WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Heart Disease Prediction: Using Machine Learning

Ankit Jaiswal[1], Rahnuma Siddique[2], Arvind Kumar[3], Akhilesh Kumar[4]

Computer Science & Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad
ankitjaiswal1436@gmail.com[1], rahnumashaikh1650@gmail.com[2], smsmps728@gmail.com[3], akhileshkumar2003up@gmail.com[4]

ABSTRACT—

Heart and cardiac disease is a major cause of death and mortality in the world. Early prediction and diagnosis of cardiac disease can lead to better treatment and outcomes for patients. This paper presents the development of a heart disease prediction system using machine learning techniques, which can help healthcare workers in identifying patients who has at high chances of getting a heart disease. The proposed system uses the Heart Disease Dataset, which contains information on 917 patients, and applies machine learning algorithm: Decision tree to predict the chances of heart disease. The results showed that the Decision tree algorithm performed the best, with an accuracy of 96%. The most important predictors of heart disease in our model were age of a person, maximum heart rate achieved by a person, chest pain type felt by a person, and exercise-induced angina.

Keywords—Heartdisease prediction system, machine learning algorithms, Decision tree algorithms, Heart Disease Dataset available. techniques and parameter tuning could improve Decision Tree's efficiency.

I. INTRODUCTION

The leading cause of death worldwide is heart diseases, with an estimated 17 million deaths in 2019 annually (National Library of Medicine). Early prediction and diagnosis of heart disease can lead to better treatment and outcomes for patients that can be achieved by Machine learning techniques.

In this paper, we present the development of a heart disease prediction system using machine learning techniques. The system is trained on the Heart Disease Dataset, which contains information on 917 patients. We used machine learning algorithms to train and test our model, namely Decision tree algorithm.

II RELATED WORK

Many studies have been conducted by using machine learning algorithms for the classification of heart disease.

• Naive Bayes classifier proposed by Miranda et al. that considered factors that are important for predicting cardiovascular diseases.

• Avinash Go lande et al. compared the accuracy of Decision Tree, KNN, and K-Means algorithms for heart disease classification and suggested that combining different classified, and it only uses one attribute at a time which can be a hectic solution for some.

• Theresa Princy R et al. conducted a survey of classification algorithms for predicting heart disease, including Naive Bayes, KNN, Decision Tree, and Neural Network, and analyzed the classifiers' accuracy for varying numbers of attributes.

III. DATA RESOURCE

This dataset consists of 12 attributes and 917 datapoints.

There will be

(A) Categorical attributes = 7

(B) Numeric attributes = 5

Attribute	Description	Range
Age	Person's age in years	28-77
Sex	Person's gender (1-M 0-F)	1,0
Chest Pain Type	Chest Pain Type	1,2,3,4
Resting Bp	Resting BP	0-200
Cholesterol	Serum cholesterol in mg/dl	0-603
Fasting BS	Fasting blood sugar in mg/dl	0,1
Resting ECG	Resting Electro cardio graphic results	0,1,2
MaxHR	Maximum heart rate achieved	60-202
Exercise Angina	Exercise Induced Angina	0,1
Old peak	ST depression induced by exercise relative to rest	-2 – 6
ST Slope	Slope of the Peak Exercise ST segment	1,2,3
Heart Disease	Class Attribute	0,1

The dataset that we used is shown in the table below.

Patients aged between 28 to 77 are selected in this dataset. In this study, gender values of 1 and 0 represent male and female patients, respectively. Among the four types of chest pain, Type 1 angina is strongly linked to heart disease, caused by narrowed coronary arteries which reduces flow of blood in muscles of our heart. It typically occurs during mental or emotional stress. Nonangina chest pain, on the other hand, may be caused by various factors other than heart disease. The fourth type, Asymptomatic, does not necessarily indicate the presence of heart disease. The next attribute Resting BP gives us resting blood pressure of a person. Cholesterol is level of cholesterol. Fasting blood sugar level is represented by Fasting BS which is very essential. 1 is the value if Fasting BS is below 120mg/dl and value of Fasting BS is 0 if it is above 120mg/dl.
Resting ECG is the resting electrocardiographic result, Maximum heart rate is given by Max HR, Exercise Angina is the exercise induced angina whose value would be 1 if there is pain and if there is no pain experienced by patient then it would be 0. Old peak is the ST depression induced by exercise, ST slope is the slope of the peak exercise ST segment, Attribute of Heart Disease which is the final result column has a value of 0 which represents normal/fine and 1 if patient is diagnosed with a heart disease, then its value would be assigned=1.

IV. APPROACH METHODOLOGY

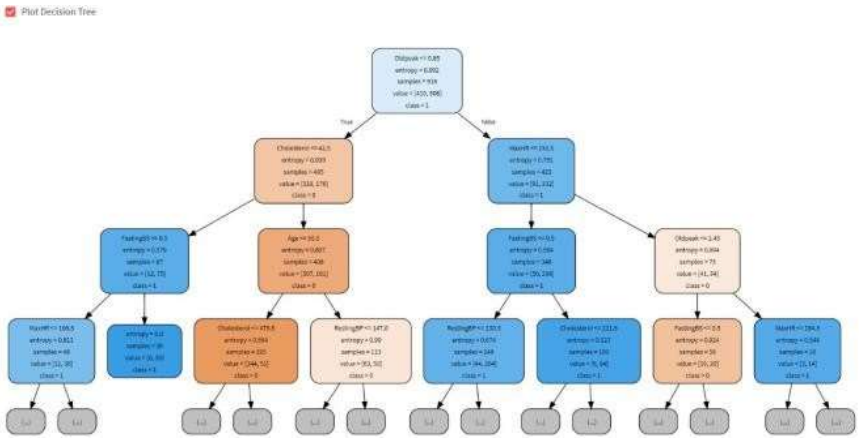
Decision Tree

The decision tree algorithm is a technique used for classification of the dataset provided, which can be used for both numerical and categorical data. It involves constructing a structure which looks like a tree and helps the user to easily analyse the data and work on it. The algorithm works by dividing the data into multiple subsets based on the most important predictors. The attribute entropy is calculated for each feature, and the data are split based on the attribute with the highest information gain or the lowest entropy.

The easiness of this algorithm/approach makes it the ideal option and its high accuracy provides an upper-hand for this algorithm as compared to other machine learning algorithms. However, a potential drawback is that the data may be over-

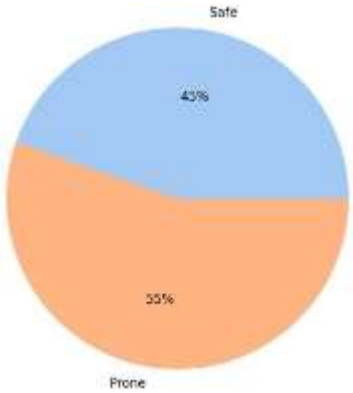
The Decision Tree algorithm yields easily interpretable results. Its accuracy is higher than that of other algorithms since it visualizes the data in a tree-like structure.

Nonetheless, there is a risk of over-classification and the algorithm only tests one attribute at a time for decision- making



V. CONCLUSION

In this paper, we presented a solution for prediction of heart disease by using one of the most efficient machine learning algorithm named Decision tree Classifier. We used decision tree classifier algorithm we used some inbuilt functions of sklearn library to train the data and return the model score. The overall result generated is shown in the form of piechart.



VI. FUTURE WORK

The future scope of this includes several aspects, including technology advancement, patient outcomes, and healthcare efficiency.

One possible area of research could be to investigate the effectiveness of using AI-powered algorithms to provide personalized recommendations and treatments plans. This could involve analyzing patient data and assessing the impact of using technology on patient outcomes, such as reducing hospital readmissions, improving medication adherence, and managing chronic conditions more efficiently.

Another area of research could be to examine the patient experience and satisfaction with using technology to access doctor's contact information and diet plans. This could involve conducting surveys or focus groups with patients to understand their preferences, concerns and feedback on using digital platforms for healthcare.

Furthermore, the future scope of this topic could involve exploring the role of healthcare providers in implementing and using technology to provide personalized care. This could involve examining the barriers and facilitators to technology adoption among healthcare providers and identifying strategies to support their use of technology in clinical practice.

REFERENCES

[1] Animesh Hazra, Arkomita Mukherjee, Amit Gupta, Asmita Mukherjee, "Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Research Gate Publications, July 2017, pp.2137-2159.

[2] V. Krishnaiah, G. Narsimha, N. Subhash Chandra, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review", International Journal of Computer Applications, February 2016.

[3] Guizhou Hu, Martin M. Root, "Building Prediction Models for Coronary Heart Disease by Synthesizing Multiple Longitudinal Research Findings", European Science of Cardiology, 10 May 2005.

[4] TMythili, Dev Mukherji, Nikita Padaila and Abhiram Naidu, "A Heart Disease Prediction Model using SVM- Decision Trees- Logistic Regression (SDL)", International Journal of Computer Applications, vol. 68, 16 April 2013.

[5] https://www.medicalnewstoday.com/articles/257484.php.

[6] Nimai Chand Das Adhikari, Arpana Alka, and Rajat Garg, "HPPS: Heart Problem Prediction System using Machine Learning".

[7] Heart disease CSV file from https://github.com/MainakRepositor/Cardiac-Arrest-Predictor

[8] C. Feng, S. Wu, and N. Liu, "A user-centric machine learning framework for cyber security operations center," in 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, China, Jul. 2017, pp. 173–175, doi: 10.1109/ISI.2017.8004902.

[9] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," Emerging artificial intelligence applications in computer engineering, vol. 160, no. 1, pp. 3–24, 2007.

[10] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine learning: a review of classification and combining techniques," ArtifIntell Rev, vol. 26, no. 3, pp. 159–190, Nov. 2006, doi: 10.1007/s10462-007-9052-3.

[11] C. Surv, M. N. Murty, P. J. Flynn, A. K. Jain, and P. J. Flynn, And. 1999.

[12] D. Sharma and N. Kumar, "A Review on Machine Learning Algorithms, Tasks and Applications," vol. 6, pp. 2278–1323, Oct. 2017.

[13] K. Pahwa and N. Agarwal, "Stock Market Analysis using Supervised Machine Learning," in 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMIT Con), Faridabad, India, Feb. 2019, pp. 197–200, doi: 10.1109/COMITCon.2019.8862225.

[14] M. Pérez-Ortiz, S. Jiménez-Fernández, P. A. Gutiérrez, E. Alexandre, C. Hervás-Martínez, and S. Salcedo-Sanz, "A Review of Classification Problems and Algorithms in Renewable Energy Applications," Energies, vol. 9, no. 8, Art. no. 8, Aug. 2016, doi: 10.3390/en9080607.

[15] Anuradha and G. Gupta, "A selfexplanatory review of decision tree classifiers," in International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), Jaipur, India, May 2014, pp. 1–7, doi: 10.1109/ICRAIE.2014.6909245.

[16] S. Patil and U. Kulkarni, "Accuracy Prediction for Distributed Decision Tree using Machine Learning approach," in 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Apr. 2019, pp. 1365–1371, doi: 10.1109/ICOEI.2019.8862580.

[17] N. S. Ahmed and M. H. Sadiq, "Clarify of the random forest algorithm in an educational field," in 2018 International Conference on Advanced Science and Engineering (ICOASE), 2018, pp. 179–184.

[18] D. Zeebaree, Gene Selection and Classification of Microarray Data Using Convolutional Neural Network. 2018.

[19] O. M. Salih Hassan, A. Mohsin Abdulazeez, and V. M. Tiryaki, "Gait-Based Human Gender Classification Using Lifting 5/3 Wavelet and Principal Component Analysis," in 2018 International Conference on Advanced Science and Engineering (ICOASE), Duhok, Oct. 2018, pp.

173–178, doi: 10.1109/ICOASE.2018.8548909.

[20] R. Zebari, A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed, "A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction," Journal of Applied Science

and Technology Trends, vol. 1, no. 2, pp. 56–70, 2020.