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

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101223

Heart Disease Prediction Using Machine Learning Algorithms and Models

Rahul Vashishta[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: Healthcare is an inescapable task to be done in human life. Cardiovascular disease is a broad category for a range of diseases that are affecting the heart and blood vessels. The early methods of forecasting cardiovascular diseases helped in making decisions about the changes to have occurred in high-risk patients which resulted in the reduction of their risks [5]. The health care industry contains lots of medical data, therefore machine learning algorithms are required to make decisions effectively in the prediction of heart diseases [4]. Recent research has delved into uniting these techniques to provide hybrid machine learning algorithms [5]. In the proposed research, data pre-processing uses techniques like the removal of noisy data, removal of missing data, filling default values if applicable, and classification of attributes for prediction [5] and decision making at different levels. The performance of the diagnosis model is obtained by using methods like classification, accuracy, sensitivity, and specificity analysis [5]. This project proposes a prediction model to predict whether people have heart disease or not and to provide awareness or diagnosis on the same [1]. 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 [1] taken to present an accurate model of predicting cardiovascular disease.

Keywords: Heart Diseases; Machine Learning; Support Vector Machines; Decision Tree Classifier; KNN Classifier; Logistic Regression; Model Interpretation.

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 [13]. 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 best outcomes. The number of people affected by heart disease increases irrespective of age in both men and women [13].

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. PROBLEM STATEMENT

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

III. LITERATURE REVIEW

Sr. No	Title -Authors	Year – Publication Site	Description
1.	"Cardiovascular Disease Prediction Using Machine Learning Models" - Atharv Nikam, Sanket Bhandari, Shamla Mantri.	Dec-2022, PuneCon	7 classifiers were used. Decision tree classifier 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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 10, Issue 12, December 2021
DOI: 10.17148/IJARCCE.2021.101223

			and training accuracy due to smaller and hard data in the dataset.
2.	"Heart Disease Prediction Using Machine Learning Techniques" - Galla Bindhika, Munaga Meghana, Manchuri Sathvika Reddy.	April – 2020, IRJET	The results showed a good accuracy standard. By introducing Random Forest classification, we can find the prediction rate without equipment. (**11 Attributes**) But the prediction results is not accurate. Data mining techniques does not help to 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 so it is computationally expensive to work on.

Table 1 Literature Review

IV. PROPOSED SYSTEM

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

V. ALGORITHMS:

5.1 K-Nearest neighbour: It's a classification algorithm. The class of a particular data point is determined based on the class which is most common among its k nearest neighbours 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.

IJARCCE
ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 10, Issue 12, December 2021
DOI: 10.17148/IJARCCE.2021.101223

VI. USE CASE DIAGRAM

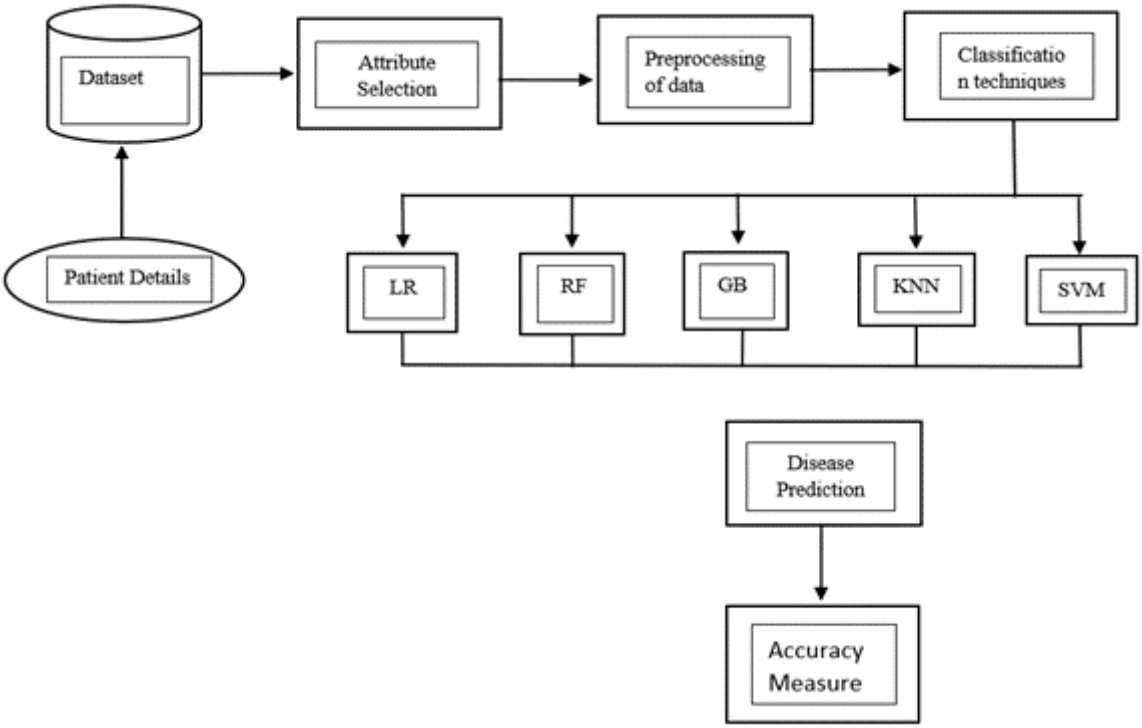

FIGURE 1 UML - SYSTEM ARCHITECTURE

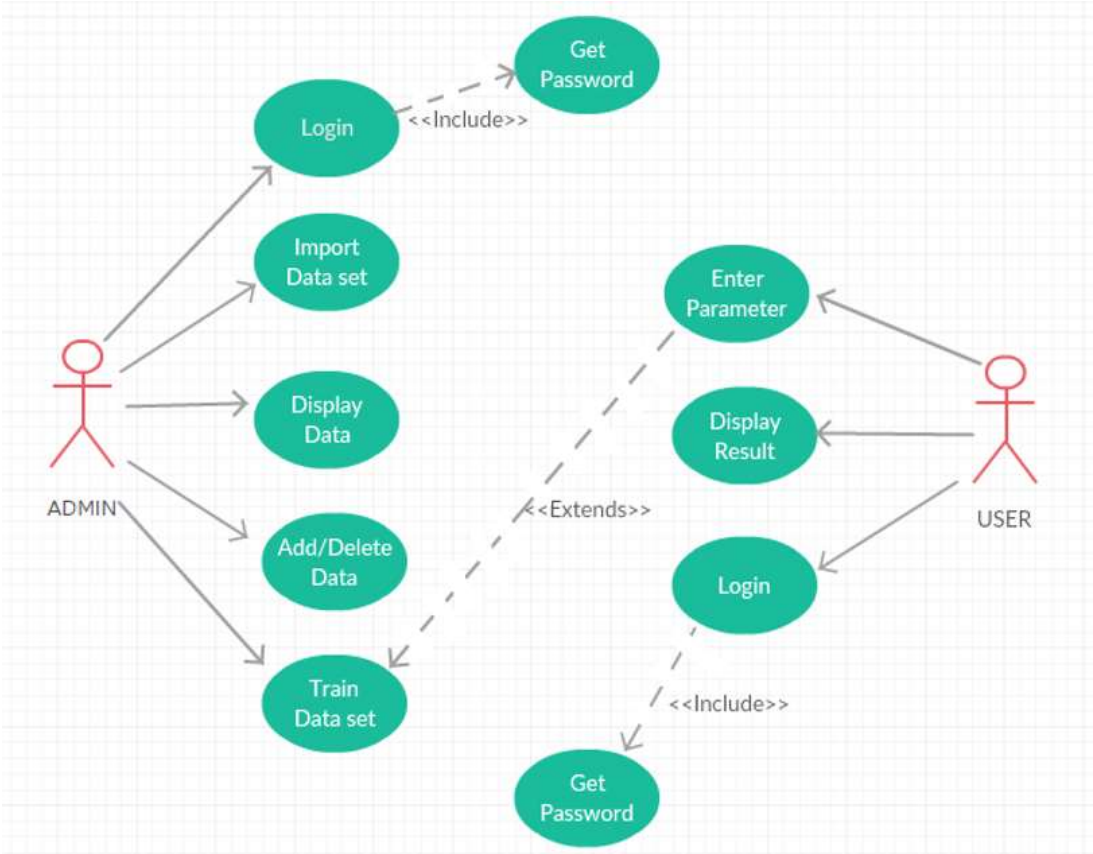

FIGURE 2 UML – CLASS DIAGRAM

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 10, Issue 12, December 2021
DOI: 10.17148/IJARCCE.2021.101223

VII. 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 medical emergency.
3. A free web platform provides a cost-effective diagnosis for patients.

VIII. 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 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.

IX. 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 on current symptoms for different diseases [2].
We have concluded that KNN, Support Vector, Decision tree, Random Forest are the best algorithms with higher accuracy rates than others for predicting and analysis among these KNN is easy to implement and require less computational resources and thus could be implemented in a web-based system effortlessly.

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 to understand this concept easily. We express our deepest gratitude, sincere thanks and deep feeling of appreciation to our Project Guide Prof. Mangesh Manke, his presence at any time throughout the Semester, 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. Finally, we extend our heartfelt gratitude to our fellows and family members.

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 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] Sanjay Kumar Sen" Predicting and Diagnosing of Heart Disease Using Machine Learning Algorithms"- International Journal of Engineering and Computer Science June 2017.
[8] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, ''Using PSO algorithm for producing best rules in diagnosis of heart disease,'' 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" Sept 2015-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.

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

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 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.