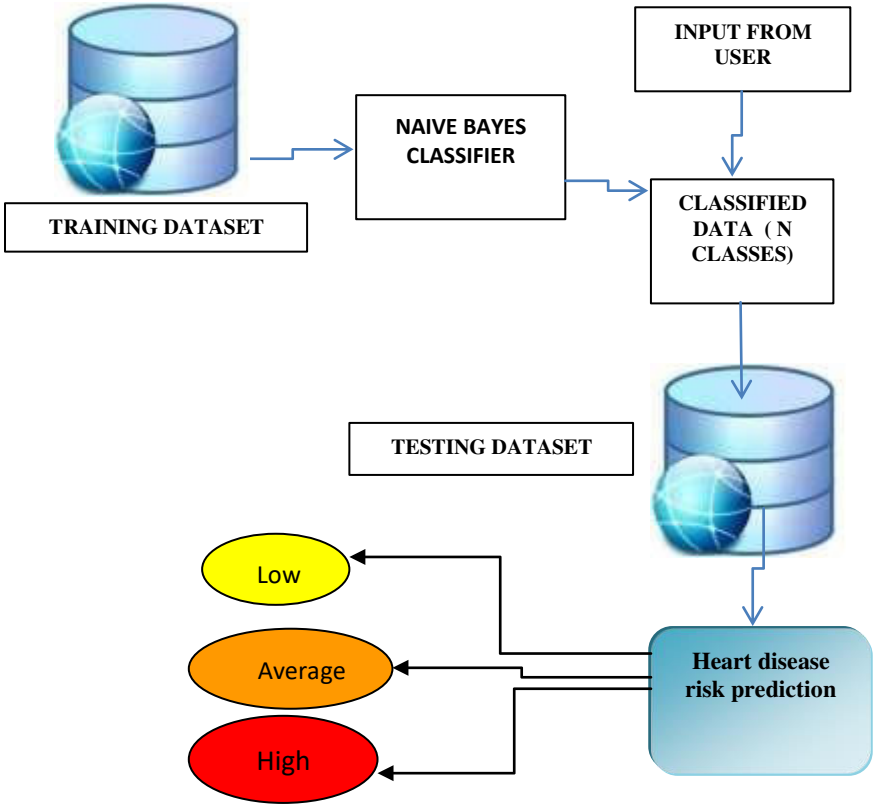
3.1Naive Bayes:

Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful in the field of medical science for diagnosing heart patients. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007100 |



IV. DATASET

The data set used in this work are clinical data set collected from one of the leading diabetic research institute in Chennai and contain records of about 500 patients. The clinical data set specification provides concise, unambiguous definition for items related to diabetes. The diabetes data set is developed to ensure people with diabetes have up to date records of their risk factors, current management, treatment target achievements and arrangements and outcomes of regular surveillance for complications, to help them monitor their care and make informed choices about their management. It will also ensure that when people with diabetes meet health care professionals the consultation is fully informed by comprehensive, up to date and accurate information. The diabetes attributes used in our proposed system and their descriptions are shown in FigTable 2.

FigTable 2: Parameters of Heart Diseases Prediction System

Name	Type	Description
Age	Continuous	Age in years
Sex	Discrete	1 = male 0 = female

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007100 |

Cp	Discrete	Chest pain type: 1 = typical angina 2 = atypical angina 3 = non-angina pa 4 =asymptomatic
Trestbps	Continuous	Resting blood pressure (in mm Hg)
Chol	Discrete	Serum cholesterol in mg/dl
Fbs	Discrete	Fasting blood sugar > 120 mg/dl: 1 = true
		0 = false
Restecg	Continuous	Resting electrocardiographic results: 0 = normal 1 = having ST-T wave abnormality 2 =showing probable or define left ventricular hypertrophy by Estes ‘criteria
Thalach	Discrete	Maximum heart rate achieved
Exang	Discrete	Exercise induced angina: 1 = yes 0 = no
Slope	Discrete	The slope of the peak exercise segment : 1 = up sloping 2 = flat 3= down sloping

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007100 |

Diagnosis	Discrete	Diagnosis classes: 0 = healthy 1= possible heart disease

V. DATA MINING TOOL

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. The experiments are conducted using the weka tool and the results are obtained. We have used the navie bayes method to perform classification by using 70% of percentage split.

VI. DATA ANALYSIS

In this system the medical data set is classified based on the classes present/absent.The results clearly states that naive bayes provides better results regarding the people affected by heart diseases. The proposed naïve bayes model was able to classify 86.4198% of the input instances correctly and the incorrect instances was 13.5802% for 70% of percentage split. With the total of 81 instances 70% was classified as correct and 11% instances was incorrect.

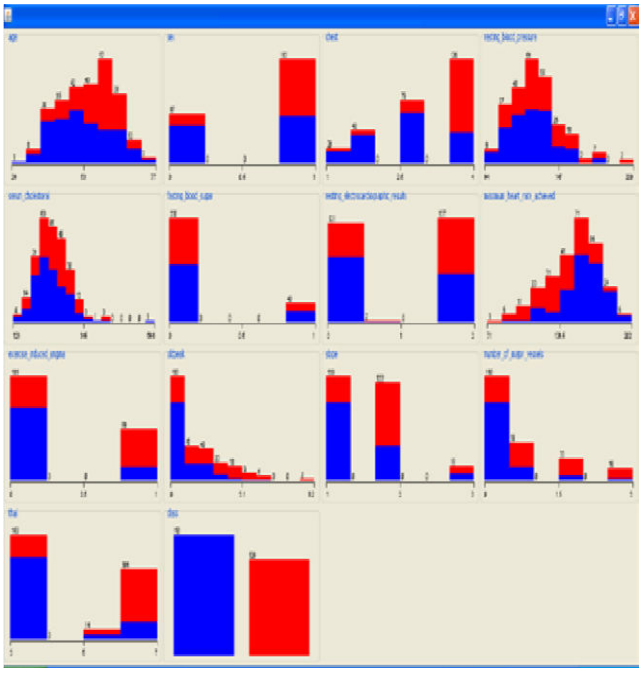


Fig.1. Attribute value distribution

The blue colored regions in the graphs in Figure 1 denote high cholesterol values. From the graphs we can see that, most of the diabetic patients with high cholesterol values are in the age group of 45 – 55, have a body weight in the range of 60 – 71, have BP value of 148 or 230, have a Fasting value in the range of 102 – 135, have a PP value in the range of 88 – 107, and have a A1C value in the range of 7.7 – 9.6**.**

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

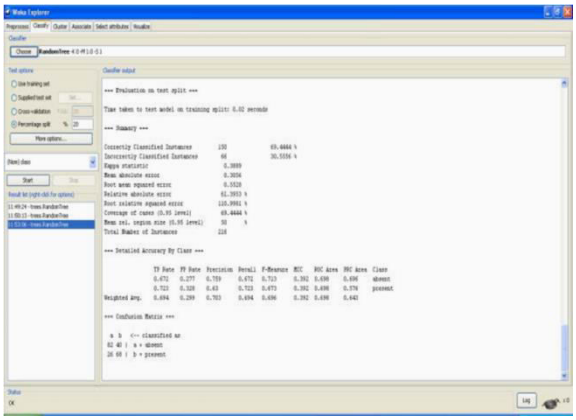
|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007100 |

VII. RESULTS AND DISCUSSIONS

The results of our experimentation are shown in Fig2.

Fig. 3 Result window of the data mining process



The proposed naïve bayes model was able to classify 74% of the input instances correctly. It exhibited a precision of 71% in average, recall of 74% in average, and F-measure of 71.2% in average. The results show clearly that the proposed method performs well compared to other similar methods in the literature, taking into the fact that the attributes taken for analysis are not direct indicators of heart disease.
The first step is the pre-processing to remove the inconsistent data. Apply the algorithm is used to classify the data.

Sl no	Attribute	Data-Type
1	Age	Numeric
2	Gender	Nominal
3	BP Systolic	Nominal
4	Diabetic	Numeric
5	BP Dialic	Numeric
6	Height	Numeric
7	Weight	Numeric
8	BMI	Numeric
9	Hypertension	Nominal
10	Rural	Nominal
11	Urban	Nominal
12	Disease	Nominal

Table 1: Data Set

VIII.WEKA TOOL

WEKA is a data mining tool. It provides the facility to classify the data through various algorithms. **WEKA** stands for **W**aikato **E**nvironment for **K**nowledge Learning and it was developed by the University of Waikato, New Zealand Software.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007100 |

IX. CLASSIFICATION MATRIX

The basic phenomenon used to classify the Heart disease classification using classifier is its performance and accuracy. The performance of a chosen classifier is validated based on error rate and computation time. The classification accuracy is predicted in terms of Sensitivity and Specificity. It compares the actual values in the test dataset with the predicted values in the trained model. In this example, the test dataset contained 208 patients with heart disease and 246 patients without heart disease.

Predicted	Classified as a Healthy(0)	Classified as a not Healthy(1)
Actual Healthy (0)	**TP**	**FN**
Actual not Healthy (1)	**FP**	**TN**

Table 2: confusion matrix

Where

- **TP- True Positive**
- **TN - true Negative**
- **FP - False Positive**
- **FN - false Negative**
- **CT- Computing Time**

For measuring accuracy rate and Error Rate, the following mathematical model is used.

Accuracy = TP+TN/ TP+FP+ TN +FN
CVError Rate = FP + FN / TP+FP+ TN +FN

Performance Comparison (C 4.5 and K-NN)

Dataset	SMO	Naïve Bayes	C4.5	J48	kNN
Heart Problem	77.34	76.3	**84.5**	73.8	83.2

Table 3: Performance Comparison

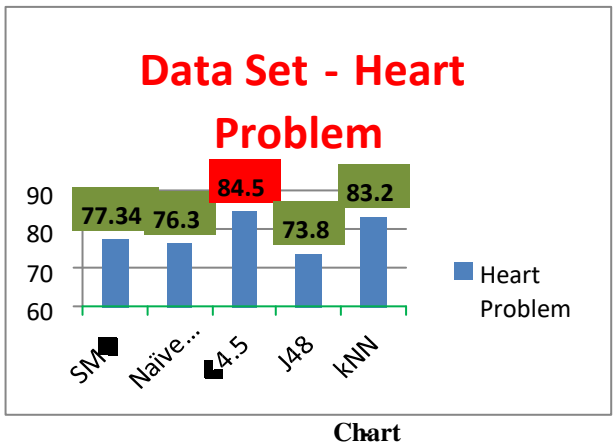


Fig4: Performance comparison bar

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.165

|| Volume 10, Issue 7, July 2022 ||

| DOI: 10.15680/IJIRCCE.2022.1007100 |

X. CONCLUSION

Data mining applications are used vastly in the medical field to detect diseases and diagnosis the heart patient based on the data set and the attributes provided. Researchers have been investigating applying different data mining techniques to help health care professionals in the diagnosis of heart disease. In the proposed work navie bayes algorithm is used to classify the data set because navie bayes provides accurate results, with these results heart diseases among people is predicted. Thus heart diseases prediction system successfully diagnose the medical data and predicts the heart diseases.The results thus obtained shows that navie bayes algorithm provides 86.4198% of accuracy with minimum time.

REFERENCES

1. S. K. Yadav, B. K. Bharadwaj& Pal, S. 2011. Data Mining Applications: A comparative study for predicting students' performance, International journal of Innovative Technology and Creative Engineering (IJITCE), 1(12).
2. Sitar-Taut, V.A., et al., Using machine learning algorithms in cardiovascular disease risk evaluation. Journal of Applied Computer Science & Mathematics, 2009.
3. Wu, X., et al., Top 10 algorithms in data mining analysis. Knowl. Inf. Syst., 2007.
4. S. K. Yadev& Pal., S. 2012. Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification, World of Computer Science and Information Technology (WCSIT), 2(2), 51-56.
5. Sitar-Taut, V.A., et al., Using machine learning algorithms in cardiovascular disease risk evaluation. Journal of Applied Computer Science & Mathematics, 2009.
6. Srinivas, K., B.K. Rani, and A. Govrdhan, Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. International Journal on Computer Science and Engineering (IJCSE), 2010. Vol. 02, No. 02: p. 250-255.
7. Agarwal, S., Pandey, G. N., & Tiwari, M. D. Data Mining in Education: Data Classification and Decision Tree Approach.
8. Merceron, A., & Yacef, K. (2005, May). Educational Data Mining: a Case Study. In AIED (pp. 467-474).
9. Bakar, A. A., Othman, Z. A., & Shuib, N. L. M. (2009, October). Building a new taxonomy for data discretization techniques. In Data Mining and Optimization, 2009. DMO'09. 2nd Conference on (pp. 132-140). IEEE.
10. Burrows, W. R., Benjamin, M., Beauchamp, S., Lord, E. R., McCollor, D., & Thomson, B.
11. (1995). CART decision-tree statistical analysis and prediction of summer season maximum surface ozone for the Vancouver, Montreal, and Atlantic regions of Canada. Journal of applied meteorology, 34(8), 1848-1862.
12. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. Information Theory, IEEE Transactions on, 13(1), 21-27.
13. Dasarathy, B. V. (1980). Nosing around the neighborhood: A new system structure and classification rule for recognition in partially exposed environments. Pattern Analysis and Machine Intelligence, IEEE Transactions on, (1), 67-71.
14. de Oña, J., López, G., & Abellán, J. (2012). Extracting decision rules from police accident reports through decision trees. Accident Analysis & Prevention.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.165

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details