International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 05 | May 2024 www.irjet.net p-ISSN: 2395-0072

A Comparative Analysis of Machine Learning Based Models for heart Disease Prediction

Ashima Gogoi, Sonali Mondal*, Biswajit Das

Department of Computer Science, Arunachal University of Studies, Namsai, Arunachal Pradesh, India

---

Abstract -: *Daily the instances of coronary heart illnesses are growing at a rapid charge and it's very essential and concerning to predict one of these illnesses beforehand. This diagnosis is a tough undertaking i.e. it has to be completed precisely and effectively. The studies paper specially makes a specialty of which affected person is much more likely to have a heart ailment based on various scientific attributes. We prepared a coronary heart disorder prediction device to predict whether the patient is probably to be identified with a heart disorder or now not the usage of the clinical history of the patient. We used one-of-a-kind algorithms of gadget learning along with logistic regression and KNN to predict and classify the patient with heart ailment. A quite useful method changed into used to alter how the model may be used to enhance the accuracy of prediction of heart assault in any individual. The strength of the proposed model become quite pleasurable and changed into capable of predict proof of having a heart disorder in a selected man or woman with the aid of the use of KNN and Logistic Regression which confirmed an excellent accuracy in contrast to the formerly used classifier which includes naive bayes and so forth. So, a quiet vast quantity of stress has been lift off via using the given version in finding the possibility of the classifier to correctly and accurately become aware of the coronary heart disease. The Given coronary heart sickness prediction device enhances hospital therapy and decreases the cost. This venture offers us sizeable expertise that can help us are expecting the patients with coronary heart disorder it's far applied on the pynb format.*

Key Words: **Heart Disease, Machine Learning, Prediction**

1.INTRODUCTION

Healthcare is one of the number one cognizance for humanity. according to WHO suggestions, right fitness is the essential right for people. it's far taken into consideration that suitable health care services have to be to be had for ordinary checkup of 1's fitness. almost 31% of all deaths are because of coronary heart related disorder in everywhere in the world. Early detection [1] and remedy of numerous coronary heart diseases could be very complicated, in particular in developing nations, due to the dearth of diagnostic centers and certified docs and different sources that affect the accurate diagnosis of heart sickness. With this challenge, in recent instances computer era and machine learning techniques are being used to make medical aid software as a support device for early analysis of coronary heart disease. identity of any coronary heart related infection at number one degree can lessen the loss of life threat. numerous ML strategies are utilized in medical statistics to recognize the sample of facts and making prediction from them. Healthcare statistics are typically big in volumes and complex in shape. ML algorithms are successful to handle the huge records and mine them to discover the significant statistics. device gaining knowledge of algorithms research from past information and do prediction on real time information. This kind of ML framework for coronary infection expectation can inspire cardiologists in taking faster moves so greater patients can get drugs within a shorter timeframe, thus saving large quantity of lives. system mastering is a branch of AI research [2] and has grow to be a totally famous aspect of statistics science. The system gaining knowledge of algorithms are designed to perform a huge wide variety of obligations consisting of prediction, classification, decision making etc. To analyze the ML algorithms, training facts is needed. After the gaining knowledge of segment, a version is produced that's taken into consideration as an output of ML set of rules. This version is then examined and demonstrated on a fixed of unseen actual time check dataset. The final accuracy of the model is then in comparison with the actual fee, which justify the overall correctness of expected end result. lots of efforts has already been accomplished to expect the coronary heart sickness the use of the ML algorithms by way of authors [3-5], but this is a further effort to do the experiment on benchmarking UCI heart sickness prediction dataset whilst comparing the four popular ML technique to test the maximum correct ML approach. The paper is based as follows: section 2 contains the details of ML techniques used in these studies paintings. segment three shows the methodology, segment four summaries with end result of this work and segment 5 list out the belief.

2. Literature Review

Sharma et all stated in their paper that, our goal in this study is to create a machine learning model that can predict cardiac disease based on a variety of factors. A benchmark dataset of 14 distinct heart disease related factors was employed by us, derived from the UCI Heart disease prediction. To train and assess our model, we used a variety of machine learning techniques, including logistic Regression, decision trees, random forest, and support

vector machines. Our goal is to determine the relationships between these factors and precisely estimate the risk of heart disease [1]. Ali et.al stated that According To our research the random forest algorithm made the best accuracy in predicting cardiac illness which make it a useful tool for doctor to use as a clinic decision support system based on the number of variables we can accurately forecast the incidence and severity of heart disease by applying machine learning algorithm. furthermore, our research use of artificial neural networks in conjunction with logistic regression analysis has improved the precision and effectiveness of our model even further. By enabling early identification and intervention for heart disease this paradigm has the potential to save lives and significantly improve the standard of patient care in the healthcare sector [2]. Ingelsson et al stated in the early stages of cardiac disease, medical practitioners can enhance patient outcomes and make well-informed decisions by utilizing machine learning. This study emphasizes the potential advantages for the healthcare industry as well as the importance of machine learning techniques in the prediction and diagnosis of cardiac disease. This can help with medical decision-making and significantly increase the specificity and accuracy of predictions of cardiovascular disease. By utilizing machine learning, we can increase the precision and effectiveness of cardiac disease prediction, allowing for prompt patient interventions and individualized treatment regimens [3]. Pichon et al., stated that heart disease diagnosis and prognosis accuracy have been demonstrated by machine learning algorithms. The prognosis and diagnosis of cardiac disease have been completely transformed by the application of machine learning techniques in the medical field. All things considered, the use of machine learning methods to the prediction of cardiovascular illness holds potential for raising the precision and effectiveness of heart disease diagnosis. When diagnosing and treating cardiac disease, these algorithms can help doctors make decisions more quickly and intelligently. Better patient outcomes result from more rapid and accurate therapies made possible by this [4]. Singh and Kumar (2020) conducted a comprehensive study on heart disease prediction utilizing machine learning algorithms [5]. Their research aimed to address the critical need for precise and accurate prediction systems in the context of heart-related diseases, given the significant role of the heart in living organisms and the increasing prevalence of heart-related deaths worldwide [5]. The study focused on evaluating the performance of various machine learning algorithms in predicting heart disease, namely k-nearest neighbor, decision tree, linear regression, and support vector machine (SVM). By utilizing the UCI repository dataset for training and testing, Singh and Kumar aimed to provide insights into the effectiveness of these algorithms for this particular task [5]. Jindal et al. (2021),[6] conducted a project on heart disease prediction using Logistic Regression, KNN, and Random Forest Classifier. Their objective was to improve prediction accuracy, achieving 87.5% accuracy with these techniques. They utilized a dataset from the UCI repository, analyzing 14 medical attributes to classify patients as at risk or not. KNN proved most efficient with 88.52% accuracy. Their approach is highlighted for its cost efficiency and potential in healthcare settings [6]. "Heart disease prediction using machine learning techniques," Shah, D., Patel, S., and Bharti, S.K. [7] aimed to define effective data mining techniques for accurate heart disease prediction with minimal attributes [7]. They focused on 14 essential attributes and applied four classification techniques: K-nearest neighbor, Naive Bayes, decision tree, and random forest. After preprocessing the data, K-nearest neighbor, Naïve Bayes, and random forest yielded the best results. K-nearest neighbor (k = 7) achieved the highest accuracy among the four algorithms [7]. Arumugam, K., Naved, M., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. [8] explored the intersection of data mining and healthcare, originating from database statistics and valuable in assessing medical treatment effectiveness [8]. Their focus was on diabetes-related heart disease, a condition affecting diabetics due to insulin production or utilization issues. Heart disease, or cardiovascular disease, encompasses conditions affecting the heart or blood vessels [8]. Salhi, D. E., Tari, A., & Kechadi, M. T. (2021) [9] conducted research on heart disease prediction from a data analytics perspective. They noted the recent emergence of heart disease prediction due to the availability of data and highlighted diverse approaches adopted by other researchers [9]. Kavitha et al.,[10] introduced a hybrid machine learning model for heart disease prediction at ICICT 2021. They leveraged the Cleveland heart disease dataset and tested three algorithms: Random Forest, Decision Tree, and a hybrid model. Results showed 88.7% accuracy, demonstrating the potential for early detection of cardiovascular diseases [10].

3. Objective

The primary objectives of our project represent:

- The Project aims to predict heart disease based on given data.
- It aims to save time such the disease can be predicted earlier.
- Achieving high accuracy in predicting the presence or absence of heart disease to ensure reliable risk assessment and clinical decision-making.

The Above point represents the main prospective of the approach work.

4. Methodology

The improved methodology will produce more accurate results and superior model performance. The dataset used for this project purpose was the Public Health Dataset. It contains predicted attributes, but all published experiments refer to using a subset of 14 of them. The “target” field refers to the presence of heart disease in the patient. It is integer-

valued 0 = no disease and 1 = disease. The first four rows and all the dataset features are shown without any preprocessing. Now the attributes which are used in this project are described as follows and for what they are used or resemble:

- Age: Age of the patient in years.
- Sex: 1 for male, 0 for female.
- Cp: Chest pain type.
- Trestbps: Resting blood pressure (in mm Hg on admission to the hospital). The normal range is 120/80. Elevated readings may indicate the need for lifestyle changes.
- Chol: Serum cholesterol level indicating the amount of triglycerides present. It should ideally be less than 170 mg/dL (may vary between labs).
- Fbs: Fasting blood sugar level, where 1 indicates a level larger than 120 mg/dL. Levels between 100 to 125 mg/dL are considered prediabetic.
- Restecg: Resting electrocardiographic results.
- Thalach: Maximum heart rate achieved. The maximum heart rate is calculated as 220 minus the patient's age.
- Exang: Exercise-induced angina, where 1 indicates presence and 0 indicates absence. Angina is a symptom of coronary artery disease.
- Old peak: ST depression induced by exercise relative to rest.
- Slope: The slope of the peak exercise ST segment.
- Ca: Number of major vessels (ranging from 0 to 3) colored by fluoroscopy.
- Thal: Thalassemia status, with values indicating normal (3), fixed defects (6), or reversible defects (7).
- Target (T): 0 indicates no disease, while 1 indicates the presence of heart disease (angiographic disease status).

5. Processing of Data

The dataset does not have any null values. But many outliers needed to be handled properly, and also the dataset is not properly distributed. Two approaches were used. One without outliers and feature selection process and directly applying the data to the machine learning algorithms, and the results which were achieved were not promising. But after using the normal distribution of dataset for overcoming the overfitting problem and then applying Isolation Forest for the outlier's detection, the results achieved are quite promising. Various plotting techniques were used for checking the skewness of the data, outlier detection, and the distribution of the data. All these preprocessing techniques play an important role when passing the data for classification or prediction purposes.

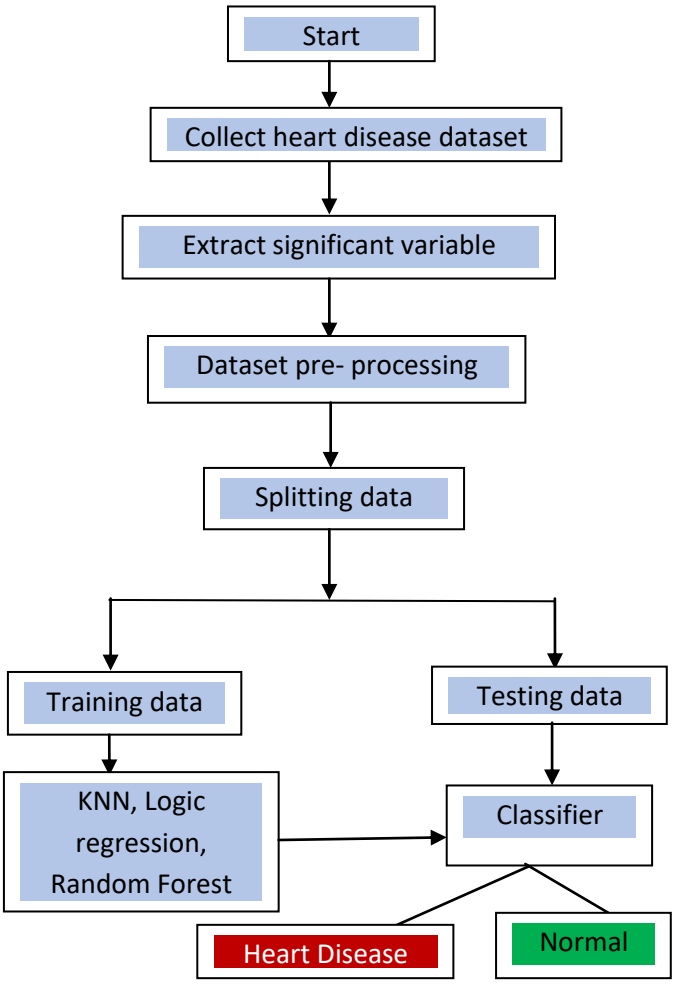

Fig1: Flow chart of Prediction System

6. RESULT AND DISCUSSION

The calculation of three different algorithms— KNN, SVM, and RF— for significant heart disease prediction.
Output is the very important to outcome so that identity to heart disease prediction using machine learning.

- **Result**
  The performance of various machine learning models in predicting heart disease was evaluated using the dataset. The following accuracy levels were achieved:

Support Vector Machine (SVM): 86.88%

K-Nearest Neighbors (KNN): 85.25%

Algorithms	Accuracy
K-Nearest Neighbor	85.24
Support Vector Machine	86.88

Table1: Accuracy Comparison of The Models

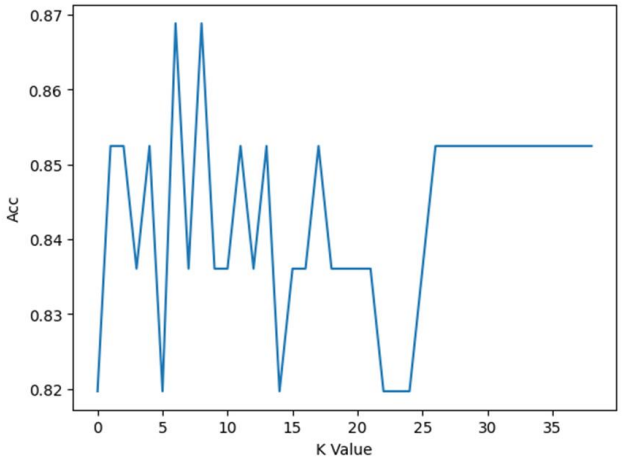

Fig2: K Value Accuracy

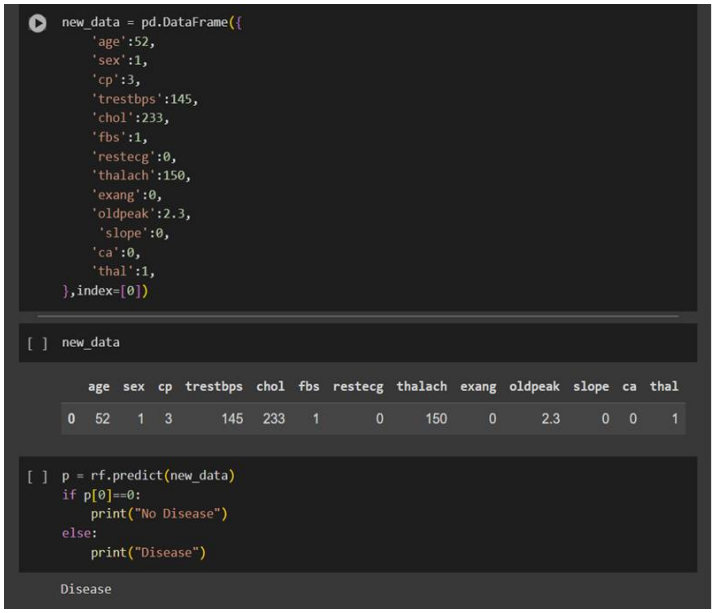

Fig3: Diagnosis with patient's data

7. Discussion

The results of our study demonstrate the efficacy of machine learning algorithms in predicting heart disease based on patient attributes. Logistic Regression exhibited the highest accuracy among the models tested, achieving a rate of 90.16%. This model's simplicity and interpretability make it a suitable choice for initial screening and diagnosis tasks. Support Vector Machine and Random Forest achieved comparable accuracies, both exceeding 86%. These models offer robustness to noise and are effective in handling complex relationships within the data. However, their interpretability may be limited compared to Logistic Regression. K-Nearest Neighbors and Gradient Boosting also performed well, with accuracies around 85%. KNN's performance improved with increasing values of K, indicating the importance of selecting an optimal value for the number of neighbors. Gradient Boosting, on the other hand, utilizes an ensemble of weak learners to improve predictive performance, making it particularly adept at handling heterogeneous data. Decision Trees exhibited the lowest accuracy among the models tested, achieving a rate of 80.33%. While Decision Trees are easy to interpret and visualize, they are prone to overfitting, especially on complex datasets like the one used in this study.

8. Conclusion and Future Work

In conclusion, our results highlight the potential of machine learning models in aiding the early detection and diagnosis of heart disease. The choice of model depends on various factors including accuracy, interpretability, and computational complexity. Further research could explore ensemble methods or deep learning architectures to improve predictive performance and generalization on larger and more diverse datasets. Additionally, validation of these models on independent datasets and clinical trials is essential to assess their real-world utility and impact on patient outcomes.

9. Reference

[1] Sharma, V., Yadav, S. and Gupta, M., 2020, December. Heart disease prediction using machine learning techniques. In 2020 2nd international conference on advances in computing, communication control and networking (ICACCCN) (pp.177181). IEEE. (DOI:10.1109/ICACCCN51052.2020.9362842)

[2] Ali, M.M., Paul, B.K., Ahmed, K., Bui, F.M., Quinn, J.M. and Moni, M.A., 2021. Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison. Computers in Biology and Medicine, 136, p.104672.
(DOI:10.1016/j.compbiomed.2021.104672)

[3] Ingelsson, E., Schaefer, E.J., Contois, J.H., McNamara, J.R., Sullivan, L., Keyes, M.J., Pencina, M.J., Schoonmaker, C., Wilson, P.W., D'Agostino, R.B. and Vasan, R.S., 2007. Clinical utility of different lipid measures for prediction of coronary heart disease in men and women. Jama, 298(7), pp.776-785.
(DOI: 10.1001/jama.298.7.776)

[4] Pischon, T., Girman, C.J., Sacks, F.M., Rifai, N., Stampfer, M.J. and Rimm, E.B., 2005. Non–high-density lipoprotein cholesterol and apolipoprotein B in the prediction of coronary heart disease in men. Circulation, 112(22), pp.3375-3383.
(DOI: 10.1161/circulationaha.104.532499)

[5] Singh, A., & Kumar, R. (2020, February). Heart disease prediction using machine learning algorithms. In 2020international conference on electrical and electronics engineering (ICE3) (pp. 452-457). IEEE.

(DOI: 10.1109/ICE348803.2020.9122958)

[6] Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). Heart disease prediction using machine learning algorithms. In IOP conference series: materials science and engineering (Vol. 1022, No. 1, p. 012072). IOP Publishing. (DOI: 10.1088/1757-899X/1022/1/012072)

[7] Shah, D., Patel, S., & Bharti, S. K. (2020). Heart disease prediction using machine learning techniques. SN Computer Science, 1(6), 345. (DOI: 10.1007/s42979-020-00365-y)

[8] Arumugam, K., Naved, M., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. (2023). Multiple disease prediction using Machine learning algorithms. Materials Today: Proceedings, 80, 3682-3685. (DOI: 10.1016/j.matpr.2021.07.361)

[9] Salhi, D. E., Tari, A., & Kechadi, M. T. (2021). Using machine learning for heart disease prediction. In Advances in Computing Systems and Applications: Proceedings of the 4th Conference on Computing Systems and Applications (pp. 70-81). Springer International Publishing. (DOI: 10.1007/978-3-030-69418-0_7)

[10] Kavitha, M., Gnaneswar, G., Dinesh, R., Sai, Y. R., & Suraj, R. S. (2021, January). Heart disease prediction using hybrid machine learning model. In 2021 6th international conference on inventive computation technologies (ICICT) (pp. 1329-1333). IEEE. (DOI: 10.1109/ICICT50816.2021.9358597)