Heart Disease Probability Test Using Classification and Regression Tree Model

[1]Dr.A.Selvakumar, [2]Dr.P.Sudha and [3] Dr.K.Jayaprakash

Assistant Professor, Department of Computer Science, Sree Saraswathi Thyagaraja College, Pollachi, Coimbatore,India.

Assistant Professor, Department of Computer Science, Sree Saraswathi Thyagaraja College, Pollachi, Coimbatore,India.

Associate Professor, Department of Commerce, Sree Saraswathi Thyagaraja College, Pollachi, Coimbatore, India.

E-mail:[1]arshariniselva@gmail.com, [2]sudha.sabariananth@gmail.com & [3]dazzlejai@gmail.com

ABSTRACT

To forecast the diagnosis of the heart disease with minimum attributes using data mining techniques is the main objective of this research. The heart diseases which affect the heart cause decreased blood flow. Data Mining is an important technique has developed in medical information to analyze the large volume of data into trends and patterns. A large number of algorithms are used to efficiently diagnose the medical data set. To forecast heart diseases, Classification and Regression Tree (CART) can be used. IBM SPSS v. 23 decision trees empower you to recognize groups, find out the relationships between them and forecast upcoming events. The decision trees can help you present categorical results and more evidently make clear analyses to the general public. The accuracy of the algorithms is measured in terms of performance and the function is analyzed and the results are calculated.

Keywords : Heart diseases, decision tree, SPSS, CART, Data Mining

DOI Number: 10.14704/nq.2022.20.9.NQ44431 **Neuro Quantology 2022; 20(9):3766-3772**

1. INTRODUCTION

The health care services provide many options for consumer's health care through rehabilitative, emergency caretaking, prevention care, hospitality, diagnoses, home care, endurance disease care and palliative guidelines. The main aim of disease prevention is to lessen the difficulty of diseases and related risk factors it sources to understand as precise, population and individual-based interventions for major and minor prevention. The importance of disease prevention is that an individual or group engages in action to reduce or eliminate the increased risks of illness, disability, and premature death. The main types of heart disease are TA = Typical Angina, ATA = Atypical Angina, NAP = Non anginal pain and ASY = Asymptomatic. Typical (classic) angina pectoris (1) Abdominal chest pain or discomfort (2) is triggered by labor or emotional stress and (3) is relieved by rest or nitroglycerin (or both). Different (possible) angina applies to chest pain when 2 of the 3 criteria of classic angina are present. Angina Chest pain is a feeling of pressure or compression that usually occurs when your heart muscle does not get enough oxygenated blood. Non-cardiac chest pain is a term used to describe chest pain that is not caused by heart disease or heart attack. In most people, non-cardiac chest pain is associated with gastroesophageal reflux disease. If asymptomatic there are no symptoms. It is caused by recovering from a disease or condition and no longer having symptoms and having an illness or condition (such as early stage hypertension or glaucoma), but no symptoms. The rise of high-performance computing has benefited many sectors to find solutions that work for their problems. In cardiovascular research, data mining techniques play a vital role. Starting from the various interpretations between healthy people and heart patients in the existing medical data is an appreciative and excellent approach in the study of heart disease classification to find hidden clinical information. The classification of heart disease contributes to the risk factor for the treatment of patients. Two main approaches, statistical and mechanical learning, are used to predict the importance of heart disease based on the expression of clinical data.

2. RELATED WORKS

A new CART algorithm has been introduced by Al-batah, M. S. (2014) to test the probability of heart disease, which plays a key role in enhancing patient outcomes and reducing drug costs. Chaithra (2020) A new classification and regression tree (CART) algorithm for predicting ischemic heart disease, successfully showing a patient's ischemic heart disease for predicted outcomes. Early detection of heart disease can increase survival rates for diagnosing heart disease using data processing techniques. The machine learning methodology supervised by Anitha, S., & Sridevi, N. (2019) was discussed. Neural Networks, Decision tree and Naïve Bayes algorithms has been developed by Dangare, C. S., & Apte, S. S. (2012) made the most accurate prediction of heart disease. Decision trees, Naïve Bayes and Neural Network were introduced by Sellappan Palaniappan (2008) and others because the results show that each technique has its unique strength to achieve the goals of the defined mining objectives. Poornima Singh (2018) established an efficient cardiovascular prognosis system that uses neural network data processing techniques to predict heart risk status. The results obtained show that a well-designed diagnostic system can effectively predict the risk of heart disease. The K-mean clustering and MAFIA algorithm was established by Kajal, E. S., & Nishika, M. (2016) for the heart disease prognosis system and achieved 89% accuracy. The random tree algorithm was discussed by Mirpouya Mirmozaffari (2017) to provide maximum accuracy for heart disease prediction and minimize errors in the best performance algorithm. Early Heart Disease Prediction Using data mining techniques has been developed by Methaila, A., Kansal, P., Arya, H., & Kumar, P. (2014). Weighing association and the MAFIA algorithm.

3. CLASSIFICATION AND REGRESSION TREES (CART) ALGORITHM

Classification and regression trees (CART) analysis creates a tree with each node R using the binary rotation partition because the partition cell and N, M, O represent the leaf or terminal nodes, after which this section represents a further segment of the data. As shown in Figure. 1 the variation associated with the description of Y is not adequately described. In descending order of priority, the graph is taken as the domain of all variables associated with our Y. The word binary means that each group of patients referred to by a "node" in a decision tree can only be separated by two groups. Therefore, each individual node can be spitted into two child nodes, where the original node is considered the superior node. Thus, each individual superior node can be resurrected into two subordinate nodes, and they each separate and produce additional subordinates. The term "partition" specifically refers to the set of data that is divided or segmented.

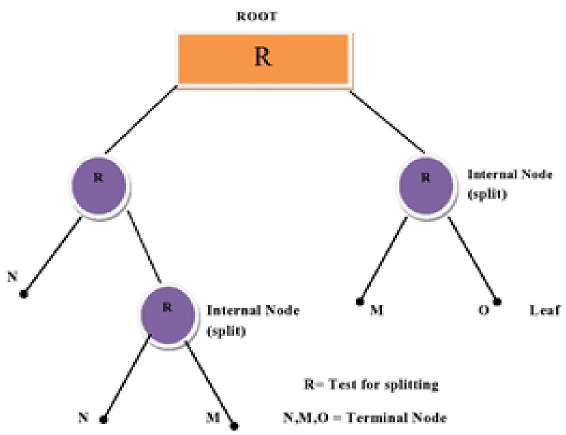

Figure : 1 CART Structure

4. DATA COLLECTION

For the experimental consideration, Heart disease dataset is taken from the following source https://www.kaggle.com/jillanisofttech/heartdisease-%20prediction-using-differenttechniques

CART classification has been used to forecast the type of chest pain among heart patients. The dataset comprises of 12 attributes namely Age, Sex, ChestPainType, RestingBP, Cholesterol, FastingBS, Resting ECG, MaxHR, ExerciseAngina, Oldpeak, ST_Slope and HeartDisease. The database consists of 917 cases. Table 1 lists the name of the attributes with its type and description.

Table 1.Heart disease data set attributes

Name	Type	Description
Age	Continuous	Age of the patients
Sex	Discrete	1 = male, 0 = female
ChestPainType	Discrete	Chest pain type: 1 = typical angina, 2 = atypical angina 3 = non-anginal pain, 4 = asymptomatic
RestingBP	Continuous	Resting blood pressure (in mm Hg)
Cholesterol	Continuous	Serum cholesterol (in mg/dl)
FastingBS	Continuous	Fasting blood sugar > 120 mg/dl: 1 = true, 0 = false
RestingECG	Discrete	Resting electrocardiographic results: 0 = normal 1 = having ST-T wave abnormality 2 = showing probable or definite left ventricular hypertrophy by Estes' criteria
MaxHR	Continuous	Maximum heart rate achieved
ExerciseAngina	Discrete	Exercise induced angina: 1 = yes, 0 = no
Oldpeak	Continuous	ST depression induced by exercise relative to rest
ST_Slope	Discrete	The slope of the peak exercise ST segment: 1 = up sloping, 2 = flat, 3 = down sloping
HeartDisease	Continuous	Heart Disease 1= yes, 0 = no

5. RESULTS AND DISCUSSION

Decision tree (Figure. 2) is a graph that uses a branching technique to demonstrate all the possible outcome of a decision. It can be drawn manually or with the help of specialized software or graphics program. Informally, when a panel makes a decision, decision trees can be very effective in structuring the discussion. Deliberately, they can be used to allocate funds /

time or other values to obtain potential results so that results can be automated. The results of the study have been statistically attained by using IBM SPSS v. 23 from heart patients' database. CART (Classification and Regression Trees) algorithm was applied in the decision tree construction. Generally CART divides a given data into different parts, which are as identical as possible depending on the dependent variable. A terminal that has the same value for all cases of the dependent variable is a homogeneous, "pure" terminal.

A decision tree to predict chest pain is developed by taking patients' chest pain type as the dependent variable (i.e., Typical Angina, Atypical Angina, Non-Anginal Pain, Asymptomatic), and the following nine factors related to the chest pain such as Gender, RestingBP, Cholesterol, FastingBS, RestingECG, MaxHR, ExerciseAngina, HeartDisease and Age of the patient are treated as the independent variables. Out of 917 valid patient chest pain data, testing data set was used to corroborate the effectiveness of the proposed research model.

CART (Classification and Regression Trees) is a regression based decision tree that begins with base node 0. It replicates the dependent variable, chest pain type of the heart patients. Overall, 54.1% of the samples have asymptomatic type of chest pain, 22% have Non-Anginal Pain, 18.9% have Atypical Angina and 5% have Typical Angina. The CART algorithm realized the primary partition through the element "Heart disease". Node 1 is obtained, which contains 36.3% of patients with an A typical Angina type of chest pain range to this category. From node 1, the algorithm executes a new hierarchization, whose most important element is "ExerciseAngina". It means that this element will be the most significant variable within the group of independent variables. It's also reinforces the significance of this element achieved in the first level of splitting and also nodes 3 and 4wereattained. Node 3 highlights the category of Atypical Angina type of chest pain. Node 4 highlights the category of asymptomatic type of chest pain, causing this node not to branch any additional nodes. From node 2, the algorithm executes a new hierarchization, whose foremost element is "ExerciseAngina". it means that this element will be the most significant variable within the group of independent variables. It's also reinforces the significance of this element achieved in the second splitting and nodes 5 and 6 were attained. Node 5 and 6 highlights the category of asymptomatic type of chest pain. From node 6the algorithm couldn't perform any further branch. From node 5, the algorithm executes one more hierarchization, whose powerful element is "Cholesterol". It means that this component will be the most significant variable within the group of independent variables. It's also reinforces the significance of this element accomplished in the second splitting and also nodes 9 and 10 are attained. Node 9 and 10 highlights the category of asymptomatic type of chest pain, causing these nodes not to branch any auxiliary nodes. From node 3, the algorithm accomplishes a new hierarchization, whose prominent factor is "RestingECG". It means that these factors will be the most significant variable within the group of independent variables. It's also emphasizes the significance of this element achieved in the second splitting and nodes 7 and 8 are attained. Node 7 highlights the category of Atypical Angina type of chest pain. Node 8 highlights the category of Non-Anginal Pain, causing this node not to branch any auxiliary nodes. From node 7, the algorithm implements one more hierarchization, whose significant element is "MaxHR". It means that this element will be the most significant variable within the group of independent variables. It's also emphasizes the significance of this element accomplished in the third splitting and nodes 11 and 12 were attained. Node 11 and 12 emphasizes the category of Atypical Angina type of chest pain. From node 12the algorithm couldn't perform any further branch. From node 11 the algorithm executes one more hierarchization, whose significant element is "Cholesterol". It means that these elements will be the most significant variable within the group of independent variables of the research work. It's also emphasizes the significance of this element accomplished in the fourth splitting and nodes 13 and 14 were attained. Node 13 emphasizes the category of Non- Anginal Pain and Node 14 emphasizes the category of asymptomatic type of chest pain, causing these nodes not to branch any further.

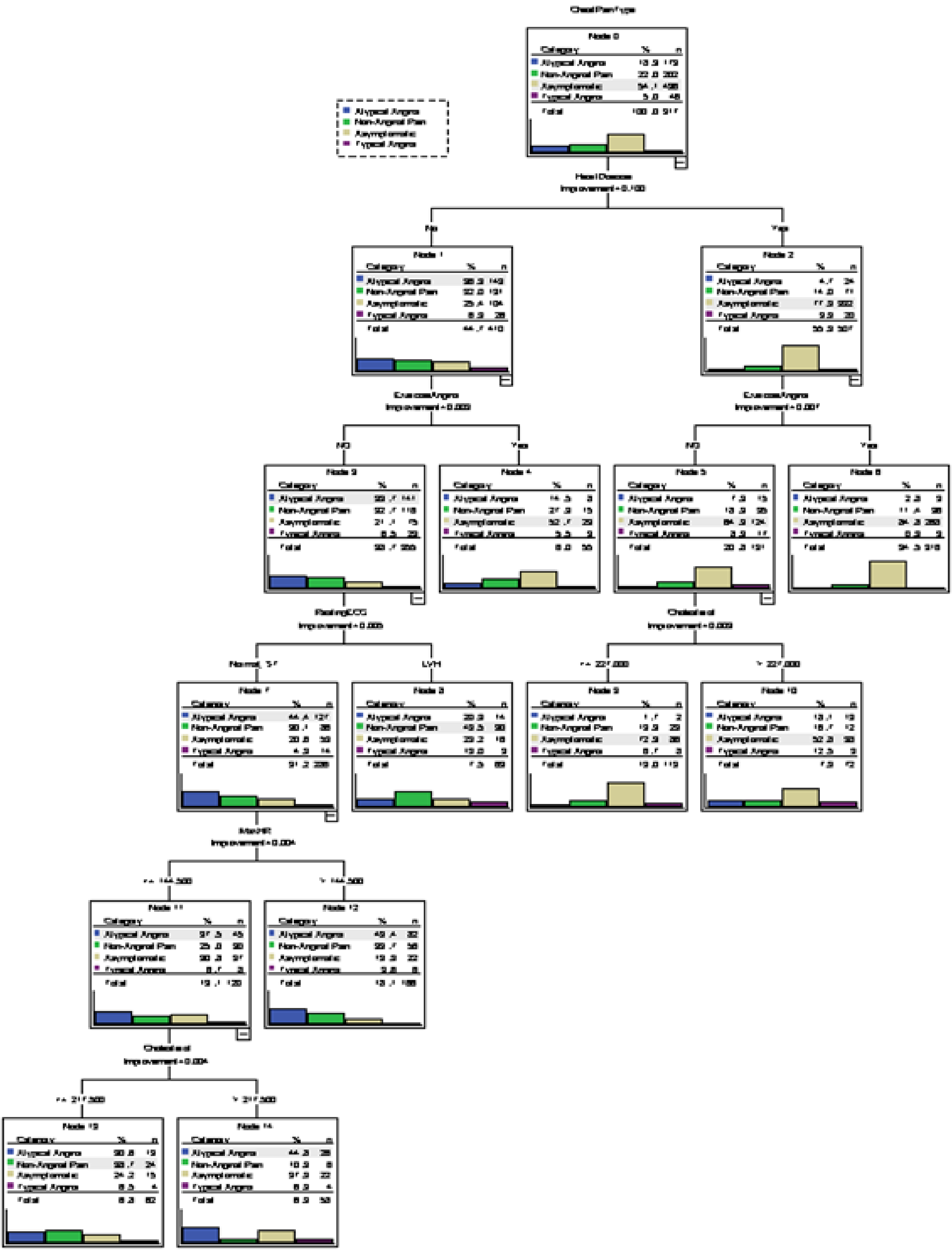

Figure. 2. CART classification tree for test samples

Table 2. Tree table for test samples

Node	Atypical Angina		Non-Anginal Pain		Asymptomatic		Typical Angina		Total		Predicted Category	Parent Node	Primary Independent Variable		
	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent			Variable	Improvement	Split Values
0	173	18.9%	202	22.0%	496	54.1%	46	5.0%	917	100.0%	Asymptomatic				
1	149	36.3%	131	32.0%	104	25.4%	26	6.3%	410	44.7%	Atypical Angina	0	HeartDisease	.100	No
2	24	4.7%	71	14.0%	392	77.3%	20	3.9%	507	55.3%	Asymptomatic	0	HeartDisease	.100	Yes
3	141	39.7%	116	32.7%	75	21.1%	23	6.5%	355	38.7%	Atypical Angina	1	ExerciseAngina	.009	NO
4	8	14.5%	15	27.3%	29	52.7%	3	5.5%	55	6.0%	Asymptomatic	1	ExerciseAngina	.009	Yes
5	15	7.9%	35	18.3%	124	64.9%	17	8.9%	191	20.8%	Asymptomatic	2	ExerciseAngina	.007	NO
6	9	2.8%	36	11.4%	268	84.8%	3	0.9%	316	34.5%	Asymptomatic	2	ExerciseAngina	.007	Yes
7	127	44.4%	86	30.1%	59	20.6%	14	4.9%	286	31.2%	Atypical Angina	3	RestingECG	.005	Normal; ST
8	14	20.3%	30	43.5%	16	23.2%	9	13.0%	69	7.5%	Non-Anginal Pain	3	RestingECG	.005	LVH
9	2	1.7%	23	19.3%	86	72.3%	8	6.7%	119	13.0%	Asymptomatic	5	Cholesterol	.003	<= 227.000
10	13	18.1%	12	16.7%	38	52.8%	9	12.5%	72	7.9%	Asymptomatic	5	Cholesterol	.003	> 227.000
11	45	37.5%	30	25.0%	37	30.8%	8	6.7%	120	13.1%	Atypical Angina	7	MaxHR	.004	<= 144.500
12	82	49.4%	56	33.7%	22	13.3%	6	3.6%	166	18.1%	Atypical Angina	7	MaxHR	.004	> 144.500
13	19	30.6%	24	38.7%	15	24.2%	4	6.5%	62	6.8%	Non-Anginal Pain	11	Cholesterol	.004	<= 217.500
14	26	44.8%	6	10.3%	22	37.9%	4	6.9%	58	6.3%	Atypical Angina	11	Cholesterol	.004	> 217.500

Growing Method : CRT   Dependent Variable : ChestPainType

6. CONCLUSION

In medicine, an important task is to diagnose heart disease. This work investigates applying the CART model to help healthcare professionals in this task. A data set of 917 cases and twelve attributes were selected as the input variables of the model. Based on the experimental database, the asymptomatic chest type cases occurs excess of cholesterol component. Therefore, the CART model gives significantly better results in predicting heart disease. This kind of research can play a significant role in enhancing patient outcomes, reducing medical fees, and further advanced clinical research. The future work will be additionally improved and developed for the automation of heart disease prognosis.

REFERENCES

[1] Al-batah, M. S. (2014). Testing the probability of heart disease using classification and regression tree model. Annual Research & Review in Biology, 1713-1725.

[2] Chaithra (2020). Classification and Regression Tree (CART) algorithm for the prediction of ischemic heart disease, International Journal of Future Generation Communication and Networking, 13(4), 196–206.

[3] Anitha, S., & Sridevi, N. (2019). Heart disease prediction using data mining techniques. Journal of analysis and Computation.

[4] Dangare, C. S., & Apte, S. S. (2012). Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications, 47(10), 44-48.

[5] Sellappan Palaniappan (2008). Intelligent heart disease prediction system using data mining techniques, IJCSNS International Journal of Computer Science and Network Security, 8(8), 343 – 350.

[6] Poornima Singh (2018). Effective heart disease prediction system using data mining techniques, International Journal of Nano medicine,121- 124.

[7] Kajal, E. S., & Nishika, M. (2016). Prediction of Heart Disease using Data Mining Techniques. International Journal of

Advance Research, Ideas and Innovations in Technology, 2(3), 1-7.

[8] Mirpouya Mirmozaffari (2017). Data mining classification algorithms for heart disease prediction", International Journal of Computing, Communications & Instrumentation Engg. (IJCCIE), 4(1), 11-15.

[9] Methaila, A., Kansal, P., Arya, H., & Kumar, P. (2014). Early heart disease prediction using data mining techniques. Computer Science & Information Technology Journal, 24,53-59.