Medical Database Mining for Heart Disease Precautions and Early Call Up

Mahima Choudhary

Assistant Professor, University Dept. of Computer Science, University of Mumbai

---***---

Abstract - Modern medicine produces a great deal of information that is stored in the medical database. Extracting useful information and making scientific decision for diagnosis and treatment of disease increasingly becomes obligatory from the database. Data mining in medicine can deal with this difficulty. It can also improve the management level of hospital information and encourage the progress of tele-medicine and community medicine. Medical field is first and foremost directed at patient care activity and secondarily as research resource. Collecting medical data is to benefit the individual patient. Mainly, storage of medical information of patients who come for heart disease hospitalization then these proposed algorithms are run on that information and result will be provided in the form of user comprehensible words and graphs.

Index Terms — Medical, Heart Disease, Naïve Bayes, Decision Tree.

I. INTRODUCTION

TODAY, data mining is an effective tool for decision making from production to health, shopping habits to marketing, information security From military applications to military applications has been used. Social media like Facebook and WhatsApp channels meet their income through data [1].

Heart is important part of our human body. More than country affected for heart disease every year some of the million people for death from heart disease. Life is itself dependent on efficient working a heart as brain, circulation of blood in body is inefficient the organs like brain suffer and if then heart is not properly within it. More than disease attack for heart. Now days many hospitals not proper treatment. But increasing the payment of bill. Some of hospitals average treatment for patients so result is better. Heart disease is a group of condition affecting the structure and function of heart and has more root causes. Heart disease is the leading cause of every year death in the world. Some types of disease occurs attack for heart. Types of disease considered are coronary heart disease, angina pectoris, congestive heart failure, cardiomyopathy, congenital heart disease, arrhythmias, myocarditis, heart attack; heart cancer etc. in this disease is particularly specific very dangers disease to cardiovascular disease or coronary heart disease. There are considered some important reasons of heart disease:

- Age
- Smoking
- Sugar
- Obesity
- Depression
- Hyper tension
- High blood cholesterol
- Poor diet
- Family history
- Physical inactivity

Types of heart diseases

a.) Angina: It can be referred to as angina pectoris. It occurs when an area of the heart muscle does not get enough oxygen. The patient well experiences in chest discomfort, tightness or pain. It is a symptom of coronary artery disease. Due to lack oxygen in the heart muscle is usually caused by the narrowing of the coronary arteries because of plaque accumulation.

b.) Arrhythmia: Arrhythmia is an irregular heartbeat. They caused problems with heart-rhythm. It happens when the heartbeats do not work properly. To make the heart beat in a better way it should not, either move too fast, slow or erratically.

c.) Fibrillation: Fibrillation occurs when the heartbeat is irregular. We experience irregular heartbeats. We feel like a fluttering or a racing heart. Precaution has to be taken when they veer too far from normal heartbeat. Irregular heartbeats can become fatal.

d.) Congenital heart disease: It refers to born with it. In the country UK it is surveyed that every 1,000 babies are born with some kind of congenital heart disease.

e.) Coronary artery disease: It causes disease or damaged because of cholesterol containing deposits. Plaque accumulation narrows the coronary arteries and the heart gets less oxygen.

f.) Myocardial infarction: The other name is known as heart attack, cardiac infarction and coronary thrombosis. Interrupted blood flow damages or destroys part of the heart muscle. It is usually caused due to blood clot. It can also occur if an artery suddenly becomes narrows.

g.) Heart failure: The other name called as congestive heart failure. It does not pump blood around the body efficiently. Left side or Right side or both side of the body might be affected. Coronary artery disease can make the heart stiff or weak to fill and pump properly.

Interpretable Machine Learning refers to machine learning models that can provide explanations regarding why certain predictions are made. In many domains where user trust in the predictions of machine learning systems is needed, merely providing traditional machine learning metrics like AUC, precision, and recall may not be sufficient. While machine learning techniques have been employed for decades, the expansion of these techniques into fields like healthcare have led to an increased emphasis for explanations of machine learning systems. Clinical providers and other decision makers in healthcare note interpretability of model predictions as a priority for implementation and utilization. As machine learning applications are increasingly being integrated into various parts of the continuum of patient care, the need for prediction explanation is imperative. Machine learning solutions are being used to assist providers across clinical care domains as well as clinical operations, and costs. Decisions based on machine learning predictions could inform diagnoses, clinical care pathways, and patient risk stratification, among many others. It follows, that for decisions of such import, clinicians and others desire to know the “reason” behind the prediction. In this tutorial, we will give an extensive overview of the various nuances of what constitutes explanation in machine learning, explore multiple definitions of explanation.

The contexts within healthcare systems where it may be prudent to ask machine learning systems for explanations vs. explanation agnostic contexts will also be explored. Thus a physician may be greatly interested in knowing why a machine learning system is suggesting a cancer diagnosis vs. a hospital ED planner would rarely be interested in knowing why a machine learning system is making predictions about hourly arrivals in ED. We also discuss how these definitions map to various machine learning systems and algorithms that are available today - all within a healthcare context. We use results from our research on performance comparison of interpretable models on real world problems like risk of readmission prediction, ED utilization prediction and hospital length of stay prediction to explore the constraints and drivers around going about using explainable machine learning algorithms in various healthcare contexts.

II. LITERATURE SURVEY

In paper [1], Chen proposed a new convolutional neural network based multimodal disease risk prediction algorithm by using structured and unstructured data of hospital. Authors invented disease prediction system for the various regions. Also, predict that whether a patient experiences from the high risk of cerebral infarction or low risk of cerebral infarction. The accuracy of disease prediction reaches up to the 94.8% with faster speed than Convolutional neural network based unimodal disease risk prediction algorithm.

In paper [2], authors designed the Alzheimer disease risk prediction system with the help of EHR data of the patient. Here they used active learning context to solve a real problem suffered by the patient. The experts identify the similar health conditions between the two patients and on the basis of that patient risk correctly evaluated.

Designed cloud-based health –Cps system in [3], which manage the huge amount of biomedical data. This system performed various operations on cloud-like data analysis, monitoring and prediction of data. With the help of this system, a person gets more information about how to handle and manage the huge amount of biomedical data in the cloud. Also, the many services related to healthcare know by this system.

In paper [4], the author proposed wearable 2.0 system in which design smart washable clothing that improves the QoE and QoS of the next-generation healthcare system. With the help of this cloth, it captured the physiological condition of the patient. And for the analysis purpose, this data is used. Discuss the issues which are facing while designed the wearable 2.0 architecture. In this, there are many applications discussed like chronic disease monitoring, elderly people care, emotion care etc.

Proposed telehealth system [5] in that author discusses how to handle a large amount of hospital data in the cloud. For this author invented new optimal big data sharing algorithm. By this algorithm, users get the optimal solution of handling biomedical data.

In paper [6], proposed a best clinical decision-making system which predicts the disease on the basis of historical data of patients. In this predicted multiple diseases and unseen pattern of patient condition. And 2D/3D graph and pie charts designed for visualization purpose.

The heart disease prediction involve consist heart or blood vessels. The Diagnosis and prediction of heart diseases are most important so that can reduce the risk of disease [7]. In healthcare research, extreme works have been done on the prediction of heart diseases.

Mostly two algorithms used like CNN and Genetic algorithm for the prediction of heart disease. And also here major factors are considering age, family history, diabetes, hypertension, cholesterol, smoking, alcohol intake, obesity or physical inactivity etc [8].

In paper [9], for the heart disease prediction system design, authors were using machine learning algorithms like Naive Bayes, Neural network and Decision tree algorithms.

III. PROPOSED WORK

Heart disease are catastrophic for anyone and makes the life difficult to lead. In this study certain parameters are taken into the account which may help in predicting whether the patient is prone to any heart disease or not.

The parameters that are being counted for the prediction and should be taken as an input for the process of prediction are:

a. Age
b. Sex
c. Chest Pain Report
d. Blood Pressure
e. Cholesterol Levels
f. Electrocardiography Blockage Reports
g. Fast Blood Sugar Levels
h. Old Peak
i. Slope
j. Thal
k. Exang

The above stated parameters are examined properly and the numerical outcome is important over all the parameters that are being analyzed for the cumulative outcome and the overall prediction for the patient.

The implemented algorithm will be first trained over the previous datasets available from the UC Irvine Machine Learning Repository.

Implemented algorithm is based over the concept of Machine Learning and Knowledge Discovery in Database.

Where the Machine Learning's concept of Naïve Bayes helps in predicting the nature of input parameter if it is prone towards the disease or not. The second section the decision tree depicts the final call on the basis of the movement of the tree towards the left region or the right region, in process of doing this a proper priority index will be required as well for the correct judgement, in this study the priority indexing is done on the scale of 0 to 2.

Priority indexing has been done on the basis of the input's basic availability and on the basis of the related theories and algorithms studied during the research study.

Table 1: Priority Indexing

Factor	Priority Index	Description
Age	1	Patient's Age
Gender	0	Patient's Gender
Chest Pain	2	Chest Pain Status: Positive or Negative
Blood Pressure	2	Blood Pressure Status Positive or Negative
Cholesterol Level	2	Cholesterol Levels: High Cholesterol is a risk
ECG Report	2	ECG blockades stats
Ca	1	Number of major vessels that colored by fluorosopy
Blood Sugar	1	Diabetic Status and Sugar Levels
Old Peak	1	ST segments that induced by exercise relative to rest
Slope	1	Peak exercise ST Segments
Thal	1	Defect values
Exang	1	Exercise induced angina

Naïve Bayes calculates the prone chances for each parameter and is generates the output on the basis of the received ratio. Naïve Bayes analyzed output is combined with the priority index value so as to add weightage to the generated received ratio. Output range for it will be {-1, 0, 1}, where -1 represents the extremely close to the disease prone, 0 is for the close to disease prone and 1 is for the safe, however the generated range is completely a flag range.

$$Ai = \frac{P(HD|F)}{P(F_{Total})}$$

Ai represents the outcome probability for P(HD) (prone to heart disease) with respect to the factor, out of the total probability of heart disease prone with respect to the total P(F_{total}) probability prone to heart disease due to all factors.

Generated range will help the decision tree to take the right movements and hence lands over the exact locations after the iterations done for the factors.

IV. EXPERIMENTAL ANALYSIS

The implemented algorithm generated the database file over the parameters and hence generated the prediction for the patients. The data that has been examined here are for the testing purposes only.

Table 2: Execution Time

Number of Dataset	Naïve Bayes (seconds)	Decision Tree
200	0.0721	0.2225
1000	0.1473	0.258
1500	0.2569	0.2863
2000	0.2994	0.4126
3000	0.4272	0.6403

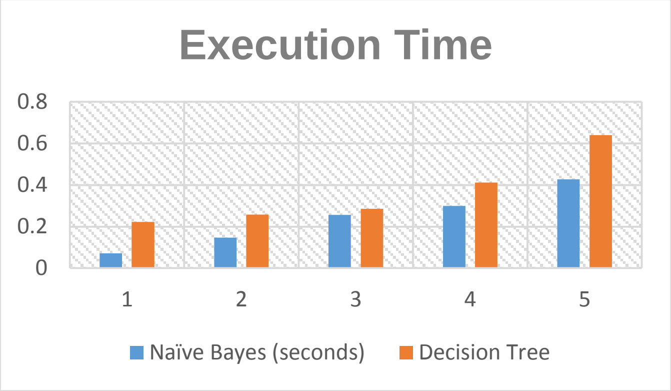
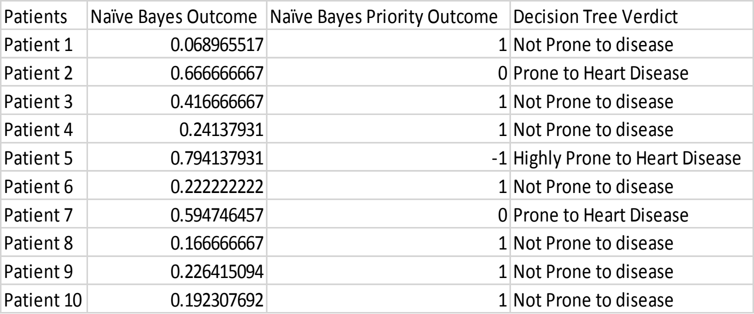


Figure 1: Execution time chart for Naïve Bayes and for Decision Tree Prediction

Table 3: Prediction Verdict



Patients	Naïve Bayes Outcome	Naïve Bayes Priority Outcome	Decision Tree Verdict
Patient 1	0.068965517	1	Not Prone to disease
Patient 2	0.666666667	0	Prone to Heart Disease
Patient 3	0.416666667	1	Not Prone to disease
Patient 4	0.24137931	1	Not Prone to disease
Patient 5	0.794137931	-1	Highly Prone to Heart Disease
Patient 6	0.222222222	1	Not Prone to disease
Patient 7	0.594746457	0	Prone to Heart Disease
Patient 8	0.166666667	1	Not Prone to disease
Patient 9	0.226415094	1	Not Prone to disease
Patient 10	0.192307692	1	Not Prone to disease

V. CONCLUSION

From the above implemented work it can be deduce that the factors affecting person's health can be analyzed very closely and a prediction can be made and if found out if the patient is prone or very close to it diagnostics can be performed and precautionary steps can be taken. However before proposing system a thorough testing is needed to be performed.

REFERENCES

[1] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," IEEE Access, vol. 5, no. 1, pp. 8869–8879, 2017.

[2] B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, " A relative similarity based method for interactive patient risk prediction," DataMiningKnowl.Discovery, vol. 29, no. 4, pp. 1070–1093, 2015.

[3] IM. Chen, Y. Ma, Y. Li, D. Wu, Y. Zhang, and C. Youn, " Wearable 2.0: Enable human-cloud integration in next generation healthcare system," IEEE Commun. , vol. 55, no. 1, pp. 54–61, Jan. 2017.

[4] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, "HealthCPS: Healthcare cyberphysical system assisted by cloud and big data," IEEE Syst. J., vol. 11, no. 1, pp. 88–95, Mar. 2017.

[5] L. Qiu, K. Gai, and M. Qiu, "Optimal big data sharing approach for telehealth in cloud computing," in Proc. IEEE Int. Conf. Smart Cloud (Smart Cloud), Nov. 2016, pp. 184–189.

[6] Ajinkya Kunjir, Harshal Sawant, Nuzhat F.Shaikh, "Data Mining and Visualization for prediction of Multiple Diseases in Healthcare," in IEEE big data analytics and computational intelligence, Oct 2017 pp.23-25.

[7] Shanthi Mendis, Pekka Puska, Bo Norrving, World Health Organization (2011), Global Atlas on Cardiovascular Disease Prevention and Control, PP. 3– 18. ISBN 978-92-4-156437-3.

[8] Amin, S.U.; Agarwal, K.; Beg, R., "Genetic neural network based data mining in prediction of heart disease using risk factors", IEEE Conference on Information & Communication Technologies (ICT), 2013, vol., no.,pp.1227-31,11-12 April 2013.

[9] Palaniappan S, Awang R, "Intelligent heart disease prediction System using data mining techniques," IEEE/ACS International Conference on Computer Systems and Applications, AICCSA 2008., vol., no., pp.108115, March 31 2008-April 4 2008.