**ABSTRACT**

Day by day the cases of heart diseases are increasing at a rapid rate and it’s very Important and concerning to predict any such diseases beforehand. This diagnosis is a difficult task i.e. it should be performed precisely and efficiently. These research paper mainly focuses on which patient is more likely to have a heart disease based on various medical attributes. We prepared a heart disease prediction system to predict whether the patient is likely to be diagnosed with a heart disease or not. We used different algorithms of machine learning such as Random Forest, KNN,SVM and Decision Tree to predict and classify the patient with heart disease. A quite Helpful approach was used to regulate how the model can be used to improve the accuracy of prediction of Heart Attack in any individual. The Given heart disease prediction system enhances medical care and reduces the cost. This project gives us significant knowledge that can help us predict the patients with heart disease

INTRODUCTION

Heart disease describes a range of conditions that affect your heart. Today, cardiovascular diseases are the leading cause of death worldwide with 17.9 million deaths annually, as per the World Health Organization reports. Various unhealthy activities are the reason for the increase in the risk of heart disease like high cholesterol, obesity, increase in triglycerides levels, hypertension, etc. are certain signs which the American Heart Association lists like the persons having sleep issues, a certain increase and decrease in heart rate (irregular heart beat), swollen legs, and in some cases weight gain occurring quite fast; it can be 1-2 kg daily. All these symptoms resemble different diseases also like it occurs in the aging persons, so it becomes a difficult task to get a correct diagnosis, which results in fatality in near future. But as time is passing, a lot of research data and patients records of hospitals are available. There are many open sources for accessing the patient’s records and researches can be conducted so that various computer technologies could be used for doing the correct diagnosis of the patients and detect this disease to stop it from becoming fatal. Many studies have been performed and various machine learning models are used for doing the classification and prediction for the diagnosis of heart disease. Heart disease is very fatal and it should not be taken lightly. Heart disease happens more in males than females, which can be read further from Harvard Health Publishing. Researchers found that, throughout life, men were about twice as likely as women to have a heart attack. \*at higher risk persisted even after they accounted for tradi tional risk factors of heart disease, including high choles terol, high blood pressure, diabetes, body mass index, and physical activity.

METHODOLOGY

In this section, the method and analysis are described, which is performed in this research work. First of all, the collection of data and selection of relevant attributes are the initial steps in this study. After that, the relevant data is pre-processed into the required format. The given data is then separated into two categories: training and testing datasets. The algorithms are then used, and the given data train the model. The accuracy of this model is obtained by using the testing data. The procedures of this study are loaded by using several modules such as a collection of data, selection of attributes, pre-processing of data, data balancing, and prediction of disease.

DATA COLLECTION

In this article, the dataset is collected from the [Kaggle](https://www.kaggle.com/datasets/arezaei81/heartcsv) . So, the first step is organizing the dataset from the to predict the heart disease and then dividing the dataset into two sections: training and testing. In this article, 80% data has been considered as a training dataset, and 20% dataset is used for testing purposes.

DATASET AND ATTRIBUTES

Attributes of a dataset are properties of a dataset, which are important to analyze and make a prediction regarding our concern. Various attributes of the patient, like gender, chest pain, serum cholesterol, fasting blood pressure, exang, etc., are considered for predicting diseases. However, the correlation matrix can be used for attribute selection to construct a model.

|  |  |  |
| --- | --- | --- |
| ATTRIBUTES | DESCRIPTION | VALUES |
| 1.Age | Patients age in years | Continuous |
| 2.Sex | Male-0 Female-1 | Male Female |
| 3.CP | Chest Pain type | Four Types |
| 4.Trestbps | Resting Blood Pressure | Continuous |
| 5.Chol | Serum Cholesterol in mg/dl | Continuous |
| 6.FBS | Fasting Blood Pressure | < or > 120mg/dl |
| 7.Restecg | Resting Electrocardiograph | Five Values |
| 8.Thalach | Maximum heart rate achieved | Continuous |
| 9.Exang | Exercise Induced Angina | Yes/No |
| 10.oldpeak | ST Depression | Continuous |
| 11.slope | Slope of Peak Exercise ST segment | Up/flat/down |
| 12.Ca | Number of major vessels | 0-3 |
| 13.thal | Defect type | Reversible/Fixed/Normal |
| 14.Targets | Heart Disease | 1=Disease 0=No Disease |