Heart Disease Prediction using Artificial Intelligence

Zaibunnisa L. H. Malik
Computer Engineering,
M.H. Saboo Siddik Polytechnic
Mumbai, India.

Momin Fatema
Computer Engineering,
M.H. Saboo Siddik Polytechnic
Mumbai, India.

Nikam Pooja
Computer Engineering,
M.H. Saboo Siddik Polytechnic
Mumbai, India.

Gawandar Ankita
Computer Engineering
M.H. Saboo Siddik Polytechnic
Mumbai , India.

Abstract— **Artificial Intelligence techniques have been widely used in clinical decision support systems for prediction and diagnosis of various diseases with good accuracy. These classifying techniques are very effective in designing clinical support systems due to their ability to get hidden patterns and relationships in medical data provided by medical professionals. One of the most important applications of such systems is in the diagnosis of heart diseases because it is one of the leading causes of deaths all over the world. Almost all systems that predict heart diseases using clinical dataset having parameters and inputs from complex tests conducted in labs. None of the systems predicts heart diseases supporting risk factors like age, case history, diabetes, hypertension, high cholesterol, tobacco smoking, alcohol intake, obesity or physical inactivity, etc. Heart disease patients have many of those visible risk factors in common which may be used very effectively for diagnosis. A system based on such risk factors would not only help medical professionals but it would give patients a warning about the probable presence of heart disease even before the patient visits a hospital or goes for costly medical check-ups. Hence this paper presents a technique for prediction of heart disease using major risk factors with help of different Classifying Algorithms. This technique involves four major classification algorithms such as K Neighbors, Support Vector, Decision Tree, Random Forest algorithms.**

Keywords—Heart, AI, K Neighbors, Risk Factors, Algorithms

I. INTRODUCTION

Cardiovascular disease (CVD) remains the leading cause of death for adults in the United States (US) with an estimated 85.6 million Americans experiencing some form of CVD [1,2]. The term CVD is used to describe disorders of the heart and blood vessels such as coronary heart disease, stroke, congestive heart failure, and arrhythmias. African Americans comprise 13.3% of the US population (46.3 million people) yet have a three-fold greater risk of developing CVD and a two-fold greater risk of CVD-related mortality than that of non-Hispanic whites and other ethnic groups [3,4,5,6].

Coronary heart disease, a condition that originates from atherosclerosis, is the most common form of CVD and the primary cause of one-third of all US adult deaths [7]. This form of CVD remains the leading cause of death for most racial and ethnic groups, including Hispanics, non-Hispanic Whites, and African Americans [5]. Stroke risk among African Americans is 2–3 times higher than that of non- Hispanic Whites and hypertension prevalence is among the highest in the world at 44% [8].

The reasons for these disparities in disease risk and mortality are multi-factorial. Differential health care access and quality, environmental or neighbourhood influences, persistent racial discrimination, health behaviours such as diet, smoking, socioeconomic status, and genetic variation have been hypothesized as contributors to African American CVD risk [9,10]. Yet, these risk factors are not influential across all situations, and many factors are modifiable, given changed circumstances or behaviours. To lower risk, knowledge and attitudes toward CVD must be assessed to develop intervention programs that will resonate with the target audience [6,11].

Previous CVD risk studies conducted among African Americans have included those living in the US as a whole, or individuals living in states within the American South. These results are limited, as they may inadvertently omit unique characteristics of other regional populations and cultural differences within populations [6,12]. Lifestyle factors in a given region, such as the South, can provide a starting point for new investigations, but may not directly apply to other geographic areas or contexts [13]. Regionally specific qualitative insights into CVD-risk related knowledge among African Americans can fill existing research gaps and permit more accurate targeting of nutrition and lifestyle-

related programming [9,14]. For example, dietary interventions based on modification of characteristically Southern or 'Soul Food' recipes may not be relevant to all African Americans in or out of the American South where 55% of African Americans reside [15]. Due to differences in geographic origin, life experience, socioeconomic status, and food accessibility, African American culture is diverse in dietand health behaviours

II. PROBLEM STATEMENT

Recent research in the field of medicine has been able to identify risk factors that may contribute toward the development of heart disease but more research is needed to use this knowledge in reducing the occurrence of heart diseases. Diabetes, hypertension, and high blood cholesterol have been established as the major risk factors of heart diseases. Life style risk factors which include eating habits, physical inactivity, smoking, alcohol intake, obesity is also associated with the major heart disease risk factors and heart disease. There are studies showing that reducing these risk factors for heart disease can actually help in preventing heart diseases. There are many studies and researches on the prevention of heart disease risk. Data from studies of population has helped in prediction of heart diseases, based on blood pressure, smoking habit, cholesterol and blood pressure levels, diabetes. Researchers have used these prediction algorithms in adapted form of simplified score sheets that allow patients to calculate the risk of heart diseases. The Framingham Risk Score (FRS) is a popular risk prediction criterion which is used in algorithms for heart disease prediction. This study aimed at developing an intelligent system based on classification algorithms for the prediction of heart disease based on risk factors' categories.

III. EXISTING SYSTEM

The existing system modules generates comprehensive report by implementing the strong prediction algorithm. In this project the input details are obtained from the patient and the doctor. Then from the doctor inputs, using ai algorithms heart disease is analysed. Now, the obtained result is compared with the result of existing models with in the same domain and found to be improved. The main aims of the existing system to compare and check the before patient whose having disease outputs and new patient disease and determine future possibilities of the heart disease to a particular patient By Implementing the above-mentioned model we will get the goal of developing a system with increased rate of accuracy of estimating the new patient getting heart attack percentage. The data of heart disease patients collected from the UCI laboratory is used to discover patterns with K Neighbours Classifier, Support Vector Classifier, decision Tree Classifier, Random Forest Classifier. The results are compared for performance and accuracy with these AI algorithms. The model which is proposed for Heart Disease Prediction System is invented for using different algorithms of AI and approach. But by using all the existing systems theaccuracy is very less.

IV. LITERATURE REVIEW

Now a day's human beings are so busy in their existence to reap what they need and earn that they overlook to take care of their fitness. Because of this, there's alternate inside the food which they devour, their life-style changes. They are greater tensed and in very much strain to earn money so this results in blood strain, diabetes and diverse different illnesses at young age. All those reasons result in negligence in their fitness which will increase the chances of heart disorder. Heart is the maximum crucial organ of the human body and if it is affected then it additionally affects the alternative most important organs of the frame. Clinical choices are frequently made primarily based on physician's intuition and experience in preference to on the knowledge rich facts hidden in the database. This practice ends in unwanted biases, errors and excessive medical expenses which affects the satisfactory of carrier supplied to patients. HD diagnosis traditionally by using medical history of patient. However, the diagnosis results are not accurately diagnosis HD. Furthermore, these methods are not reliable in terms of accuracy and computation. There are a number of publications that propose different techniques for the extraction of features from the heart sounds and classify them using neural networks. In the late 80's Mohamed and Raafat developed a mathematical model to describe the heart sounds and murmurs by a finite number of parameters. In this case, features were extracted based on fourth order linear prediction of the cardiac cycle frames, where classification was carried out based on the minimum distance between the features of the measured pattern and the reference patterns. Patil and Kumaraswamy proposed an intelligent heart attack prediction system based on Data Mining and Artificial Neural Network. In this method, the parameters vital to the heart attack are computed by using K-means clustering algorithm to the available data. These frequent patterns are mined from the data, with the aid of the Maximal Frequent Itemset Algorithm (MAFIA). The patterns are then selected based on the computed significant weight age. Although the above study reported that this method is capable of predicting the heart attack using MAFIA algorithm, the prediction accuracy was not reported for the work. Furthermore, this technique uses features corresponding to the behavioural habits of the subject, such as smoking and alcohol consumption, instead of feature characteristics of the heart sound signal itself. In this project we have implemented AI algorithms such as:
1. K Neighbours Classifier
2. Support Vector Classifier
3. Decision Tree Classifier
4. Random Forest Classifier
Which can predict heart disease and we are first taking input from doctor about heart related information that is smoking, cholesterol, high blood pressure etc and then our system will predict the heart disease from given algorithms and will define that which algorithm is best for prediction of Heart disease.

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NREST - 2021 Conference Proceedings

V. PROPOSED WORK

The problem with risk factors related to heart disease is that there are many risk factors involved like age, usage of cigarette, blood cholesterol, person's fitness, blood pressure, stress and etc. and understanding and categorizing each one according to its importance is a difficult task. Also, a heart disease is often detected when a patient reaches advanced stage of the disease. Hence the risk factors were analyzed from various sources. The dataset was composed of 12 important risk factors which were sex, age, family history blood pressure, Smoking Habit, alcohol consumption, physical inactivity, diabetes, blood cholesterol, poor diet, obesity. The system indicated whether the patient had risk of heart disease or not.

VI. PROPOSED METHODOLOGY

The data for 50 people was collected from surveys done by the American Heart Association. Most of the heart disease patients had many similarities in the risk factors. The TABLE I below shows the identified important risk factors and the corresponding values and their encoded values in brackets, which were used as input to the system.

TABLE I
RISK FACTORS VALUES AND THEIR ENCODINGS

	Risk Factors	Values
1	Sex	Male (1), Female (0)
2	Age (years)	20-34 (-2), 35-50 (-1), 51-60 (0), 61-79 (1) , >79 (2)
3	Blood Cholesterol	Below 200 mg/dL - Low (-1) 200-239 mg/dL - Normal (0) 240 mg/dL and above - High (1)
4	Blood Pressure	Below 120 mm Hg- Low (-1) 120 to 139 mm Hg- Normal (0) Above 139 mm Hg- High (-1)
5	Hereditary	Family Member diagnosed with HD -Yes (1) Otherwise –No (0)
6	Smoking	Yes (1) or No (0)
7	Alcohol Intake	Yes (1) or No (0)
8	Physical Activity	Low (-1) , Normal (0) or High (-1)
9	Diabetes	Yes (1) or No (0)
10	Diet	Poor (-1), Normal (0) or Good (1)
11	Obesity	Yes (1) or No (0)
12	Stress	Yes (1) or No (0)
Output	Heart Disease	Yes (1) or No (0)

Data analysis has been carried out in order to transform data into useful form, for this the values were encoded mostly between a range [-1, 1]. Data analysis also removed the inconsistency and anomalies in the data. This was needed. Data analysis was needed for correct data preprocessing. The removal of missing and incorrect inputs will help the neural network to generalize well.

The proposed application is developed using python and is capable of identifying if a patient has heart disease or not. There are number of factors which increased risk of heart diseases, like family history of heart disease, smoking, cholesterol, high blood pressure, obesity, lack of physical exercise etc. Heart disease is a major health problem in today's time. Thus, there is necessity to develop a system which will predict the heart disease using Artificial intelligence. In this project we have implemented AI algorithms such as

1. K Neighbours Classifier
2. Support Vector Classifier
3. Decision Tree Classifier
4. Random Forest Classifier

Which can predict heart disease So in this project we are first taking input from doctor about heart related information that is smoking, cholesterol, high blood pressure, or whether the patient has Diabetes etc. and based on the factors our system will predict the heart disease from given algorithms and it will generate a detailed report of the heart disease So from this we can define that which algorithm is best for prediction of Heart disease.
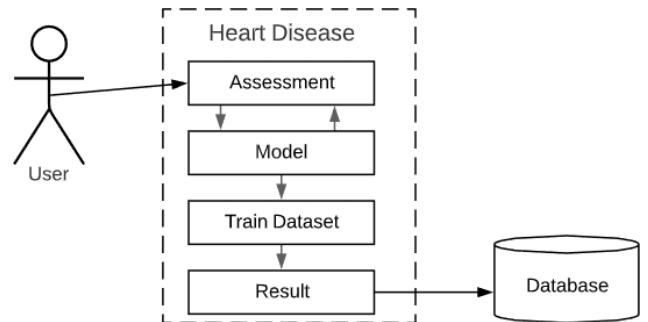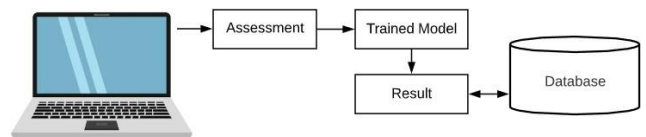

Fig 1. Block Diagram

Fig 2. System Architecture Diagram

VII. LIMITATIONS

- Need more datasets, to increase the accuracy of thealgorithms.
- The proposed application can only be used byMedical Personnel.
- The proposed application is Web-based, hencecannot be used in Mobile devices.
- The result of the application depends upon theaccuracy of the algorithms

VIII. CONCLUSION AND FUTURE ENHANCEMENT

The proposed application uses Risk Factors, which need to be identified by Medical Professionals before using the application. The result may vary based on the identified Risk Factors. If the Risk Factors identified are less accurate or wrong, the application may give wrong results. The application may use different AI techniques to capture and correct response based on past experiences. The result of the application depends on the accuracy of the Classification Algorithms. If the accuracy is low, the result generated may be wrong or less accurate. Increasing the dataset, may result in more accurate results. We can build an intelligent system

Special Issue - 2021

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NREST - 2021 Conference Proceedings

which could predict the disease using risk factors hence saving cost and time to undergo medical tests and check-ups and ensuring that the patient can monitor his health on his own and plan preventive measures and treatment at the early stages of the diseases.

REFERENCES

[1] "Global atlas on cardiovascular disease prevention and control", WHO, 2011.

[2] Mozaffarian D, Wilson PW, Kannel WB, Beyond established and novel risk factors: lifestyle risk factors for cardiovascular disease. Circulation 117: 3031–3038, 2008.

[3] Poirier P, Healthy lifestyle: even if you are doing everything right, extra weight carries an excess risk of acute coronary events. Circulation 117:3057–3059, 2008.

[4] Wood D, De Backer, Prevention of coronary heart disease in clinical practice: recommendations of the Second Joint Task Force of European and other Societies on Coronary Prevention. Atherosclerosis 140: 199–270, 1998.

[5] Anderson KM, Odell PM, Cardiovascular disease risk profiles. Am Heart J 121: 293–298. 1991.

[6] Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. Am Heart J., 121: 293–298, 1991.

[7] Kannel WB, An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol 110: 281–290, 1979.

[8] R. Das, I. Turkoglu, and A. Sengur, Effective diagnosis of heart disease through neural networks ensembles, Expert Systems with Applications, Elsevier, pp. 7675–7680, 2009.

[9] T. Porter and B. Green, "Identifying Diabetic Patients: A Data Mining Approach," Americas Conference on Information Systems, 2009.

[10] S. Panzarasa et al, "Data mining techniques for analyzing stroke care processes," in Proc. of the 13th World Congress on Medical Informatics, 2010.

[11] L. Li, T. H., Z. Wu, J. Gong, M. Gruidl, J. Zou, M. Tockman, and R. A. Clark, "Data mining techniques for cancer detection using serum proteomic profiling," Artificial Intelligence in Medicine, Elsevier, 2004.

[12] V. A. Sitar-Taut et al., "Using machine learning algorithms in cardiovascular disease risk evaluation," Journal of Applied Computer Science and Mathematics, 2009.

[13] K. Srinivas, 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), vol. 2, no. 2, pp. 250-255, 2010.

[14] H. Yan, et al., "Development of a decision support system for heart disease diagnosis using multilayer perceptron," in Proc. of the 2003 International Symposium on, vol. 5, pp. V-709- V-712.

[15] M. C. Tu, D. Shin, and D. Shin, "Effective Diagnosis of Heart Disease through Bagging Approach," Biomedical Engineering and Informatics, IEEE, 2009.

[16] Hai H.Dam, Hussain A.Abbass and Xin Yao, "Neural – Based Learning Classifier Systems", IEEE Transactions on Knowledge and Data Engineering, Vol.20, No.1, pp.26-39, 2008.

[17] Shantakumar B.Patil and Y.S.Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network", European Journal of Scientific Research, Vol.31, No.4, pp.642-656, 2009.

[18] Polat , K., S. Sahan, and S. Gunes, Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour based weighting preprocessing. Expert Systems with Applications 2007. 32 p.625–631.

[19] Das, R., I. Turkoglu, and A. Sengur, Effective diagnosis of heart disease through neural networks ensembles. Expert Systems with Applications, Elsevier, 2009. 36 (2009): p. 7675–7680.

[20] Latha Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological and Life Sciences, Vol.3, No.3, pp.157-160, 2007.