



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 4, April 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com



Heart Disease Prediction System using Machine Learning

Dr. Sunil B. Wankhade ¹, Gaurav Jain ², Shrushti Jadhav ³, Neeraj Gupta ⁴, Sagar Dapurkar ⁵.

Head of Department, Department of Information Technology, Rajiv Gandhi Institute of Technology (Mumbai University), Mumbai, Maharashtra, India¹

U.G. Student, Department of Information Technology, Rajiv Gandhi Institute of Technology (Mumbai University), Mumbai, Maharashtra, India^{2,3,4,5}

ABSTRACT: The heart is the fundamental and most important part of the human body. The life of a human being mostly depends on the heart. The heart is a very complex part of the human body; hence, understanding it precisely is a very big deal for doctors as well as researchers. Sometimes it gets very difficult for doctors to understand the problems related to the heart of the patient, resulting in the criticality of the patient's health. In Asia, 18.9% of the deaths recorded were due to cardiac arrest, of which 20% were male and 17.8% were female. One in every five deaths nowadays is caused by heart disease. Hence, we can understand the seriousness of cardiological diseases. That's why, to overcome these problems, diseases related to the heart can be predicted by using artificial intelligence, deep learning, machine learning, and neural networks. It invites automation in the health industry for predicting patients' health and also helps doctors monitor the issues in a more precise and easy way. So, we are building a system for predicting heart diseases with the help of machine learning and deep learning. There are various algorithms for building such models, such as KNN (K-Nearest Neighbors), Decision Tree, SVM algorithm, Random Forest, K-means, Logistic Regression, and many more. But here we are using KNN (K-nearest neighbors) for building the model for heart disease prediction. Our goal is to improve the model's accuracy and provide a very precise and correct outcome of whether the person will have heart disease or not by using inputs from their daily life habits. We have used a built-in dataset from Kaggle. It consists of around 18 attributes such as gender, BMI (body mass index), blood pressure, smoker, drinker, etc. This paper proposes a model for heart disease prediction using KNN (K-Nearest Neighbors) algorithm and deployed it in a web application.

KEYWORDS: Heart disease, Prediction, Machine learning, K- nearest neighbors

I. INTRODUCTION

Heart disease is the primary cause of death of humankind nowadays. It is reported in USA that cardiovascular death is about one-third of overall death. There are various mining algorithms for building such models, such as KNN (K-Nearest Neighbors), Decision Tree, SVM algorithm, Random Forest, K-means, Logistic Regression, and many more. But here we are using KNN (K-nearest neighbors) for building the model for heart disease prediction.

Heart disease prediction system using machine learning is a system that uses data on individuals to predict the likelihood of them developing heart disease. The system uses the K-Nearest Neighbors (KNN) algorithm, which is a simple yet powerful machine learning algorithm that is used for classification and regression tasks. The KNN algorithm works by first training the model on a dataset of known cases, which includes individuals who have and have not developed heart disease. The system then takes in new data from an individual, such as their age, sex, blood pressure, and other medical indicators, and uses the KNN algorithm to find the Heart disease of an individuals in the training dataset that are most similar to the new data point. The algorithm then predicts whether or not the individual is likely to develop heart disease based on the outcome of the K-nearest neighbors in the training dataset. The prediction is made based on the most common outcome among the K-nearest neighbors.

II. RELATED WORK

Several studies have been conducted using machine learning data to predict heart disease. Researchers have used different data mining techniques to achieve different levels of accuracy. Avinash Goland and colleagues studied heart



disease classification using several ML algorithms, including Decision Tree, ANN, and K-Means. They found that Decision Tree has the highest accuracy and suggested that it can be further optimized by combining the technique and parameter tuning. T Nagamani and his team proposed a system that uses data mining techniques and the MapReduce algorithm to achieve greater accuracy compared to artificial fuzzy neural networks. Fahd Saleh Alotaibi designed an ML model that compared five different algorithms and found that Decision Tree was the most accurate. Anjan Nikhil Repaka and colleagues proposed a system using NB techniques for classification and AES algorithms for secure data transmission. Teresa Prince R. conducted research on various classification algorithms including Naive Bayes, KNN, Decision Tree and Neural Network. Finally, Nagaraj M. Lutimath and his team performed heart disease prediction using Naive Bayes classifier and SVM and found that SVM outperformed Naive Bayes in terms of accuracy. To determine the best algorithm for predicting heart disease, different classification algorithms were analyzed based on their accuracy, precision, recall, and F-measure scores. We analyzed decision tree, random forest, logistic regression, k-nearest neighbor, support vector, and Naive Bayes classification algorithms based on accuracy, precision, recall, and F-Measure scores and determined the best classification algorithm to use. prediction of heart disease.

III. METHODOLOGY

The aim of this project to give a model which will effectively give accurate results regarding heart disease. The Process Flow diagram is shown below. It shows the working of the machine learning model and also all the steps involved in the model. First the dataset is stored in the model then it is cleaned and significant variables are extracted from it. Data Pre-processing is done following splitting of data for training and testing. Then the algorithm is applied and the result is shown on the web interface.

IV. EXPERIMENTAL RESULTS

Data collection: -

To collect and pre-process the data, we used the Heart Disease Dataset, which comprises of four different databases, but we only utilized the UCI Cleveland dataset. This dataset contains a total of 76 attributes; however, for this study, we only used a subset of 14 features as referred to in published experiments. Therefore, we opted to use the pre-processed UCI Cleveland dataset, which is available on Kaggle. The details of the 14 attributes used in this study are outlined in Table shown below.

SR No	Attribute description	Distinct values of attributes
1	Age- represent the age of a person	Multiple values between 18 and above
2	Sex- describe the gender of person (0- Female, 1-Male)	0,1
3	Race – It shows the origin of the patient	Multiple values i.e. American Indian/Alaskan native, Asian, Black, Hispanic, White, Other
4	Health- It shows the general health of the patient	Excellent ,Fair, Good, Poor, Very Good
5	BMI- Body Mass Index	15 to 35
6	Smoking- Tells whether the patient smokes or not	0,1
7	Alcohol drinking- Tells whether the patient consumes alcohol or not	0,1
8	Stroke- Tells whether the patient has experienced stroke or not	0,1
9	Physical Health- physical health of the patient	0 to 10



10	Mental Health- mental health of the patient	0 to 10
11	Walking- It shows whether the patient has difficulty in walking or not	0,1
12	Diabetic- It shows whether the patient is diabetic or not	0,1
13	Physical activity- physical activity of the patient	0,1
14	Sleep Time- patient's duration of sleep	3 to 12 hours
15	Asthma- Tells whether the patient is diagnosed with Asthma or not	0,1

Data pre-processing: -

The objective was to convert raw data into a more useful and efficient format, which involved addressing missing or insignificant data through data cleansing. This process included checking for missing values and imputing them as necessary, as well as handling any noisy data that may have been generated by errors in data collection or entry. Once the data was cleaned, it was transformed to ensure it was in a suitable form for the mining process, with 82% of this involving normalization, discretization, and converting lower-level attributes to higher-level attributes in the hierarchy. Finally, the data was processed and split into training and test sets before being fed into the machine learning model.

Algorithm used: -

The aim of this research is to predict heart disease using instant measurement parameters through the implementation of KNN (K-nearest neighbors), one of the most popular data mining algorithms utilized in disease prediction studies. KNN is a simple and basic classification technique that is often applied when limited or no information is available about the data distribution. It is a non-parametric algorithm, meaning it does not assume any specific data distribution for analysis. This makes it a practical choice for real-world scenarios where data may not follow theoretical statistical models, such as the normal distribution. KNN is considered a "lazy" algorithm as it requires only a quick training phase and does not make any generalizations, instead it maintains all training data. To calculate the similarity between training and test data, KNN commonly uses Euclidean distance in its classifier.

Training the machine learning model: -

To train the model, the dataset was divided into two subsets, with 80% allocated to training data and 20% to test data. The training data was used to identify the patterns or relationships between features and their corresponding target column, which were then utilized to train the machine learning model. The model was then evaluated and produced an accuracy rate of 85% for the training data and 82% for the test data. Utilizing the trained model, a predictive system was developed, and various tests were conducted to evaluate the system's performance by inputting data that was not included in the original dataset.

Saving the machine learning model: -

We've saved the model with the help of the `fix` module in python. The `fix` module implements an abecedarian, still important formula for reissuing and serializing a Python object structure. Pickle model provides the posterior functions of `fix dump` and `fix cargo`.

V. FUTURE SCOPE

Improving the accuracy of health diagnosis can often involve effectively handling diverse category labels within the prediction method, which represents a promising direction for future research. With ongoing advancements in data processing and machine learning algorithms, we can continue to enhance this approach to achieve a level of accuracy that provides reliable predictions. This will enable us to make informed decisions based on the prediction results for various patients. To reduce the information size and obtain the optimal set of attributes for heart disease prediction, we plan to utilize a Genetic algorithm. The advantage of using this algorithm is that it can facilitate more accurate and consistent results in a shorter amount of time, due to the reduced dataset used for prediction.

VI. CONCLUSION

Heart diseases can be extremely dangerous and can have fatal consequences if left untreated. They are complex and difficult to predict, leading to a significant number of deaths each year. Neglecting the early signs of heart disease can result in a rapid deterioration of the patient's condition. In today's world, a sedentary lifestyle and increased stress have made the situation even worse. It is important to detect the disease as early as possible to keep it under control. Leading a healthy lifestyle and incorporating healthy habits is essential for overall well-being. Unhealthy habits like tobacco use and an unhealthy diet increase the risk of stroke and heart diseases. Consuming fruits and vegetables daily is a healthy practice that can help prevent these conditions. While classification techniques are typically used to predict heart disease, more research should focus on exploring various data cleaning and sanitization techniques that prepare datasets and make them more suitable for mining. Properly sanitized datasets have been found to provide higher accuracy than those with missing values. Choosing the right data cleaning technique, along with the appropriate classification algorithm, can lead to the development of predictive systems that significantly improve accuracy. In the future, intelligent systems may be developed to select the appropriate treatment for patients diagnosed with heart disease. Much work has already been done to create models that predict whether a patient is likely to develop heart disease. Data mining can be of great help in deciding which processing method to follow by extracting knowledge from relevant databases.

REFERENCES

1. "Heart Disease Prediction using effective Machine Learning Techniques" by Avinash Golande in International Journal of Recent Technology in the year 2019.
2. "Heart Disease Prediction using Data Mining with Mapreduce Algorithm" by T.Nagamani, S.Logeswari, B.Gomathy in International Journal of Innovative Technology and Exploring Engineering(IJITEE) in the year Jan 2019
3. "Implementation of Machine Learning Model to Predict Heart Failure Disease" by Fahd Saleh Alotaibi in International Journal of Advanced Computer Science and Applications in th year 2019.
4. "Design And Implementation Heart Disease Prediction Using Naives Bayesian" by Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin in International Conference on Trends in Electronics and Information in the year 2020.
5. "Prediction Of Heart Disease using Machine Learning" by Nagaraj utimath in International journal Of Recent Technology and Engineering in the year 2019.
6. "Prediction of Heart Disease Using Machine Learning Algorithms" by Santhana Krishnan. J, Geetha S. in International Conference on Innovations in Information and Communication Technology (ICIICT) in the year 2020.
7. "Efficient heart disease prediction system" by Purushottam, Kanak Saxena and Richa Sharma in Procedia Computer Science in the year 2016.
8. "Effective heart disease prediction using hybrid machine learning techniques" by Chandrasegar Thirumalai, and Gautam Srivastava in IEEE in the year 2019.
9. "An optimized stacked support vector machines based expert system for the effective prediction of heart failure" by Ali, Liaqat in IEEE in the year 2019.
10. "Biomarkers in cardiovascular disease: Statistical assessment and section on key novel heart failure biomarkers." By Dhingra, Ravi, and Ramachandran S in Trends in cardiovascular medicine in the year 2017.



SJIF Scientific Journal Impact Factor

Impact Factor: 8.423



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details