Journal of Cardiovascular Disease Research

ISSN: 0975-3583, 0976-2833 VOL 13, ISSUE 05, 2022

Heart Failure Detection Using Different Machine Learning Algorithms

Raghav Sharma sharmaraghv917@gmail.com

Mayuri Mukewar mukewarmayuri1@gmail.com

Rohini Pardeshi rohinipardeshi1998@gmail.com

Pranita Kumbhar pranitakumbhar249@gmail.com

Prof. Asmita Manna asmita.nag@gmail.com

Abstract:- Being an important and dominant part of the human body, the heart is responsible for proper blood circulation to other parts of the body. It is very essential that the heart functions at a good and proper rate so that the circulation of blood is constant all over the human body. But, nowadays there are many heart diseases that are majorly responsible for the degradation of heart rate. Among these heart diseases, Coronary Artery Disease(CAD) is the most popular and very common heart disease which is caused by plaque buildup in the walls of coronary arteries. As per a WHO study, around 29.2% (11.7 million) of people on earth suffer from numerous heart diseases and most of them die due to the effect of Coronary Artery Disease(CAD). In order to study these heart diseases, many habitual methods and examinations have been taken by medical scientists. However, these methods can be reduced and understood in a better & easier way by the use of ML and AI techniques. A model can be created using ML and AL techniques in order to overcome the insufficiency of data analytic tools to determine Heart disease. In this paper, we have used algorithms in ML techniques to predict whether the person has heart disease or not (especially CAD or not) by just taking values of some important attributes like type of chest pain, Resting Blood Pressure, Maximum heart rate, etc.

Keywords: Comparative study, Artificial Intelligence, Exploration, Naive Bayes, K-Nearest Neighbor, Random Forest, Decision tree, K-means, Cardiovascular Artery Disease, type of chest pain, Resting Blood Pressure, Maximum heart rate.

1. INTRODUCTION:

It is been observed by professionals that the risk rate of the heart is rising uninterruptedly. According to WHO, 80% of 17.9 million deaths are majorly due to cardiovascular heart diseases. Due to this, major countries are being affected a lot. Thus, prior prediction of the risk to the heart is very important. Currently, in order to predict health issues priorly, AI and ML techniques are in demand. For our work, we have referred to various research works related to AI and

ML algorithms which have helped in the prior prediction of some health conditions. In our work, we have used ML algorithms like Decision trees, Random Forest, K-nearest Neighbor, and K-means. We have calculated the accuracies and scores of these algorithms and the one which gave the highest accuracy, that we have used to further predict whether the person has heart disease or not.

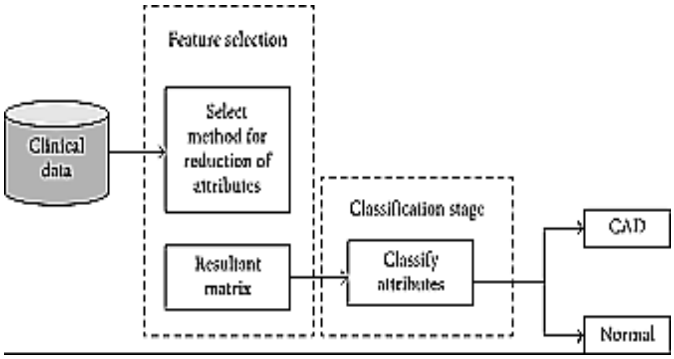
2. LITERATURE REVIEW:

ISSN: 0975-3583, 0976-2833 VOL 13, ISSUE 05, 2022

Author	Title	Method	Pros	Cons	Remark
Rajesh N, T Maneesha, Shaik Hafeez, Hari Krishna (2018)	Machine Learning Algorithms for Heart Disease Prediction	Naive Bayes	Combination of AI methods such as Naive Bayes and K-means gives better and favorable results.	In a few cases, the Naïve Bayes does not give sufficient results so it is needed to consider the effectiveness of different ML methods.	If the input given data is clean and well maintained then Naive Bayes Algorithm gives accurate results
Hadia Amin, Abita Devi, Nida Ul Amin (2019)	K-means and the Apriori Algorithm are used to predict heart disease.	K- Means and Apriori	Accurate results are shown by the K-Means algorithm	Apriori algorithm alone is not efficient enough to give expected results	Fusion of Apriori and K-means algorithms produce better results.
K.Vembandasamy, R.Sasipriya.and E.Deepa,(2015)	Heart Diseases Detection Using Naive Bayes Algorithm	Naive Bayes	The risk rate of the heart can be easily determined using the Naive Bayes Algorithm.	In some cases when small datasets are dealt with, Decision Trees give less accurate results.	In this paper the Naive Bayes 86.4% of the correctness while consuming less time
Zerina Masetic, AbdulHamit Subasi(2016)	Congestive heart failure detection using random forest classifier	Random Forest	Both simple and congestive heart failure can be treated with machine literacy approaches (CHF).		The results of different trials were observed in this paper using different statistical criteria and it was found that arbitrary timber produces 100 delicacies.
Chithambaram T, Logesh Kannan N, Gowsalya M (2019)	Heart Disease Detection Using Machine Learning	K-Nearest Neighbor, Random Classifier, Correlation, SVM	The author has used the Gini indicator system along with decision tree and hyperplane in SVN algorithm and it shows major gain in characteristics and visualize powerful representation of the decision tree algorithm		Among all other classification algorithms in machine learning, SVM provides better prediction accuracy.
Keshav Srivastava, Dilip Kumar Choubey(2020)	Heart Disease Prediction using Machine Learning and Data Mining	Decision Trees, KNN, Naïve Bayes, Random Forest, SVM.	A web application is erected utilizing a beaker, and these collections are utilized to build prognostications grounded on the information offered by the stoner. Unborn experimenters can intensify their delicacy by engaging KDD ways to recoup retired information from samples.		In their trial, they used th dataset of Cleveland Heart Disease from the UCI depository to initialize information with unknown merits and utilized methodologies related to Decision Tree, K-Nearest Neighbour, Support Vector Machines, and Random Forest to gain delicacy of 79%, 87%, and 83%, independently.

Brief about Literature survey

Architectural Diagram



Algorithmic Survey

There are a few of the methodologies that we have linked for gaining better accurate results relating to the disquisition of CAD.

The basic difference of algorithms:

Parameters	K-Nearest Neighbors	Decision tree	Naive Bayes	Random Forest
Definition	K-nearest neighbors (KNN) algorithm is a type of supervised ML algorithm that can be used for both classifications as well as regression predictive problems	Decision Trees are a supervised Machine Literacy system in which data is continually divided according to a set of rules.	Naive Bayes classifiers are a collection of classification algorithms grounded on Bayes'Theorem. It isn't a single algorithm but a family of algorithms where all of them partake in a common principle, i.e. every brace of features being classified is independent of each other.	Random Forest is an extractor that holds several decision trees on distinct subsets of a dataset and parses them to increase the dataset's vatication delicacy.
Accuracy	Gives less accurate results	Gives less accurate results.	Accuracy depends on the type of data. If data is clean then It gives more accurate results.	Gives more accurate results.
Speed	Slow	Faster	Very fast	Faster than Decision tree and K means

Algorithms	K-Nearest Neighbors	Decision tree	Naive Bayes	Random Forest
Algorithmic Step	1. The first step is to load training and test data. 2. Next, we need to choose the value of K i.e. the nearest data points. K can be any integer. 3. Now for every tested point, we need to implement the following steps- • Using Euclidean distance calculate the distance between test data and each row of training data. • The next step is to sort distance values in ascending order. • Now, select the top K row from the sorted array, • The next step is assigning the class to test points from regular classes of rows.	1. S starts the section with the root-knot, which involves the entire information set. 2. Utilizing the Trait Choice estimate, find the stylish trait in the dataset (ASM). 3. Bisect the S into subgroups that involve the stylish trait's implicit merits. 3. Build a knot in a decision tree that grasps the stylish trait. 4. Recursively build new decision trees using the subgroups of the information set fabricated in step-3. 5. Continue this procedure until you can no longer classify the bumps any farther and relate to the last knot as a splint knot.	Step 1: Separate By Class. Step 2: Summarize Dataset. Step 3: Summarize Data By Class. Step 4: Gaussian Probability Density Function. Step 5: Class Probabilities.	1. Choose K data points arbitrarily from the training set. Produce decision trees for the data points you've chosen (Subsets). 2. Choose N for the number of decision trees you wish to produce. 3. Way 1 and 2 should be repeated. 4. Find the validations of each decision tree for new data points, also allocate the new data points to the order with the most votes.

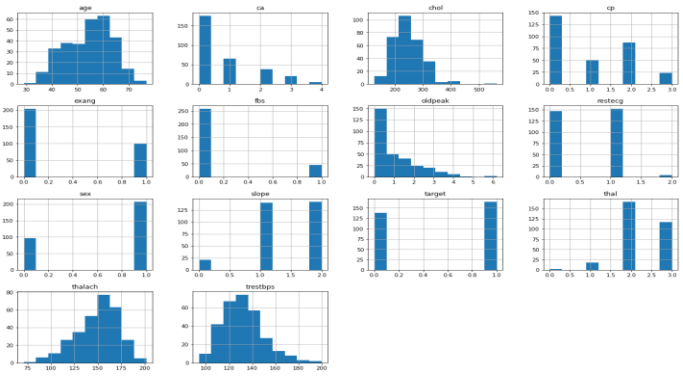
Source for the dataset:

https://data.world/informatics-edu/heart-disease-prediction

https://www.kaggle.com/ronitf/heart-disease-uci

Here's a histogram representation of the attributes given in the dataset

ISSN: 0975-3583, 0976-2833 VOL 13, ISSUE 05, 2022



Data representation of some of the attributes

Main attributes:

Age	Resting BP
Cholesterol level	Fasting Blood Sugar
Slope	Resting ECG
Calcium score	Maximum heart rate
Gender	Exercise Angina
Target	Old Peak
Chest Pain Type	

Brief about all attributes:

Attributes	**Data Type**	**Effect of attribute on heart**
Age	Int	Aged people aged 65 and aged are at lesser threat of heart attack than youngish people, videlicet heart problems, blood vessels, or both. Aging can beget changes in the heart and blood vessels that may increase a person's threat of developing cardiovascular complaints.
Cholesterol level	String	There was a significant positive correlation between heart rate and systolic and diastolic blood pressure, blood cholesterol situations, triglycerides, and blood sugar, as well as significant contrasts between heart rate and HDL-cholesterol situations, indeed after compliance. and the confusing aspects below.
Slope	Float	Pitch member changes related to the constituents that ensure cardiovascular exercise, and ST/ heart rate (voice position), has been suggested as a more accurate ECG determinant to diagnose significant coronary thruway (CAD)

ISSN: 0975-3583, 0976-2833 VOL 13, ISSUE 05, 2022

		complaints.
Calcium score	Int, float	Calcium- score netting heart test (coronary calcium test) uses motorized tomography (CT) to check the calcium deposit in the major highways of your heart. Bettered coronary calcium situations suggest that you have an increased threat of severe coronary roadway complaint and an increased threat of a preterm heart attack.
Gender	Character	Men frequently develop CVD at an early age and have an advanced threat of developing heart complaints (CHD) than women. Women, by the discrepancy, are more likely to have strokes, which frequently do latterly in life.
Target	Character	The target of the supervised model is a special type of point. The target column for training data contains the factual values used for model training. The column aimed at test data contains factual values for which the vaticination is compared.
Chest pain Type	Float	Chest pain is the most common presenting complaint of acute myocardial infarction. The classic incarnation of ischemia is generally described as heavy casket pressure or squeezing, a " burning" feeling, or difficulty in breathing. The discomfort or pain frequently radiates to the left shoulder, neck, or arm.
Resting BP	Float	The heart is good at performing under pressure. It constantly pumps blood throughout your blood vessels, furnishing every part of your body with the oxygen you need to survive. Blood pressure measures the force of your blood pushing against the walls of your blood vessels
Fasting Blood	int, float	A fasting blood sugar position of lower than 100

ISSN: 0975-3583, 0976-2833 VOL 13, ISSUE 05, 2022

Sugar		mg/ dL (5.6 mmol/ L) is normal. A fasting blood sugar position from 100 to 125 mg/ dL (5.6 to6.9 mmol/ L) is considered prediabetes.However, you have diabetes If it's 126 mg/ dL (7 mmol/ L) or advanced on two separate tests.
Resting ECG	int	The resting phase is also known as Phase 4. The membrane eventuality is at-90 mV. Phase 0 fast sodium channels open and sodium flows into the cell (depolarizes). This results in a rapid-fire upstroke.
Maximum heart rate	in	The traditional system, also known as HRmax, is a simple way of gauging your maximum heart rate. Start by abating your age from 220. Also, use the result to calculate your range. For illustration, if you are 50 times old, the computation would be 220-50 = 170 (HRmax).
Exercise Angina	int, float	Angina is a pain in the chest that comes on with exercise, stress or other effects that make the heart work harder. It's an extremely common
		symptom of coronary roadway complaint, which is caused by cholesterol-congested coronary
Old Peak	int, float	ST depression convinced by exercise relative to rest

After Implementing we got the following results for accuracy
Accuracy of K nearest neighbor algorithm: 68%
Score of K nearest neighbor algorithm: 69%

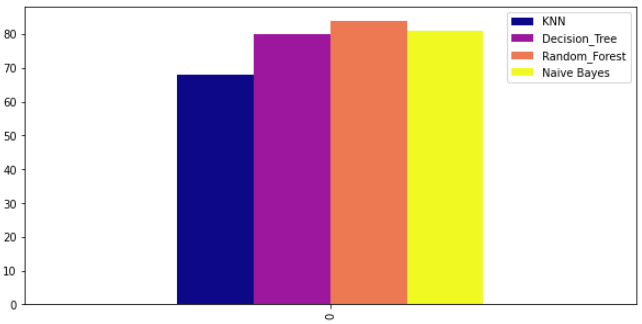
Accuracy of Random Forest algorithm: 84%
Score of Random Forest algorithm: 83%

Accuracy of Naive Bayes algorithm: 81%
Score of Naive Bayes algorithm: 81%

Accuracy of Decision tree algorithm: 80%
Score of Decision tree algorithm: 80%

Accuracy Comparison table and graph:

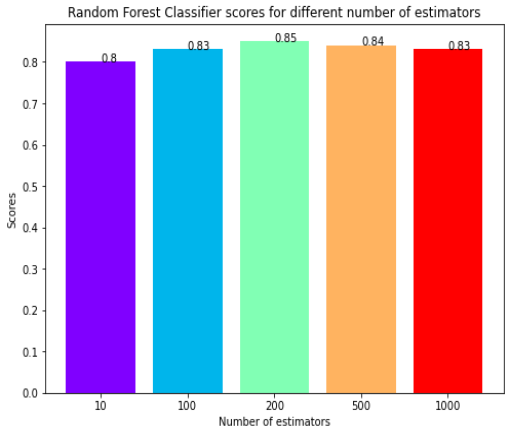
KNN	Decision_Tree	Random_Forest	Naive Bayes
68.0	80.0	84.0	81.0



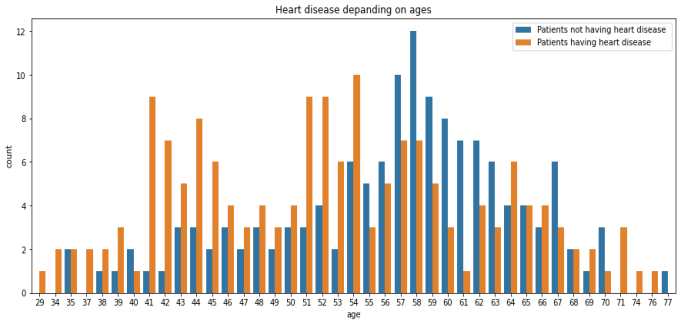
ISSN: 0975-3583, 0976-2833 VOL 13, ISSUE 05, 2022

Looking at the result we found that among 4 algorithms, the Random Forest algorithm offers the highest accuracy which we improved after applying the following methods.

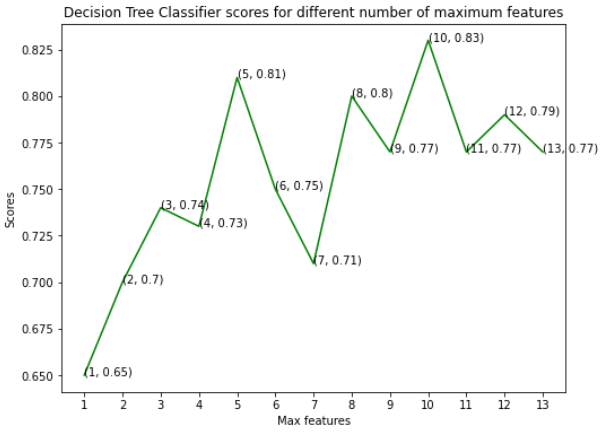
1. Use additional data (high quality) and feature engineering
2. Tune the algorithm hyper parameters
3. Try different algorithms.



Comparison bar graph between patients having/not having heart disease

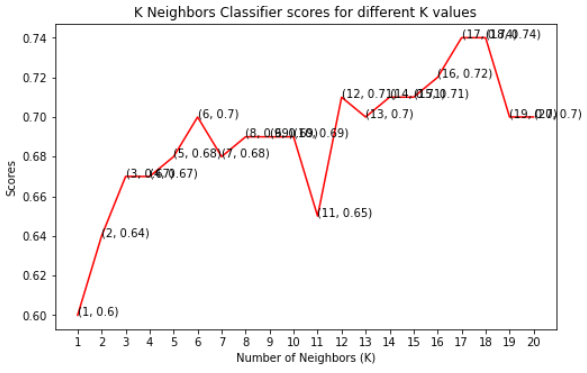


3. Decision tree



Graphs of the used algorithm:

1. K nearest neighbor:



2. Random Forest

Conclusion:-

After reviewing the literature, we plant that among all the methods available in the literacy and data mining industry, 100% random forest, 88.5% nearby neighbor K and 86% Naive Bayes algorithm has proven to be the best. algorithms are very important and effective in determining the threat of a cardiovascular condition. While working we found the accuracy of all 4 algorithms as 69%, 46%, 81%, and 80%. Further in the project, we will try to improve the accuracy of the given algorithms using various methods such as using additional (high-quality) data and feature engineering. adjusting the algorithm hyper parameters and developing an accurate system that can be used to predict the risk level of Cardiovascular Artery Disease in a patient quickly and easily.

ISSN: 0975-3583, 0976-2833 VOL 13, ISSUE 05, 2022

REFERENCES:

1. Hadia Amin, Abita Devi, Nida UI Amin, "Predictive Analysis of Heart Disease using K-Means and Apriori Algorithms", Journal of Applied science and Computations, ISSN NO:1076-5131, Aug 2019

2.Zerina Masetic, AbdulHamit Subasi, "Congestive heart failure detection using random forest classifier", Computer Methods and Programs in Biomedicine, Volume 130, July 2016.

3.K.Vembandasamy, R.Sasipriya*PP*and E.Deepa, "Heart Disease detection using Naive Bayes Algorithm", IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 9, September 2015.

4.Dilip Kumar Choubey, Keshav Srivastava, "Detection of Heart Disease using Machine Learning Techniques", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-9 Issue-1, May 2020

5. Edward Choi, 欽 recursively create new decision trendy Schuetz, Walter F Stewart, Jimeng Sun "Using recurrent neural network models for early detection of heart failure onset", Journal of the American Medical Informatics Association, Volume 24, Issue 2, March 2017

6. Vishal Dineshkumar Soni, "Detection Of Heart Disease Using Machine Learning Techniques", INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 9, ISSUE 08, AUGUST 2020, ISSN 2277-8616

7. Rajesh N, T Maneesha, Shaik Hafeez, Hari Krishna, "Prediction of Heart Disease using Machine Learning Algorithm", *International Journal of Engineering & Technology, 7 (2.32) (2018) 363-366*

8. Chithambaram T, Logesh Kannan N, Gowsalya M, "Heart Disease Detection using Machine Learning", DOI: https://doi.org/10.21203/rs.3.rs-97004/v1