International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

# HEART DISEASE PREDICTION USING MACHINE LEARNING

T.SWATHI, U.JYOTHI, P.CHARITHA, M.SEVITHA, S.POOJA, G.SAI YASASWINI

**BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE & ENGINEERING(AI & ML)**

MALLA REDDY UNIVERSITY, HYDERABAD

**Under the guidance of**

Prof. K. Senthil Kumar

**ABSTRACT:** Heart disease is a major cause of death worldwide, making early diagnosis and prevention essential. Predictive models have gained significant attention in recent years, with several algorithms being employed to develop these models. However, there are challenges in implementing heart disease prediction models, including data quality, model accuracy, ethical concerns, and limited data. Therefore, this project aims to develop a heart disease prediction model and analyse different algorithms used in disease prediction. In order to increase the predictive accuracy of machine learning algorithms, this study compares six algorithms, including KNN (K-Nearest Neighbour), Decision Tree, Random Forest, Support Vector Machines, Logistic Regression, and Neural Network. 13 attributes, including age, sex, and cholesterol, are used, and ensemble methods like boosting and bagging are used. The accuracy, recall, f1 score, and precision of each algorithm are calculated to determine the most accurate model. Additionally, this study identifies the limitations of heart disease prediction models and their implications for patient diagnosis and treatment, by developing and analysing heart disease prediction models. In conclusion, while heart disease prediction models have the potential to be financially feasible and be useful in the future, their current limitations and challenges mean that they cannot be relied upon as the sole means of diagnosis or treatment decisions

**Key Words:** Heart Diseases, Machine Learning Algorithms, Logistic Regression, Random Forest, Decision Tree.

**1.INTRODUCTION:** Heart disease are the leading cause of death globally, accounting for approximately 18 million deaths annually, which is about 32% of all global deaths. According to the World Health Organization (WHO), approximately three-quarters of CVD deaths occur in low- to middle-income there has been increasing interest in developing predictive models for early detection of heart disease. Due to the availability of large amounts of health data and advancements in machine learning, there has been a growing interest in utilizing machine learning algorithms for the purpose of early detection of heart disease. Various algorithms, such as logistic regression, KNN (K-Nearest Neighbours), decision tree, support vector machines, random forest, and neural network, have been employed to develop these models. Despite the potential benefits of machine learning algorithms in predicting heart disease, several challenges remain. These include data quality, model accuracy, ethical concerns, and limited data. For instance, several important risk factors such as family history, lifestyle, smoking status, ranking status, and medication history cannot be fully accounted for by current algorithms. To address these limitations, research efforts are underway to improve the accuracy and effectiveness of machine learning algorithms in predicting heart

disease. One approach involves incorporating additional data sources, such as genetic and environmental factors, to improve the predictive power of the models. Another approach involves developing interpretable machine learning models that can provide insight into the underlying mechanisms of heart disease. The results suggest that while machine learning algorithms have the potential to be financially feasible and useful, there are significant limitations to their current capabilities, including their inability to account for certain risk factors such as family history, lifestyle, smoking status, drinking status, and medication history. As such, while these models may be useful for calculating the probability of heart disease, they should not be used solely for diagnosis or treatment decisions. The findings of this project highlight the need for continued research and development in this field, with a focus on addressing the current limitations of machine learning algorithms in predicting heart disease.

**2.Description :**Heart disease prediction involves using various methods to assess an individual's risk of developing cardiovascular issues. It typically includes a combination of medical history evaluation, risk factor assessment, and sometimes predictive models based on machine learning algorithms. Here's a breakdown of how it generally works. Heart disease prediction is crucial for early intervention and prevention. Identifying high-risk individuals enables healthcare providers to implement strategies for lifestyle modification, prescribe medications, and schedule regular monitoring to reduce the risk of cardiovascular events. By amalgamating medical expertise with technological advancements, heart disease prediction aims to mitigate the prevalence and impact of cardiovascular issues by identifying risks at an early stage, thereby promoting better health outcomes. Sure, heart disease prediction involves a multi-faceted approach that combines various methods and tools to assess an individual's risk of developing cardiovascular issues. Let's delve deeper into each component. In this paper, the heart disease dataset of UCI repository is taken and subjected to various classification and clustering algorithms using python. The main focus is to target all possible combinations of the attributes against various algorithms. Then of all the techniques it is the technique that works the best to predict the heart disease at an early stage is identified. Implementing 3 algorithms such as Decision tree, Random forest and Logistic regression would make it easier to identify and sort out the disease. Dataset is used to classify and train the model. After training the model, the most accurate and successful algorithm was later used to predict the disease.

**3.Literature Review:** There are numerous works has been done related to disease prediction systems using different data mining techniques and machine learning algorithms in medical centres .K. Pola et al, [7] proposed Prediction of Heart Disease using Multiple Regression Model and It proves that Multiple Linear Regression is appropriate for predicting heart disease chance. The work is performed using training data set consists of 3000 instances with 13 different attributes which has mentioned earlier. The data set is divided into two parts that is 70% of the data are used for training and 30% used for testing. Based on the results, it is clear that the classification accuracy of Regression algorithm is better compared to other algorithms. Marjia et al, [8] developed heart disease prediction using K Star, j48, SMO, and Bayes Net and Multilayer perception using WEKA software. Based on performance from different factor SMO and Bayes Net achieve optimum performance than K Star, Multilayer perception and J48 techniques using k-fold cross validation. The accuracy performances achieved by those algorithms are still not satisfactory. Therefore, the accuracy's performance is improved more to give better decision to diagnosis disease. S. Seema et al,[9] focuses on techniques that can predict chronic disease by mining the data containing in historical health records using Naïve Bayes,

Decision tree, Support Vector Machine(SVM) and Artificial Neural Network(ANN). A comparative study is performed on classifiers to measure the better performance on an accurate rate. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy. S. Seema et al,[9] focuses on techniques that can predict chronic disease by mining the data containing in historical health records using Naïve Bayes, Decision tree, Support Vector Machine(SVM) and Artificial Neural Network(ANN). A comparative study is performed on classifiers to measure the better performance on an accurate rate. From this experiment, SVM gives highest accuracy rate, whereas for diabetes Naïve Bayes gives the highest accuracy.

**4.Proposed Approach:** People all across the globe now live luxury lives, and they work like machines to gain a lot of money and recognition. People fail to care on their health because of their hectic schedules. As a result, the food they eat has changed, as has their way of life. Because of the strain and stress in their lives, they develop high blood pressure, diabetes, and a variety of other disorders at an early age. All of these factors contribute to the onset of heart disease. We are employing the Logistic Regression approach to develop an effective heart attack prediction system in this system. We can provide input to the system in the form of a CSV file or by entering it manually. Logistic Regression is an algorithm that is applied after accepting input. The operation is carried out after accessing the data set, and an effective heart attack level is created. The suggested method would incorporate some more criteria relevant to heart attacks, such as weight, age, and priority levels, after consultation with specialist doctors and medical professionals. The heart attack prediction system is intended to assist in identifying different risk levels of heart attack, such as normal, low, or high, as well as providing prescription data based on the expected outcome. The user can also register and login or else user can quickly predict the heart disease by just one click, but after quickly predicting the result the user's data will not be stored and neither user can access his/her previous record. After Successful registration and login the user can feed the values on the website and click on predict, on the very next screen the user can see the result. It will show the probability of heart disease in the percentage. If the percentage is greater than 60%, it will display "Your chance of having heart disease is high" and will provide prevention and symptoms as well as a doctor's contact information for further assistance; if the percentage is less than 60%, it will display "Your chance of having heart disease is low." The user data is saved in the database using their name and username, and any doctor who wants to add his/her details can do so using the admin panel. Our major goal is to save the prior record so that the user may see how much progress has been made by the user.

**5.Methodology:** This paper shows the analysis of various machine learning algorithms, the algorithms that are used are SVM, Logistic Regression and Random Forest Classifiers, which can be helpful for practitioners or medical analysts for accurately diagnosing Heart Disease. This paperwork includes examining the journals, published paper and the data of cardiovascular disease of the recent times. Methodology gives a framework for the proposed model. The methodology is a process which includes steps that transform given data into recognized data patterns for the knowledge of the users. The proposed methodology includes steps, where first step is referred as the collection of the data than in second stage it extracts significant values than the 3rd is the processing stage where we explore the data. Data processing deals with the missing values, cleaning of data and normalization depending on algorithms used . After processing of data, classifier is used to classify the processed data the classifier used in the proposed model are SVM. Logistic Regression, Random Forest Classifier. Finally, the proposed model is undertaken, where

we evaluated our model on the basis of accuracy and performance using various performance metrics. Here in this model, an effective Heart Disease Prediction System (EHDPS) has been developed using different classifiers. This model uses 13 medical parameters such as chest pain, fasting sugar, blood pressure, cholesterol, age, sex etc. for prediction .

**6.References:** Avinash , Pavan Kumar T, ”Heart Disease Prediction Using Effective Machine Learning Techniques”, International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019. T. Nagamani, S. Logeswari, B .Gomathy,” Heart Disease Prediction using Data Mining with Algorithm”, International Journal of Innovative Technology and Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019. Fahd Saleh Alotaibi,” Implementation of Machine Learning Model to Predict Heart Failure Disease”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019. Anjan Nikhil , Sai Deepak , Ramya G Franklin, ”Design And Implementation Heart Disease Prediction Using Bayesian”, International Conference on Trends in Electronics and Information(ICOEI 2019). Theresa Princy R,J. Thomas ,’Human heart Disease Prediction System using Data Mining Techniques’, International Conference on Circuit Power and Computing Technologies,Bangalore,2016. Nagaraj M, Chethan C ,Basavaraj S Pol., ’Prediction Of Heart Disease using Machine Learning’, International journal Of Recent Technology and Engineering,8,(2S10), pp 474-477, 2019. afar Alzubi, Anand Nayyar, Akshi Kumar. "Machine Learning from Theory to Algorithms: An Overview", Journal of Physics: Conference Series, 2018 Fajr Ibrahem . and Mamoon Younes ‘Analysis and classification of heart diseases using heartbeat features and machine learning algorithms ’,Journal Of Big Data,2019;6:81. Internet source [Online]

**7.Future Enhancement**: To enhance heart disease prediction using machine learning, consider incorporating advanced algorithms like ensemble methods (e.g., Random Forests), feature engineering for relevant variables, and fine-tuning model hyperparameters. Additionally, explore integrating data from diverse sources, such as genetic information or wearable devices, to improve prediction accuracy. Regularly update the model with new data to ensure its relevance and effectiveness.

**8.Research Questions:**

1. How effective are machine learning algorithms in predicting the risk of heart disease based on patient demographic data?

2. Can the integration of genetic information improve the accuracy of machine learning models for heart disease prediction?

3. What is the comparative performance of different machine learning algorithms in predicting specific types of heart diseases?

4. To what extent does incorporating lifestyle factors, such as diet and exercise, enhance the predictive capabilities of machine learning models for heart disease?

5. How does the size and diversity of the dataset impact the generalization ability of machine learning models for heart disease prediction?

6. What role do novel biomarkers play in improving the early detection and prediction of heart disease through machine learning?

7. How well do interpretable machine learning models perform in providing insights into the factors contributing to heart disease risk?

**9.Hypothesis**: A hypothesis for predicting heart disease using machine learning could be: "By leveraging a dataset of diverse patient health records and applying advanced machine learning algorithms, we hypothesize that we can develop an accurate predictive model for identifying

individuals at risk of heart disease based on key risk factors such as age, blood pressure, cholesterol levels, and lifestyle factors."

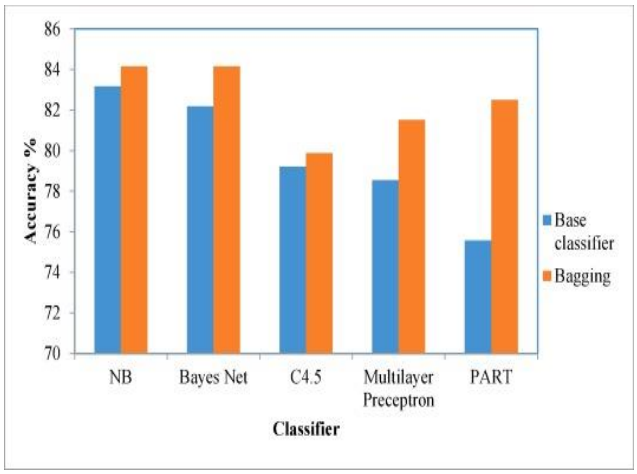
**10.Problem Background:** Heart disease remains a leading cause of morbidity and mortality globally, necessitating advanced tools for early detection and prevention. The conventional approach to identifying individuals at risk relies on established risk factors, medical history, and clinical assessments. However, machine learning presents an opportunity to enhance predictive accuracy by leveraging vast and diverse datasets. The background of the problem lies in the complexity of factors contributing to heart disease, encompassing genetic predispositions, lifestyle choices, and demographic characteristics. Traditional risk assessment methods often struggle to capture the intricate interplay of these variables. Machine learning models offer the potential to discern subtle patterns and interactions within these multifaceted datasets, thereby improving the precision of risk predictions. The significance of this is underscored by the potential to shift from a reactive to a proactive healthcare paradigm. Early identification of individuals at risk can enable timely interventions, lifestyle modifications, and personalized healthcare strategies. However, the deployment of machine learning in healthcare necessitates addressing ethical considerations, ensuring model interpretability, and maintaining patient privacy. In essence, the problem background revolves around leveraging machine learning to revolutionize heart disease prediction, providing a more nuanced and individualized approach to preventive healthcare. This aligns with the broader objective of enhancing public health outcomes and reducing the societal burden imposed by heart-related ailments.

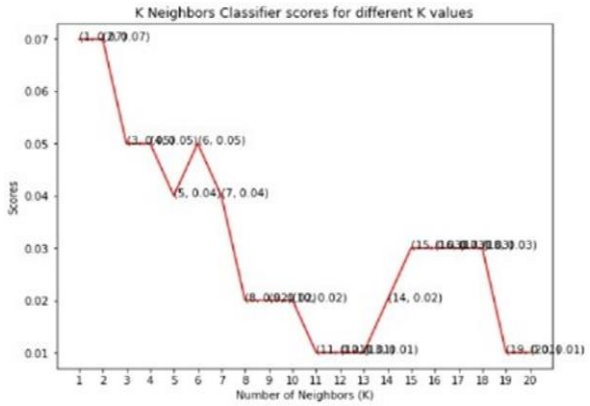
**11.Acknowledgement:** Acknowledging medical professionals, clinicians, and researchers who shared their expertise, provided guidance, and offered valuable insights into the clinical aspects of heart disease. Their input is crucial for ensuring the accuracy and relevance of the predtive model. Expressing gratitude to institutions, research groups, or individuals who provided access to datasets, patient records, or research materials essential for training and validating the predictive algorithms. Detailing their contribution enhances transparency and acknowledges their role in the project. Recognizing the collaboration and guidance received from experts in data science, machine learning, and statistics who contributed their knowledge and skills to refine the predictive model. Their input might include refining algorithms, conducting analyses, or offering technical expertise. Acknowledging the financial support and resources provided by funding agencies, research grants, or institutional support. This acknowledgment demonstrates appreciation for the resources that enabled the project's execution, including access to facilities, equipment, and personnel. Thanking individuals who participated in beta testing, offering their time and feedback, thereby contributing to the improvement and validation of the predictive model. Recognizing the end users who will ultimately benefit from the tool emphasizes its practical application and impact in clinical settings. Emphasizing the collective effort and expressing sincere gratitude for the invaluable contributions. Highlighting how each contribution impacted the project's development and eventual success adds depth to the acknowledgment.

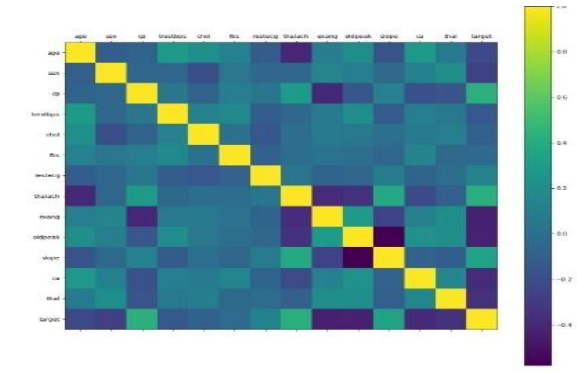
**12.Advantages:** Timely Intervention Predictive models enable the early identification of individuals at risk of heart disease, allowing for timely interventions and preventive measures. Focusing on prevention reduces the likelihood of cardiovascular events, potentially improving health outcomes and quality of life for individuals at risk. Preventive measures and early interventions may lead to reduced healthcare costs associated with managing advanced stages of heart disease. Implementing a robust heart disease prediction system not only benefits individuals by enabling personalized care but also has broader implications for healthcare systems, resource

allocation, public advancements in preventive medicine.

**13.Experiment Result:**



**14.Conclusion:** The development and implementation of a heart disease prediction system stand as a pivotal advancement in healthcare. By harnessing the power of data analytics, machine learning, and medical expertise, this predictive model offers a proactive approach to managing cardiovascular health. The ability to identify individuals at risk of heart disease early on enables personalized interventions, ultimately leading to improved health outcomes and a reduction in the burden of cardiovascular diseases. This predictive system not only aids healthcare providers in making informed decisions but also empowers individuals to take proactive steps towards better heart health. Its impact extends beyond individual care, influencing resource allocation, public health strategies, and contributing to ongoing advancements in preventive medicine. As we continue to refine and improve these predictive models, incorporating feedback, updating algorithms, and ensuring ethical and equitable access to these tools, the goal remains clear: to reduce the prevalence of heart disease, enhance the quality of care, and ultimately, save lives .The journey towards mitigating the impact of heart disease is ongoing, and the development of robust predictive systems stands as a beacon of hope in the realm of preventive healthcare, offering a path towards a healthier future for individuals and communities alike

**15.References**: [I] Anooj,P.K.2012.Clinical decision support system: risk level prediction of heart disease using weighted fuzzy rules. J.KingSaudUniv.-ComputerInf.Sci.24(1),27-40.

[2] Bhatla.N.Jyoti,K.,2012.Ananalysis of heart disease prediction using different data mining techniques. Int. J. Eng. 1(8),1-4.

[3] Chaurasia, V.,Pal.S.,2013.Early prediction of heart diseases using data mining techniques. Carib, J. SciTech.1,208-217.

[4] Dey .A. Singh...Singh .N., 2016. Analysis of supervised machine learning algorithms for heart disease prediction with reduced number of attributes using principal component analysis .Analysis 140(2),27-31.

[5] Khemphila.A.,Boonjing, V.,2011.Heart disease classification using neural network and feature selection.In:21" International Conference Engineering(ICSEng) IEEE, Las Vegas,pp.406-409. on Systems