IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114206

Heart Disease Prediction Using MachineLearning Algorithms and Models – Website Implementation

Rahul Vashistha[1], Aditya Randive[2], Pallavi Gade[3], Gaurav Pardeshi [4]

Student, Computer Science, DYPIET Dr. D. Y. Patil Institute of Engineering, Ambi, Pune, Maharashtra, India[1,2,3,4]

Abstract: People are undergoing a routine and busy schedule that leads to stress and anxiety. In addition to this, the percentage of people who are obese, stressed, and addicted to cigarettes is going up drastically [4]. This is leading to heart diseases. Heart diseases are one of the utmost causes of death in the world. The number of people affected by heart disease increases irrespective of age in both men and women [4]. The challenge behind these diseases is their timely prediction. While factors like gender, diabetes, and BMI also contribute to this disease, the chances of having heart disease also increase with the age. Men have a greater risk of heart disease. However, women also have the same possibility after menopause. Leading a stressed life can increase the chance of coronary heart disease.
In the proposed research, to pre-process data we've used techniques like the removal of noisy data, removal of missing data, filling default values if applicable, and classification of attributes for prediction and decision making at different levels. The performance of the diagnosis model is obtained by using methods like classification, accuracy, sensitivity, and specificity analysis [16]. This project proposes a prediction model to predict whether people have heart disease or not and to provide awareness or diagnosis on the same [16]. This is done by comparing the accuracies of applying rules to the individual results of Support Vector Machine, KNN classifier, Decision Tree Classifiers, and logistic regression on the dataset taken to present an accurate model of predicting cardiovascular disease.

Keywords: Coronary Heart Disease; Decision Tree Classifier; K Nearest Neighbor; Machine Learning; Naive Bayes; Support Vector Machine.

I. INTRODUCTION

People are undergoing a routine and busy schedule that leads to stress and anxiety. In addition to this, the percentage of people who are obese stressed, and addicted to cigarettes goes up drastically [4]. This is a main contributing factor that leads to heart diseases.
Cardiovascular diseases (CVDs) are the number one cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide [14]. Four out of 5 CVD deaths are due to heart attacks and strokes, and one-third of these deaths occur prematurely in people under 70 years of age [14]. Heart failure is a common event caused by CVDs and this paper is can be used to predict a possible heart disease using various algorithms to find their accuracies and select the one with the best outcomes. The number of people affected by heart disease increases irrespective of age in both men and women [4].
People with cardiovascular disease or who are at high cardiovascular risk need early detection and management wherein a machine learning model can be of great help.

II. MOTIVATION

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs [15]. Quality service implies diagnosing patients correctly and administering effective treatments. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information or decision support systems. Most hospitals today employ some sort of hospital information system to manage their healthcare or patient data [12]. These systems typically generate huge amounts of data which take the form of numbers, text, charts, and images.

III. PROBLEM STATEMENT

In this era of stress and anxiety, chronic diseases are a normal occurrence. The only way to control the mortality rate is through early prediction and proper treatment of these. So, we are trying for an early prediction and analysis of heart

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114206

disease by considering multiple various parameters like age, gender, blood pressure, heart rate, diabetes, and more. Since numerous factors are involved, the prediction is a bit challenging.

IV. PROJECT IDEA

We have used python, Tkinter, pandas, and sklearn library operations to perform heart disease classification for multiple algorithms on a collection of various data sets obtained from the UCI repository, Kaggle, and other open sources. This provides an easy-to-use visual representationof the dataset while building predictive analytics. ML process starts from the pre-processing of the data phase and is then followed by feature selection based on data cleaning, and at the last, classification of modeling performance based on the evaluation.

V. PURPOSE

In this project, we are trying to predict and analyze heart disease chances by considering various parameters like age, gender, blood pressure, cholesterol, and more. Since numerous factors are involved, the prediction is very challenging.

VI. OBJECTIVE

The main objective of this research is to develop a prototype Heart Disease Prediction System (HDPS) using multiple data mining and machine learning techniques, namely, Decision Trees, Logistic Regression, Support Vector Machine, and K-nearest neighbors' algorithm. HDPS can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions that traditional decision support systems cannot. Providing effective treatments also helps to reduce treatment costs. To enhance visualization and ease of interpretation, it displays the results both in tabular and graphical forms.

VII. LITERATURE REVIEW

Table 1 Literature Review

Sr. No	Title -Authors	Year – Publication Site	Description
1.	"Cardiovascular Disease PredictionUsing Machine Learning Models" - Atharv Nikam, Sanket Bhandari, Shamla Mantri.	Dec-2022, PuneCon	7 classifiers were used. DTC generated the highest accuracy of 73.12%. It's found that BMI plays an essential role. (**11 Attributes**) But there was a difference in testing and training accuracydue to small and hard data in the dataset.
2.	"Heart Disease Prediction Using Machine Learning Techniques" - Galla Bindhika, Munaga Meghana, Manchuri Sathvika Reddy.	April – 2020, IRJET	Results showed a good accuracy standard. Using Random Forest classification, we can find the prediction rate withoutequipment. (**11 Attributes**). But prediction results are not accurate. Data mining techniques do not provide effective decision-making. It cannot handle enormous datasets.
3.	"Early and Accurate Prediction of Heart Disease Using Machine Learning Model" - Sairabi H. Mujawar	Oct-2018, IJIRCCE	By using 2 different algorithms, the one with the 93%for presence of HD. 89%for absence of HD accuracy was determined. (**14 attributes)** While it doesn't work well with large data and accuracy is low, it also follows the black box nature soit is computationally expensive towork on.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114206

VIII. PROPOSED SYSTEM

After evaluating the results from the existing methodologies, we have used python and pandas operations to perform heart disease classification [7]. It provides an easy-to-use visual representation of the data set, working environment, and building predictive analytics [2]. ML process starts from a pre-processing data phase followed by feature selection based on data cleaning, and classification of modelling performance evaluation [2]. The random forest technique is applied to improve the accuracy of the result.

IX. PROPOSED WORK

In this system, the input details are obtained from the patient. Then from the user inputs, using ML techniques various factors are analyzed to satisfy the model equation and thus estimate and predict the heart disease possibility factor. Now, the obtained results are compared with the results of existing models within the same domain and found to be improved. The data of heart disease patients collected from the UCI ML repository is used to discover patterns with KNN,
DT, SVM, and Naive Bayes. The results are compared for performance and accuracy with these algorithms.

X. ALGORITHMS:

5.1 K-Nearest neighbor: It's a classification algorithm. The class of a particular data point is determined based onthe class which is most common among its k nearest neighbors where k is a small positive integer.

5.2 Support vector machine: It's an algorithm that is used in machine learning for classification and regression techniques. It is regularly used as a classification technique due to its efficiency when compared with the other algorithms. This technique plots a hyperplane for every attribute as a coordinate that is present in the dataset.

5.3 Logistic regression: It's a predictive analysis technique that is used when the target variable is dichotomous (binary). The logistic Regression model explains the relationship between one dependent binary variable and one or more independent variables.

5.4 Decision Tree Classifier: It organizes the characteristics to inferences about the target value. The classification trees are the tree models in which the target parameter can acquire a finite set of values. In these, the class labels are signified by the leaves, and the branches describe the concurrences of features that guide those class labels. The regression trees are the decision trees in which the target parameter can take the continuous value.

XI. USE CASE DIAGRAM

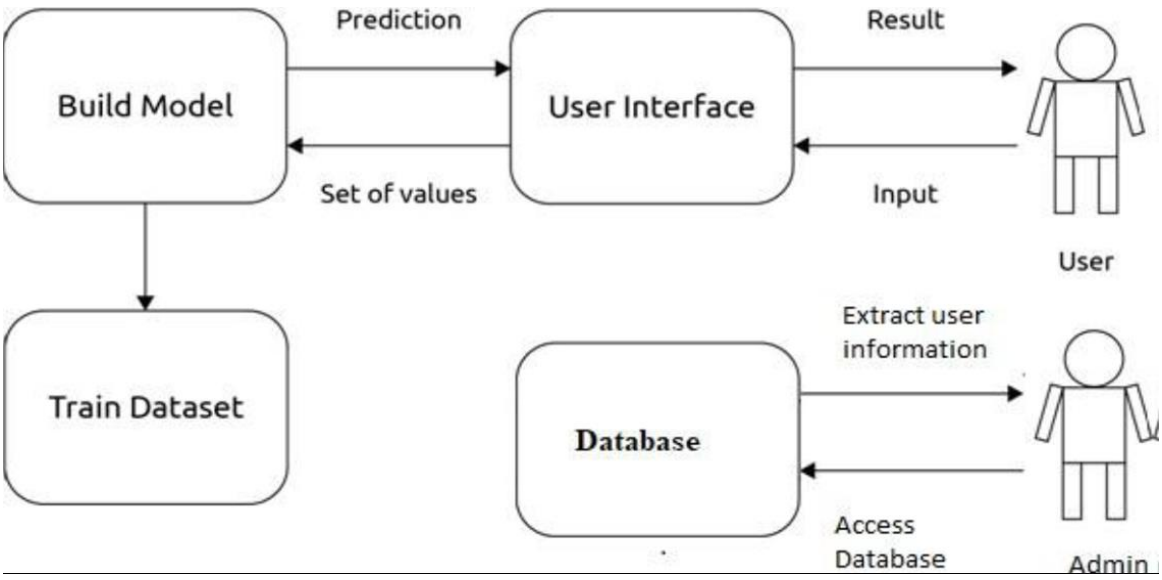

FIGURE 1 UML - ARCHITECTURE DESIGN

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114206

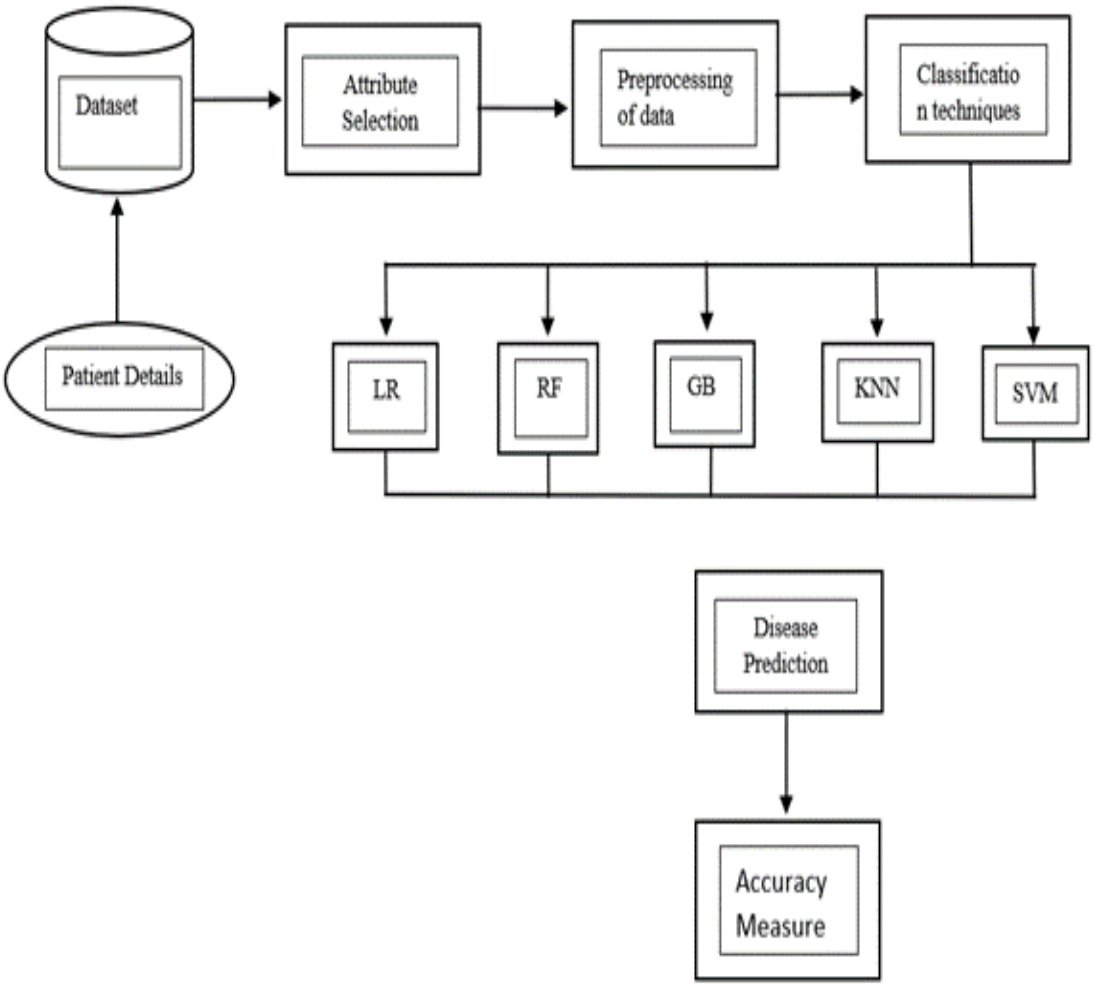

FIGURE 2 UML – PROPOSED ARCHITECTURE SYSTEM

XII. MATHEMATICAL MODEL

KNN falls in the supervised learning algorithms. This means that we have a data set with labels training measurements (x,y) and would want to find the link between x and y.
Our goal is to discover a function h:X→Y so that having an unknown observation x, h(x) can positively predict the identical output y.
In the classification problem, the K-nearest neighbor algorithm essentially said that for a given value of K, the algorithm will find the K nearest neighbor of an unseen data point, and then it will assign the class to the unseen data point by having the class which has the highest number of data points out of all classes of K neighbors.

XIII. MATHEMATICAL MODEL EQUATIONS

For distance metrics, we will use the Euclidean metric.

$$d(x, x') = \sqrt{(x_1 - x'_1)^2 + \ldots + (x_n - x'_n)^2}$$

$$P(y = j|X = x) = \frac{1}{K} \sum_{i \in \mathcal{A}} I(y^{(i)} = j)$$

Finally, the input x gets assigned to the class with the largest probability

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114206

XIV. PROJECT IMPLEMENTATION

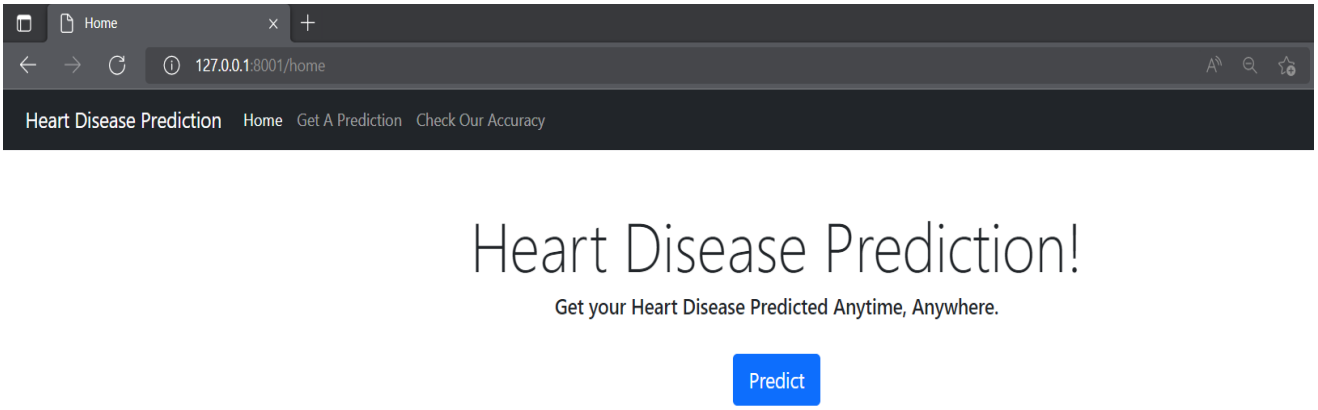

Fig 1: Home page of our website

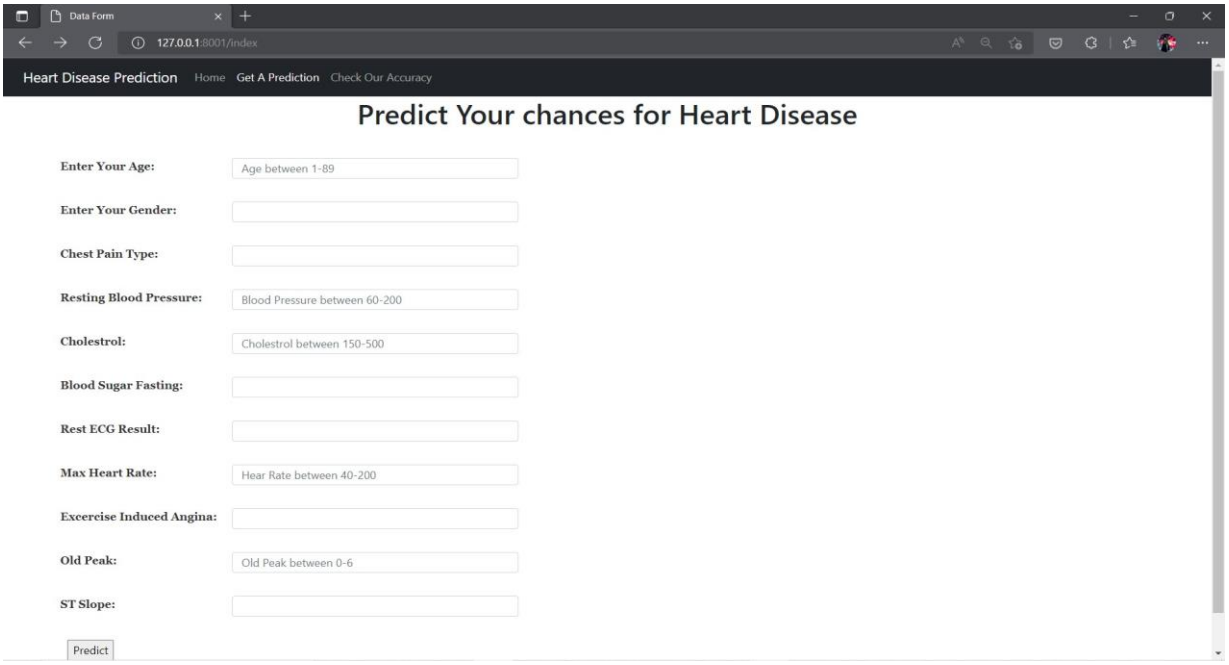

Fig 2: Data form

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114206

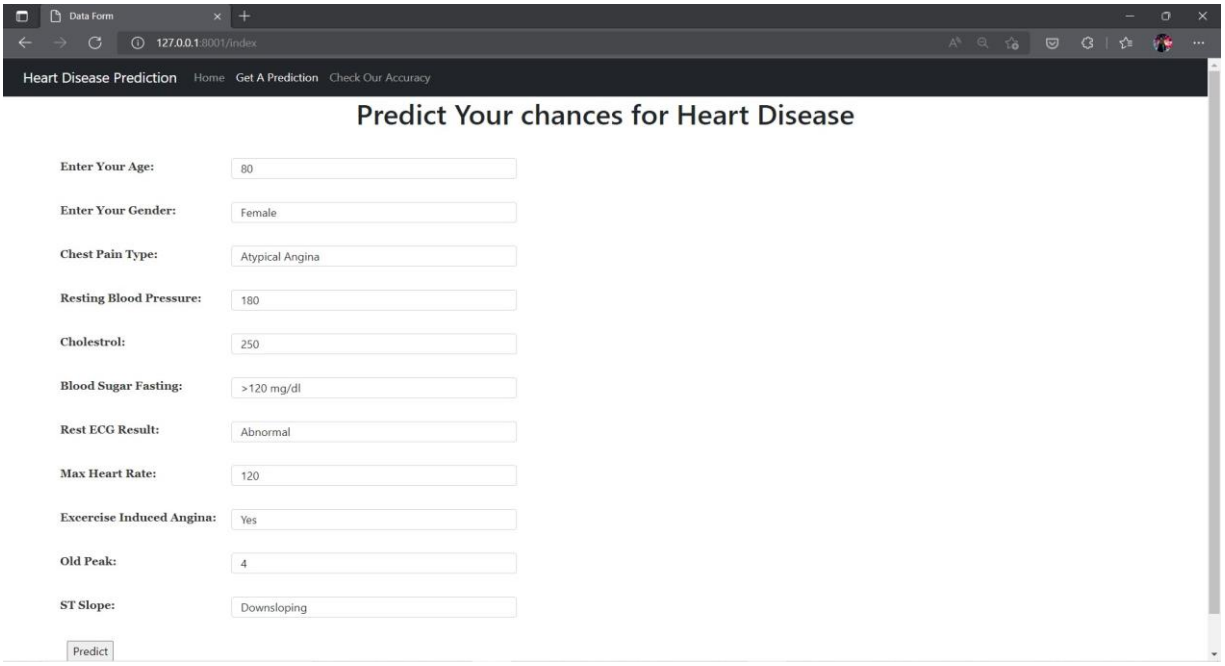

Fig 3: Filled data form

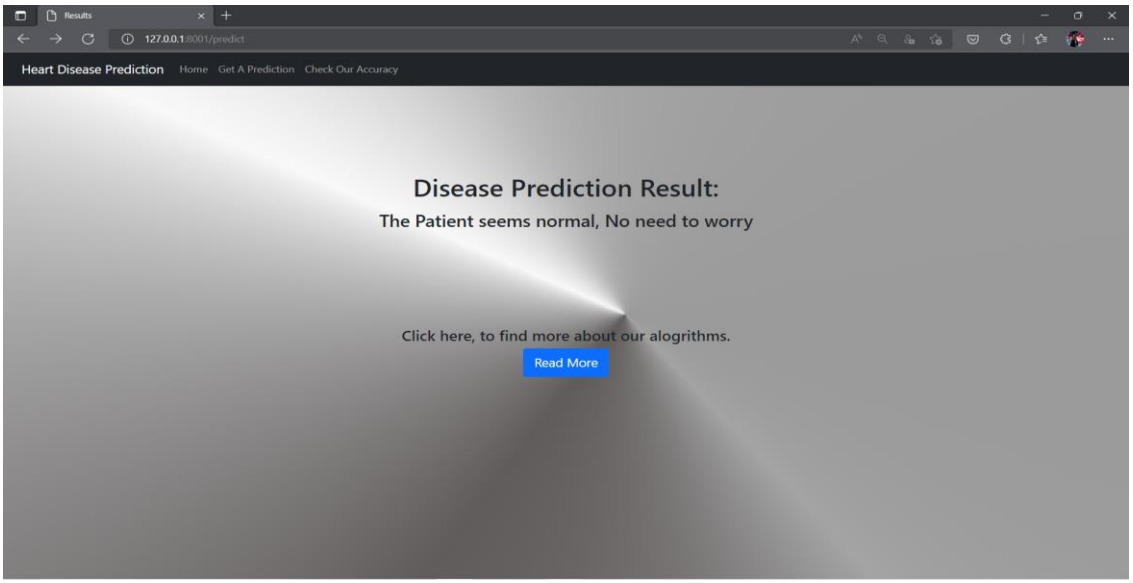

Fig 4: Prediction Results

XV. ADVANTAGES

1. By trying multiple algorithms we've increased accuracy and thus giving more effective heart disease prediction.
2. Early and online prediction can prove very useful in case of a medical emergency.
3. A free web platform provides a cost-effective diagnosis for patients.

XVI. DISADVANTAGES

1. A computerized system alone does not ensure accuracy since the prediction system is not fully automated, we still need the user to enter a wide variety of data for diagnosis, and the warehouse data is not faultless and substantiate.
2. The model cannot handle immeasurable datasets of patient records and data processing for prediction.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 7.39 ⁜ Vol. 11, Issue 4, April 2022

DOI: 10.17148/IJARCCE.2022.114206

XVII. APPLICATIONS

Websites and web apps: The disease prediction system can be combined with a health care website or consultancy app which will give an idea about the possible diseases the user has and recommend him requisite doctor or medication or tests.
IOT: The prediction system can be instrumented in an IoT device (wristband) which can easily monitor the predictive factors and give an analysis of the current condition and chances for diseases.
Hospital Pre-analysis System: This system can be implemented in hospitals where after a checkup the user data can be constructed and used for multi-disease prediction easily.

XVIII. CONCLUSION

Heart disease prediction is a necessity as well as exigent work in the medical field. The mortality rate can be reduced if the disease is recognized at the initial stages, and precautions and proper treatment are possible.
The algorithms are tested using various features. Accurate forecasting of the diseases is the goal of the proposed method. The decision classifier approach proved to be very efficacious to predict the diseased using features like age, BMI, cholesterol, and more. Adding feature BMI improved the accuracy of prediction.
Thus, by assessing the results, the suggested approach generates a more precise prediction of cardiovascular diseases. Our project focuses on analyzing and designing a system where patients' real-time information can be processed and evaluated based on previous symptoms and current symptoms for different diseases [2].
We have concluded that KNN, Support Vector, Decision tree, and Random Forest are the best algorithms with higher accuracy rates than others for predicting and analysis among these KNN is easy to implement and requires less computational resources and thus could be implemented in a web-based system effortlessly.

IX. FUTURE SCOPE

In the future, these techniques can be applied to a real-time database of the individual patient and by using the same attributes or by adding some more attributes we can determine the prediction of multiple diseases like kidney-related and lungs related diseases. We can also implement the algorithms used previously for better results. This can be deployed to
android and web platforms to analyze and predict using real-time data & by collaborating with doctors or medical organizations or as a common platform for predicting diseases. As an extension to this work and some sort of limitation to the work performed here, different types of classifiers can be included in the analysis, and more in-depth sensitivity analysis can be performed on these classifiers, also an extension can be made by applying the same analysis to other diseases datasets, and see the performance of these classifiers to classify and predict these diseases [16].

X. ACKNOWLEDGMENT

This paper is supported by many people, some of whom have a direct role and some of them have an indirect way by publishing their research online which helped us to understand this concept easily. We express our deepest gratitude, sincere thanks, and a deep feeling of appreciation to our **Project Guide Prof. Sagar Dhanake**, his presence at any time throughout our project, important guidance, opinion, comment, critics, encouragement, and support greatly improved this project work. We are obliged to the college administration for providing the necessary infrastructure and technical support. Goals make us do the work; vision is more important than the goal which makes us do work in the best way to make work equally the best. Thanks to all the colleagues for their extended support and valuable guidance. I would like to be grateful to all of our friends and family members for their consistent support, help, and guidance.

REFERENCES

[1] Atharv Nikam, Atharv Nikam, etc, December 2020, "Cardiovascular Disease Prediction Using Machine Learning Models", (PuneCon).
[2] Galla Siva Sai Bindhika, Munaga Meghana, Manchuri Sathvika Reddy, etc, April 2020, "Heart Disease Prediction Using Machine Learning Techniques", (IRJET).
[3] Garima Choudhary, Dr. Shailendra Narayan Singh, April 2020, "Prediction of Heart Disease using Machine Learning Algorithms", (ICSTCEE 2020).
[4] M. Marimuthu, S. Deivarani, Gayathri R., "Analysis of Heart Disease Prediction using Various Machine Learning Techniques" – IEEE 2019.
[5] Dinesh Kumar G, Santosh Kumar G, Arumugaraj K, "Prediction of Cardiovascular Disease Using Machine

DOI: 10.17148/IJARCCE.2022.114206

Learning Algorithms", IEEE ICCTCT 2018.

[6] H. A. Esfahani and M. Ghazanfari, ''Cardiovascular Disease Detection using a new ensemble classifier,'' in Proc. IEEE 4th Int. Conf. Knowledge Based Eng. Innovation (KBEI), Dec. 2017.

[7] Supriya Kaskar, Sonali Gandewar, Heart Disease Prediction based on Machine Learning and Deep Learning Techniques", IRJET, Aug 2021, p-ISSN: 2395-0072.

[8] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, ''Using PSO algorithm for producing best rules in diagnosis of heartdisease,'' in International Conference Computer Application (ICCA), Sep. 2017.

[9] Sonam Nikhar, A.M. Karandikar, " Prediction of Heart Disease Using Machine Learning Algorithms"- Vol-2 Issue-6, June 2016.

[10] Jaymin Patel, Prof. Tejal Upadhyay, Dr. Samir Patel, "Heart Disease Prediction using Machine learning and Data Mining Technique" Sept2015-March 2016.

[11] Sana Bharti, Shailendra Narayan Singh" Analytical study of heart disease comparing with different algorithms": Computing, Communication & Automation (ICCCA), 2015 International Conference.

[12] S. Chaitrali, Dangare and S. Apte Sulabha, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", IJCA, June 2012.

[13] V Ramalingam, Ayantan Dandapath, M Karthik Raja, "Heart disease prediction using machine learning techniques: A survey", March 2018 International Journal of Engineering & Technology (IJET).

[14] Hongzu Li, Pierre Boulanger, "A Survey of Heart Anomaly Detection Using Ambulatory Electrocardiogram (ECG)", May 2020.

[15] Obenshain, M.K: "Application of Data Mining Techniques to Healthcare Data", Infection Control and Hospital Epidemiology, 25(8), 690–695, 2004.

[16] Rahul Vashistha, Aditya Randive, et, "Heart Disease Prediction Using Machine Learning Algorithms and Models", IJARCCE, Dec 2021, 10.17148/IJARCCE.2021.101223.

BIOGRAPHY

Rahul Vashistha: Bachelors of Engineering in Computer Science. **Interests** – Virtual\Mixed Reality, Extended Reality, Cyber Security, Ethical Hacking, Web Site Development, Software Development.

Aditya Randive: Bachelors of Engineering in Computer Science. **Interests** – Machine learning, Robotics Processing Automation, Web Development, Ethical Hacking, Java Development.

Pallavi Gade: Bachelors of Engineering in Computer Science. **Interest** – Web Development, Ethical Hacking, Java Development, ML, Android Development.

Gaurav Pardeshi: Bachelors of Engineering in Computer Science. **Interest** – Web Development, Java Development.