Heart Disease Detection and Prediction Using ML Algorithms in Python

V. Ramesh[1(✉)], M. Swamy Das[2], and B. Nageshwar Rao[1]

[1] Osmania University, Hyderabad, Telangana, India
rameshvoruganti36@gmail.com
[2] Department of Computer Science and Engineering, CBIT, Hyderabad, Telangana, India

Abstract. Cases of cardiac disease are growing at a disturbing rate. It is critical to diagnose and anticipate any such disorders. This conclusion is a difficult task that must be completed unambiguously and successfully. The research efforts on which patients are prospective to have heart disease are based on numerous medicinal features and health state factors. Developed a method for detecting and predicting heart disease by analyzing a patient's medical times past to govern if the patient would be diagnosed with heart disease. Machine learning approaches such as linear regression, logistic regression, and K-Nearest Neighbors to forecast and categorize a patient with heart disease. A supportive method was utilized on the way to control how the model may be used to boost the precision of heart Risk in each single. The presented model's ability hushed up fulfilling and could forecast proof of taking heart disease in a certain separate by applying KNN and logistic regression, which demonstrated great precision in contrast to previously utilized classifiers such as naive bays. As a result, by using the offered technique in determining the probability of the classifier properly and precisely recognizing a heart condition, a crucial measure of stress has been relieved. The Provided heart disease prediction system enhances clinical consideration while lowering costs. This research provides important information that can assist us anticipate patients with heart disease, and this one is written in Python.

Keywords: Heart disease · KNN · logistic regression · and probability

1 Introduction

Machine learning is an unusually massive and varied science, and that one application implementation is expanding in our daily lives. To forecast and determine the correctness of a given dataset, machine learning integrates ensemble learning classifiers supervised and unsupervised [1]. It can include the information in this HDPS document because it will benefit many people. "Machine Learning is a technique for manipulating and extracting implicit, previously unknown/known, and possibly relevant data information" [2]. Cardiovascular disorders are mutual these days; they define a wide enclosed of harms that can upset your heart. Permitting to the World Health Organization, (cardiovascular diseases) CVD deaths 17.9 million people globally. That is the basic explanation of

B. Raj et al. (Eds.): ICETE 2023, AER 223, pp. 347–354, 2023.
https://doi.org/10.2991/978-94-6463-252-1_38

fatalities among adults. It can assist forecast individuals who are prospective to develop heart disease based on their clinical history [3]. It determines who is experiencing any side effects of heart syndrome, for instance, chest discomfort or high blood pressure, and can support in identifying sickness by fewer clinical trials and potent drugs, to make them feel much better similarly. This study is primarily concerned with three machine-learning techniques: Random Forest Classifier, KNN, and logistic regression. The accuracy is 87.5%, which is higher than in the prior system, which only utilized one machine-learning approach [4]. As a result, employing additional machine learning approaches improved efficiency and HDPS accuracy. Logistic regression belongs to the kind of supervised learning [5]. Distinct characteristics are employed in cutting-edge logistic regression. A goal of this study is headed for determine if the patient is likely to be identified with some cardio-vascular heart problems established on clinical factors for instance age, gender, fasting sugar level, chest discomfort so on [6]. The UCI vault is searched for a dataset including the patient's clinical history and characteristics. A forecast of whether the patient has heart disease using this dataset. A utilize 14 clinical variables of a patient to predict whether the patient will develop heart disease [7]. Three techniques are used to train these medical qualities: logistic regression, Random Forest Classifier, and KNN. It is the maximum successful algorithm, with an accuracy of 88.52% [8]. To end, organize people who are in danger of developing heart disease or not, and this technique is entirely cost-operational.

2 Related Work

This work was prompted by a critical quantity of effort linked to the detection of Cardiovascular Heart Disease utilizing machine-learning methods. This study offers a short literature survey [9]. Several algorithms, including Random Forest Classifier, KNN, Logistic Regression, and others, were used to make a compelling cardiovascular disease prediction. Each algorithm's ability to register the given objectives can be selected in the Results [10]. The model combining IHDPS was able to determine the decision limit by combining new and old deep learning and machine learning techniques. It worked with the most important and critical variables/information, such as family heritage connected using any heart disease [11]. Nevertheless, the precision of such IHDPS techniques was significantly less than that of the new future model for diagnosing heart disease. Using an artificial neural network, machine, and deep learning technologies to treat heart disease. McPherson et al. discovered the crucial parts of heart disease or atherosclerosis utilizing the inbuilt execution algorithm employing Neural Network methods and were simply precisely equipped to forecast whether the investigation subject is suffering from the provided disorder or not. R. Subramanian et al. colleagues used neural networks to detect and predict heart disease and blood pressure, among former things. [12]. A deep neural network was fabricated by combining a specified disease quality to generate the result, which was finished by the result perceptron and comprised almost 120 hidden layers, which is the maximum crucial and pertinent approach on behalf of assuring an accurate result of taking heart disease.

A supervised network has been recommended for heart disease diagnosis [13]. Although a specialist of doctors using new data tested the model, the model was cast-off,

produced from previous scholarly data, and projected the outcome, thus computing the correctness of the supplied techniques.

3 Data Source

A predetermined dataset of separate picked based on their past heart problems and former medical diseases. Heart ailment mentions different illnesses that affect the heart. Allowing to the World Health Organization (WHO), cardiovascular disorders are the major cause of transience among middle-aged persons [14]. They use an information source that includes the clinical history of 304 diverse individuals of varied ages. A dataset provides the clinical qualities of the patient, for example, age, resting blood pressure, and fasting sugar level. Which aids us in determining if the patient has heart disease or not. [15]. This dataset comprises 13 clinical features of 304 individuals to aid us in determining whether the patient is at the hazard of evolving a coronary illness or not, on top of describing patients who are at risk of developing coronary heart disease and those who are not [16]. This coronary heart disease dataset was collected from the UCI repository. This dataset sources the example that triggers the location of a patient in danger of coronary heart disease. These proceedings are divided into two categories: testing and training. This dataset has 303 lines and 14 sections, with each column compared to a separate record (Table 1).

4 Methodology

This research examines several machine learning methods, including Random Forest Classifier, Logistic Regression, and K closest neighbors (KNN) which might be valuable for specialists or clinical examiners in accurately evaluating the cardiac condition. This documentation includes reviewing journals, published articles, and current data on cardiovascular illness [17]. The technique provides a system for the suggested model. The approach is a process of converting provided data into perceived data patterns for the consumers' understanding. It is divided into phases, with the first being the collection of statistics, the second being the taking out of significant values, and the third being the pre-processing side by side where we discover the data. Depending on the techniques employed, data pre-processing provides missing values, record cleansing, and normalization. The classifier is charity to classify the pre-processed facts when it has been pre-processed. The classifier utilized within the suggested version is Random Forest Classifier, KNN, and Logistic Regression. Ultimately, the proposed version is implemented, and we assessed our version based on accuracy and overall performance using a variety of overall performance measures. An effective "Heart Disease Detection and Prediction System" (EHDPS) has formed in this model with the use of unique classifiers. This version predicts using thirteen scientific factors such as chest discomfort, cholesterol, blood pressure, fasting sugar, intercourse, age, and so on.

Table 1 Represents the various used are listed

S. No	Observation	Description	Values
1	Age	Age in Year	Continuous
2	Sex	Sex of Subject	Male/Female
3	FBS	Fasting Blood Sugar	<, or > 120 mg/dl
4	Restecg	Resting Electrocardiograph	Five values
5	Slop	The slope of the Peak Exercise ST segment	Up/Flat/Down
6	Thalach	Maximum Heart rate Achieved	Continuous
7	Exang	Exercise Induced Angina	Yes/No
8	Trestbps	Resting Blood Pressure	Continuous
9	CP	Chest pain	Four types
10	Oldpeak	ST Depression when Workout Compared to the Amount of Rest Taken	Continuous
11	Thal	Defect type	Reversible/Fixed/Normal
12	Ca	Gives the number of major Vessels Colored by Fluoroscopy	0–3
13	Num(Disorder)	Heart Disease	The four major types
14	Chol	Serum Cholesterol	Continuous

5 Results and Discussion

According to the results, even if the majority of the researchers utilize different algorithms. For instance, a decision tree for the detection of sufferers identified using heart disease, Random Forest Classifier, KNN, and Logistic regression. They provide a higher result to out rule them. The methods utilized are more precise and save a portion of the money. Considerably greener and wilder than the techniques used by previous studies. In addition, the greatest precision obtained using KNN and Logistic Regression is the same at 88.5%, which is greater than, or almost equal to, the accuracy obtained from the prior researcher. To summarize, our accuracy has improved because the better scientific features from the dataset are used. The research also shows that KNN and logistic regression outperform random forest classifiers in terms of predicting whether a patient has coronary heart disease. This demonstrates that Logistic Regression and KNN have a better prognosis for coronary heart disease. The subsequent 'parent 2', 'parent 3', 'parent 4', and 'parent 5' indicates a plot showing the variety of sufferers that are been separated and expected using a classifier reliant upon Sex, Chest Pain, age group, and Resting Blood Pressure. Figure 1 shows the classifier Heart diseases in KNN, Logistic Regression, Random Forest algorithms. Figure 2 depicts the risk of having a risk of heart dependent on their age. Figure 3 depicts the risk of heart dependent scheduled

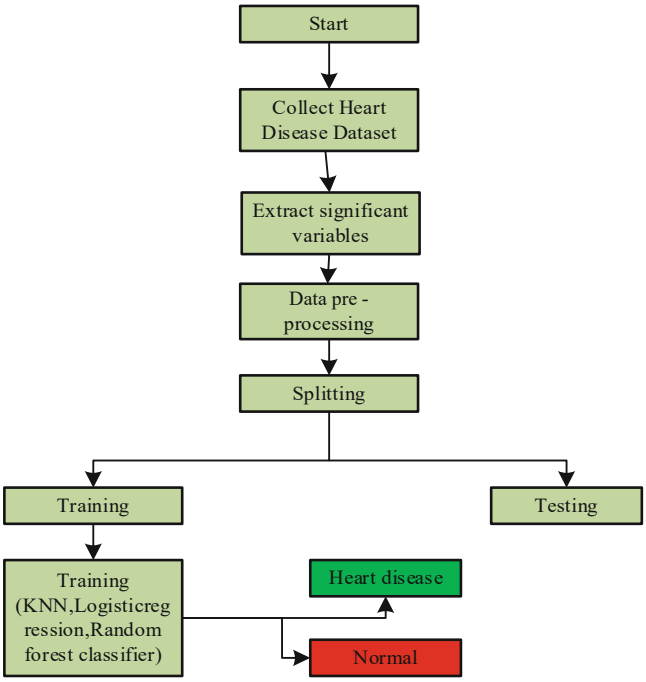

Fig. 1. shows the classifier Heart diseases in KNN, Logistic Regression, Random Forest algorithms

an individual's resting blood pressure. Figure 4 displays the individuals devising or not having heart disease depending off kind of chest pain.

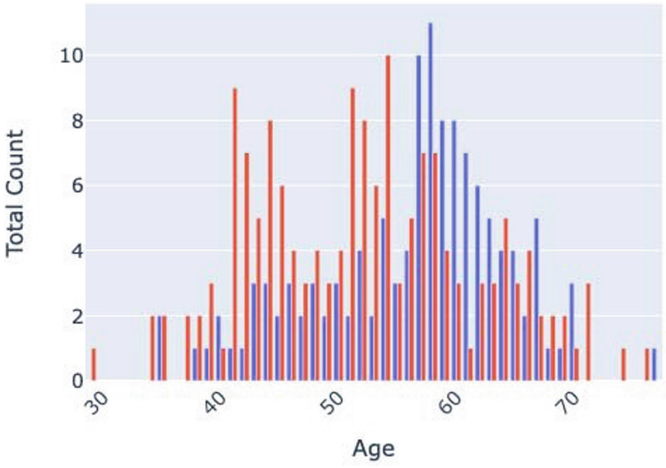

Fig. 2 Displays the risk of having a heart attack dependent off their age

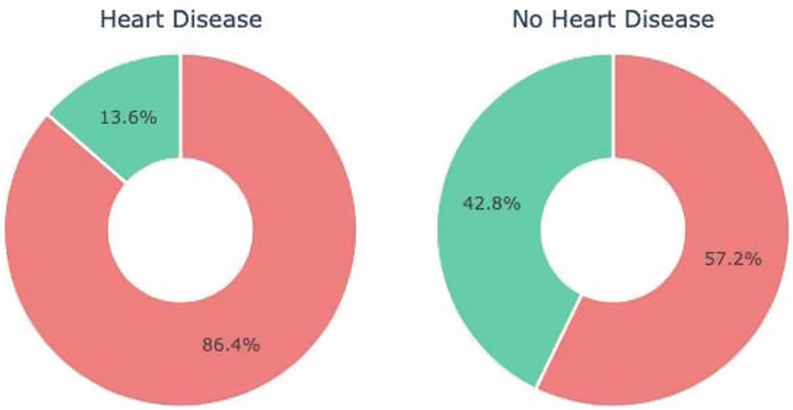

Fig. 3 Based on their resting blood pressure, this graph shows their risk of having a heart attack

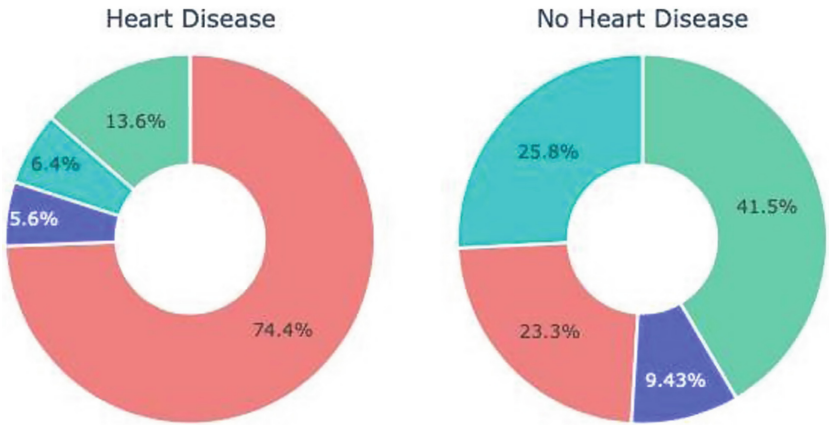

Fig. 4 Displays the individuals devising or not having heart disease depending off kind of chest pain

6 Conclusion

As a cardiovascular infection recognition version has been superior to the usage of 3 ML type modeling techniques. This venture guesses humans with cardiovascular sickness via way of means of extracting the affected person's scientific records that end in an incurable coronary heart sickness from a dataset that consists of patients' scientific records, which include chest pain, blood pressure, and sugar level. A heart disease detection Device supports a patient largely centered entirely on the medical history of having been detected with a previous coronary heart sickness. The set of rules applied in generating the given version is Random Forest Classifier, Logistic regression, and KNN. Our version is 87.5% accurate. The use of extra schooling statistics guarantees the better possibilities of the version to appropriately expect whether or not the given man or woman has a coronary heart disorder or not [9]. By the use of these, pc aided strategies

we will are expecting the affected person to speedy and higher and the value may be decreased very much. There are some scientific databases that we will paint on as those Machine gaining knowledge of strategies are higher and they can are expecting higher than a man or woman which allows the affected person in addition to the doctors. The result by sprucing up the dataset and employing logistic regression and KNN hopes to acquire a precision of 87. 5% on the method, which is improved from the prior technique' accuracy of 85%. Furthermore, it is established that the precision of KNN is the peak of the algorithms utilized, i.e., 88.52%.

References

1. Soni J, Ansari U, Sharma D & Soni S (2011). Predictive data mining for medical diagnosis: an overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43–8
2. 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), 44–8.
3. Ordonez C (2006). Association rule discovery with the train and test approach for heart disease prediction. IEEE Transactions on Information Technology in Biomedicine, 10(2), 334–43.
4. Shinde R, Arjun S, Patil P & Waghmare J (2015). An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm. International Journal of Computer Science and Information Technologies, 6(1), 637–9.
5. Bashir S, Qamar U & Javed M Y (2014, November). An ensemble-based decision support framework for intelligent heart disease diagnosis. In International Conference on Information Society (i-Society 2014) (pp. 259–64). IEEE.
6. Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014). A coronary heart disease prediction model: the Korean Heart Study. BMJ open, 4(5), e005025.
7. Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Ärnlöv J & Ingelsson E (2013).Multilocus genetic risk scores for coronary heart disease prediction. Arteriosclerosis, thrombosis, and vascular biology, 33(9), 2267–72.
8. Jabbar M A, Deekshatulu B L & Chandra P (2013, March). Heart disease prediction using lazy associative classification. In 2013 International Mutli-Conference on Automation, Computing,Communication, Control and Compressed Sensing (iMac4s) (pp. 40–6). IEEE.
9. Dangare Chaitrali S and Sulabha S Apte. "Improved study of heart disease prediction system using data mining classification techniques." International Journal of Computer Applications 47.10 (2012): 44–8
10. Soni Jyoti. "Predictive data mining for medical diagnosis: An overview of heart disease prediction." International Journal of Computer Applications 17.8 (2011): 43–8.
11. Chen A H, Huang S Y, Hong P S, Cheng C H & Lin E J (2011, September). HDPS: Heart disease prediction system. In 2011 Computing in Cardiology (pp. 557–60). IEEE.
12. Parthiban, Latha and R Subramanian. "Intelligent heart disease prediction system using CAN-FIS and genetic algorithm." International Journal of Biological, Biomedical and Medical Sciences 3.3 (2008).
13. Wolgast G, Ehrenborg C, Israelsson A, Helander J, Johansson E & Manefjord H (2016). Wireless body area network for heart attack detection [Education Corner]. IEEE antennas and propagation magazine, 58(5), 84–92
14. Patel S & Chauhan Y (2014). Heart attack detection and medical attention using motion sensing device -kinect. International Journal of Scientific and Research Publications, 4(1), 1–4.

15. Zhang Y, Fogoros R, Thompson J, Kenknight B H, Pederson M J, Patangay A & Mazar S T (2011). U.S. Patent No. 8,014,863. Washington, DC: U.S. Patent and Trademark Office.
16. Raihan M, Mondal S, More A, Sagor M O F, Sikder G, Majumder M A & Ghosh K (2016, December). Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design. In 2016 19th International Conference on Computer and Information Technology (ICCIT) (pp. 299–303). IEEE.
17. Buechler K F & McPherson P H (1999). U.S. Patent No. 5,947,124. Washington, DC: U.S. Patent and Trademark Office.