Heart Attack Prediction Using Machine Learning

Anand Mohanrao Magar, Dhanashri Rajput, Aryan V Dharmadhikari, Roshani D Dhembare, Tanish N Dhangar, Ayush S Dhangar, Manish S Dhane
Department of Engineering, Sciences and Humanities (DESH) Vishwakarma Institute of Technology, Pune, 411037, Maharashtra, India
F.Y.B.Tech Students' Engineering Design and Innovation (EDAI2)Project Paper, SEM 2A.Y.2022-23
Vishwakarma Institute of Technology, Pune, India

***Abstract*-In recent times, Heart-Disease prediction is one of the most complicated tasks in medical field. In the modern era, approximately one person dies per minute due to heart disease. As heart disease prediction is a complex task, there is a need to automate the prediction process to avoid risks associated with it and alert the patient well in advance. Data Science and machine learning (ML) can be very helpful in the prediction of heart attacks in which different risk factors like high blood pressure, high cholesterol, abnormal pulse rate, diabetes, etc... can be considered. The objective of this study is to optimize the prediction of heart disease using ML.**

***Keywords* - Disease prediction, Machine Learning, Performance measures, Classification Accuracy, *Data pre-processing, coronary, statistics*.**

1.INTRODUCTION

Human heart is the principal part of the human body. Basically, it regulates blood flow throughout our body. Any irregularity to heart can cause distress in other parts of body. Any sort of disturbance to normal functioning of the contemporary world, heart disease is one of the primary reasons for occurrence of most deaths. Heart disease may occur due to unhealthy lifestyle, smoking, alcohol and high intake of fat which may cause hypertension. Cardiovascular disease generally refers to narrowed or blocked blood vessels, which can also lead to heart attack, chest pain or stroke. In general, blood pressure, cholesterol and pulse rate are the main reasons for a heart attack. Heart attack is the main heart disease.

2.LITERATURE REVIEW

A lot of research is done in the field of heart disease prediction. Following work was referred in the study of the proposed system.

[1] First System, The goal of this study is to develop a machine learning (ML) model for heart disease prediction using the relevant parameters. For this research, a benchmark dataset of UCI Heart disease prediction is employed, which consists of 14 different heart disease-related factors. Machine learning algorithms including Random Forest, Support Vector Machine (SVM), Naive Bayes, and Decision Tree have been utilised to create the model. For the doctors in their clinic, using this model as a decision support system may be useful.

[2] In the Second System, The goal is to predict cardiac disorders using a variety of machine learning techniques, including artificial neural networks (ANN), decision trees, random forests, support vector machines, naive Baye (NB), and closest neighbour algorithms. Additionally, the results of these algorithms are compiled to forecast heart disease. These algorithms, such as the hybrid grid search algorithm and the random search algorithm, among others, were helpful when choosing the features to predict heart disease.

[3] The Third System focuses on a model that uses ML to forecast cardiac illness. Initially, ML was used to determine the degree of heart failure, but it was also employed to locate important characteristics that have a bearing on heart disorders by employing correlation approaches. Age, blood pressure, ejection fraction, salt creatinine, and other variables are some of those that contribute to heart disease. By using machine learning techniques, they suggested a strategy in this study for identifying significant traits. The goal of the effort was to build and develop feature ranking machine learning for heart disease prediction. Here, they created a multi-model system where they tested both deep learning and machine learning techniques like KNN Naive

Bayes. In this case, CNN's prediction of heart disease turned out to be extremely accurate.

[4] Forth system, This paper is about The dataset implemented for both training and testing purposes affects how accurate machine learning algorithms are. The accuracy of machine learning methods for predicting cardiac disease is therefore calculated in this research using a model. The algorithms used include k-nearest neighbour, decision trees, linear regression, and support vector machines. They used Python to achieve this since it includes a variety of libraries and header files that improve the accuracy and precision of the operation.

heart can be classified as a Heart disease. In today's contemporary world, heart disease is one of the primary

3.METHODOLOGY

A] Components-

We put machine learning to use to forecast cardiac disease. We employed two algorithms in this project: the first was a decision tree, and the second was a KNN. In addition, the visualisation tools plotly, seaborn, and matplot.lib have been employed. Later, the high-quality dataset was imported and divided into several entities, as illustrated in the diagram below.

#	Column	Non-Null Count	Dtype
0	age	303 non-null	int64
1	sex	303 non-null	int64
2	cp	303 non-null	int64
3	trestbps	303 non-null	int64
4	chol	303 non-null	int64
5	fbs	303 non-null	int64
6	restecg	303 non-null	int64
7	thalach	303 non-null	int64
8	exang	303 non-null	int64
9	oldpeak	303 non-null	float64
10	slope	303 non-null	int64
11	ca	303 non-null	int64
12	thal	303 non-null	int64
13	target	303 non-null	int64

Fig 1: Attributes that taken into consideration.

After performing the numerical analysis, we checked to see if the dataset had any null values. Visualisation has made use of the matplot.lib function. In this approach, graphs have been created to help with attribute knowledge.

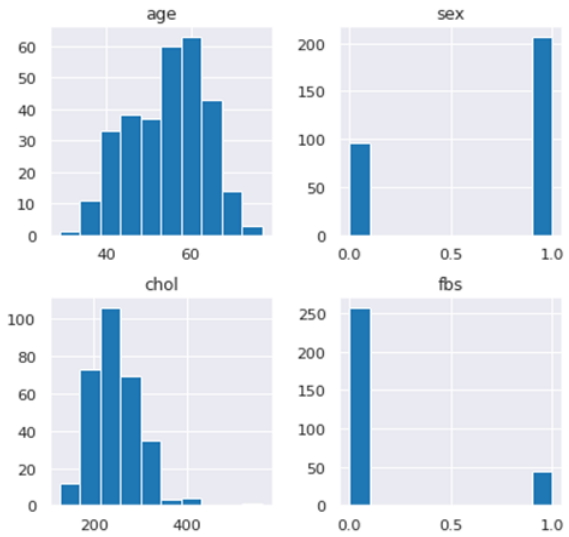

Fig 2: Person Correlation with attributes (1).

We saved certain values in x and y during the preparation of the data. The data has now been divided into training and prediction groups using the KNN and Decision tree algorithms.

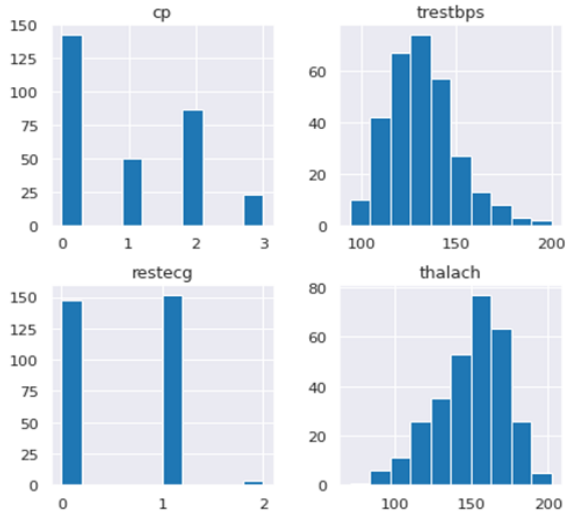

Fig 2: Person Correlation with attributes (2).

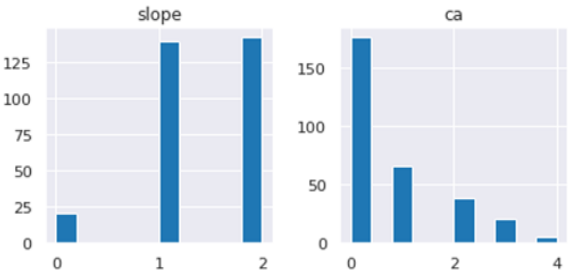

Fig 2: Person Correlation with attributes (3).

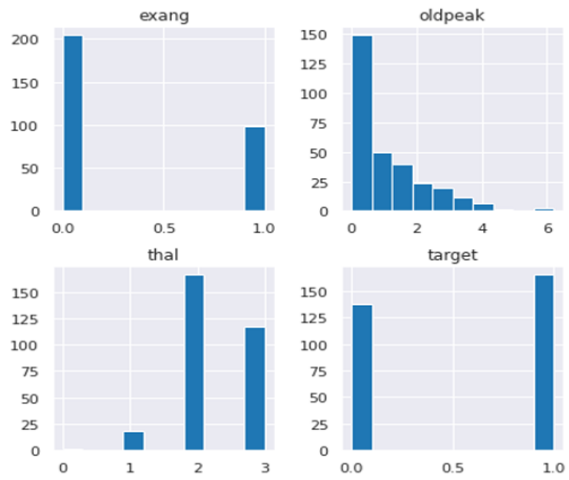

Fig 2: Person Correlation with attributes (4).

B] Flowcharts-

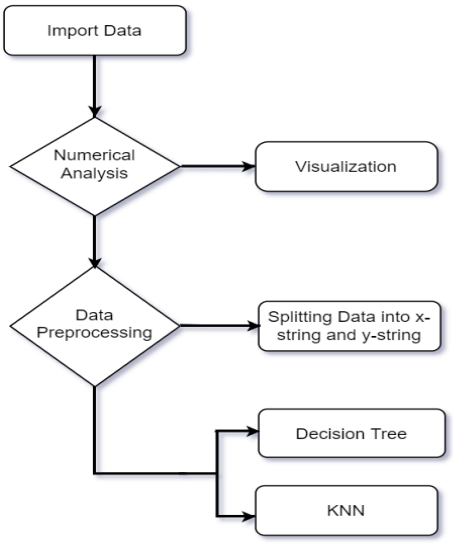

Fig 3: Workflow with dataset

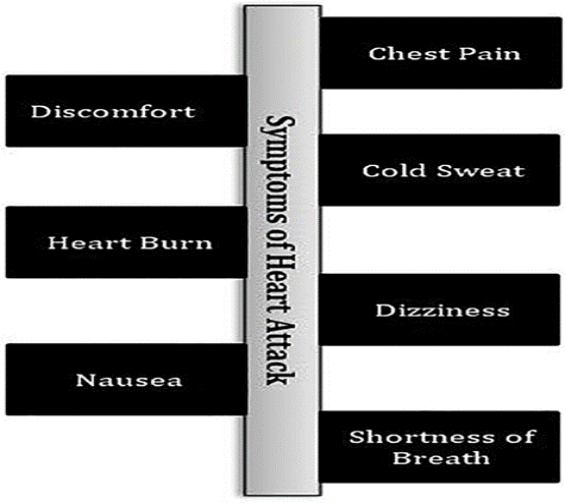

Fig 4: Symptoms of heart-attack

C] Data Balancing-

Data balancing is crucial for obtaining an accurate result since it shows that both classes are equal according to the data balance graph. The target classes in Fig. 5 are represented by "0" for patients without cardiac disease and "1" for those who do.

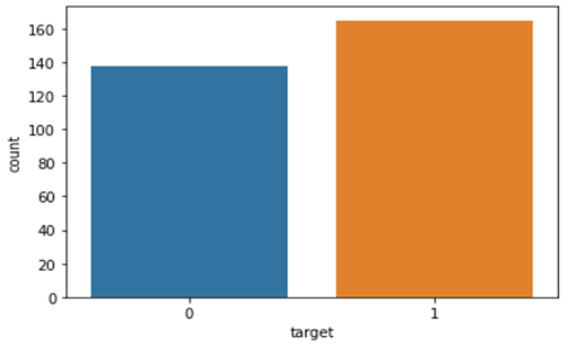

Fig 5: Target class view

D] Algorithms-

The dataset has already been imported, hence there is no need for label encoding because the data has already been imported as numerical values. Then, using the libraries that we loaded, a numerical analysis of the dataset was done. In the third phase, graphs were created using the Matplotlib tool for visualisation so that we could better comprehend the relationships between the heart disease-related characteristics. The data was then divided into X-strings and Y-strings using the data preprocessing approach. as an X-string. All attribute values are contained in the X-string, and all target values are contained in the Y-string. The two strings were then split into training and testing segments. 30% of the data was supplied to the testing module and 70% went to the training module. The decision tree algorithm was then put into practise. The logistic Regression (Supervised learning-Classification) technique has been used to the decision tree classifier algorithm. Since there are only two possible outcomes—"YES" or "NO"—logistic regression primarily focuses on predicting the next event based on the data that has already been presented. Then, we developed a function to determine which characteristic type is primarily responsible for heart disease. The KNN algorithm was then put into practise. After that, we developed a function that, when used, will provide the optimal value of K, improving the precision of our heart prediction model. The StandardScaler algorithm was used to determine the value of K. When the attributes of the input dataset

are measured in multiple units and the discrepancy between them is too great, StandardScaler is quite helpful. The distribution of data is scaled using the StandardScaler function so that the mean of the observed values is 0 and the standard deviation is 1. The comparison of the KNN algorithm and the decision tree has been completed.

*Data preprocessing –
Data Set: The suggested system uses open-source data sets, namely those that were taken from the Kaggle bug repository. There are 303 patient records in the dataset overall. For the properties of the given dataset, multiclass variables and binary classification are introduced. The presence or absence of cardiac disease is determined using the multi-class variable. If the patient has heart disease, the value is set to 1, otherwise it is set to 0 to indicate that the patient is heart disease-free. By transforming medical records into diagnosis values, data is preprocessed.

*Decision Tree –
In contrast, a decision tree is a graphical representation of the data and a type of reliable and practical machine learning algorithms.

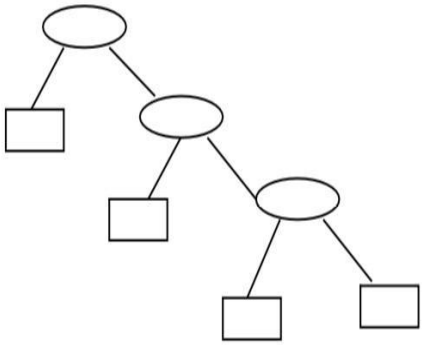

Fig 6: Decision Tree

We use the entropy of the data attributes to build the tree, and the root and other nodes are created based on the attributes.

*KNN (K Nearest Neighbors)-
It operates based on distance between data locations, and on this basis, various types of data are categorised with one another. The user determines the number of neighbors for each other's data sets, which is a very important factor in the analysis of the dataset.

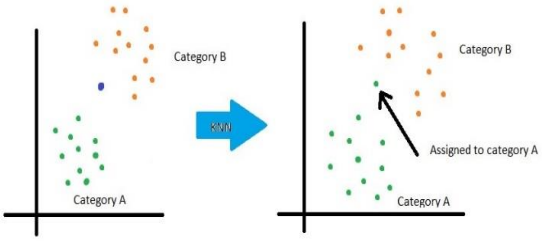

Fig 7: KNN where K = 4

*SVC -
A SVC (Support Vector Classifier)'s job is to appropriately fill out the data that the user has provided. It is primarily utilized for classification issues. It is an algorithm for supervised machine learning.

*LGBM Classifier-
Data is disseminated fast and with excellent performance in the LGBM Classifier, whose structure is built on the decision tree algorithm. It can be applied to a variety of tasks, including ranking and other machine learning applications.

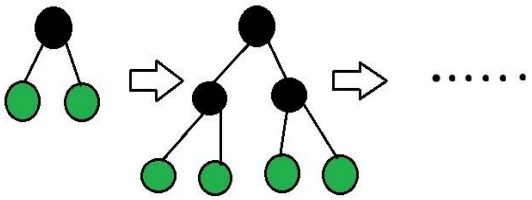

Fig 8: Level wise tree growth

4.RESULT

The final analysis demonstrates that KNN outperformed Decision Tree and SVC by a wide margin. The accuracy of the model created using KNN is 85.71%, which is 7% better than the accuracy of the Decision Tree and even better than the accuracy of the SVC model. Decision Tree might have been more accurate, but we discover that it is not appropriate for our data. SVC and LGBM Classifier, however, provide the same accuracy.

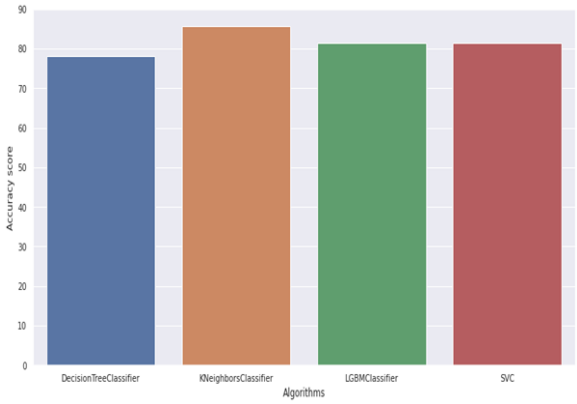

Fig 8: Accuracy of algorithm

5.FUTURE SCOPE

For prediction, we can combine a variety of machine learning algorithms. Additionally, it is preferable to use new search algorithms to choose the features before using machine learning techniques to improve the suggested system's accuracy and diagnose heart illnesses at an early stage, thereby saving numerous lives.

6.ADVANTAGES

• Heart disease prediction is a great project idea since it gives you a clear indication and evaluation of the condition of your cardiovascular system.
• It is a convenient all-in-one solution.
• The user can discuss their heart condition and receive a prompt diagnosis.

7.CONCLUSION

Since the heart is the most important and vital organ in the human body and heart disease prediction is a major concern for people, algorithm accuracy is one of the factors considered when evaluating how well an algorithm performs. The dataset utilised for both training and testing purposes affects how accurate machine learning algorithms are. Compared to the decision tree approach, the KNN algorithm is more accurate.

ACKNOWLEDGEMENT

We would like to thank our mentor Prof. Dr. C.M. Mahajan for timely guidance and assistance during the course of this EDI project. We would also thank to our college for giving this opportunity and freedom to do this project.

REFERENCE

[1] Pushpavathi T P, Santhosh Kumari, Kubra N K *"Heart Failure Prediction by Feature Ranking Analysis in Machine Learning"* Telematics Inform., vol36, pp915 923 A, Mar 2021
[2] Archana Singh, Rakesh Kumar, "Heart Disease Prediction using Machine Learning Algorithm" IEEE 2020
[3] Vijeta Sharma, Shrinkhala Yadav, Manjari Gupta *"Heart Disease Prediction using Machine Learning Techniques"* ISBN: 978-1-7281-8337-4/20/$31.00 ©2020 IEEE 2020 2nd International Conference on Advances in Computing, Communication Control and Networking
[4] Rahul Katarya, Polipireddy Srinivas *"Predicting Heart Disease at Early Stages using Machine Learning"*: A Survey Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) IEEE
[5] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia and J. Gutierrez, *"A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease",* 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, 2017
[6] S. Dhar, K. Roy, T. Dey, P. Datta and A. Biswas, *"A Hybrid Machine Learning Approach for Prediction of Heart Diseases"* 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018,
[7] Santhana Krishnan J and Geetha S, *"Prediction of Heart Disease using Machine Learning Algorithms"* ICIICT, 2019.
[8] Aditi Gavhane, Gouthami Kokkula, Isha Panday, Prof. Kailash Devadkar, *"Prediction of Heart Disease using Machine Learning",* Proceedings of the 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA), 2018.
[9] Senthil Kumar Mohan, Chandrasekar thirumalai and Gautam Srivastav, *"Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques"* IEEE Access 2019.

[10] Himanshu Sharma and M A Rizvi, "Prediction of Heart Disease using Machine Learning Algorithms: A Survey" International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 8, IJRITCC August 2017.

[11] M. Nikhil Kumar, K. V. S. Koushik, K. Deepak, "Prediction of Heart Diseases Using Data Mining and Machine Learning Algorithms and Tools" International Journal of Scientific Research in Computer Science, Engineering and Information Technology ,IJSRCSEIT 2019.

[12] Amandeep Kaur and Jyoti Arora,"Heart Diseases Prediction using Data Mining Techniques: A survey" International Journal of Advanced Research in Computer Science, IJARCS 2015-2019.

[13] Pahulpreet Singh Kohli and Shriya Arora, "Application of Machine Learning in Diseases Prediction", 4th International Conference on Computing Communication And Automation (ICCCA), 2018.

[14] M. Akhil, B. L. Deekshatulu, and P. Chandra, "Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm," Procedia Technol., vol. 10, pp. 85–94, 2013.

[15] S. Kumra, R. Saxena, and S. Mehta, "An Extensive Review on Swarm Robotics," pp. 140–145, 2009.

[16] Hazra, A., Mandal, S., Gupta, A. and Mukherjee, " A Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review" Advances in Computational Sciences and Technology , 2017.