Literature Survey

Visualizing and Predicting Heart Diseases with an Interactive Dash Board

Heart disease has become a global trend illness, with high mortality and death toll in modern society. Presently, diagnosis and treatment processes are highly challenging due to several risk factors like high blood pressure, high blood cholesterol level, diabetes, obesity, unhealthy diet and stress etc. Early diagnosis of heart disease is significant to minimize the heart related issues and to protect it from serious risks. The invasive techniques are implemented to diagnose heart diseases based on medical history, symptom analysis report by experts, and physical laboratory report. Moreover, it causes delay and imprecise diagnosis due to human intervention. It is time consuming, computationally intensive and expensive at the time of assessment. Data Analytics is considered as a cost effective technology in the recent past and it plays an essential role in healthcare. The analysis is carried out using a publicly available data for heart disease. Based on the accumulated case data, data analysis can be used to predict potential heart disease patients more accurately and optimize the treatment. In this study, some exploratory analysis on the heart disease data set and the relationship between each variable will be analyzed. Then symptom factors related to heart diseases will be obtained. This can help medical professionals predict heart disease status based on the patient's clinical data and improve patient treatment efficiency. Also, the study gives hints about the causes of the heart diseases, which will be useful in preventing this fatal illness.

Paper 1:   Heart Disease Prediction using Exploratory Data Analysis

- **Publication Year :** July 2020
- **Author :** Indrakumari Ranganathan, T.Poongodi, Soumya Ranjan Jena
- **Journal Name :** International Conference on Smart Sustainable Intelligent   Computing and Applications under ICITETM2020

Summary :

In this paper, the risk factors that causes heart disease is considered and predicted using K-means algorithm and the analysis is carried out using a publicly available data for heart disease. The dataset holds 209 records with 8 attributes such as age, chest pain type, blood pressure, blood glucose level, ECG in rest, heart rate and four types of chest pain. To predict the heart disease, K-means clustering algorithm is used along with data analytics and visualization tool. The paper discusses the pre-processing methods, classifier performances and evaluation metrics. In the result section, the visualized data shows that the prediction is accurate.

Reference Link :

https://www.sciencedirect.com/science/article/pii/S1877050920315210

Paper 2: Heart Disease Prediction model using Tree based method

- **Publication Year :** May 2022
- **Author :** Yanran Li, Yitong Liu, Jin Luo, Xiao Sun
- **Journal Name :** 2nd International Conference on Applied Mathematics, Modelling and Intelligent Computing (CAMMIC 2022)

Summary :

This study develops two heart disease prediction models using two kinds of tree methods and compares the error of the two models. Comparing with other research, this study uses the most basic algorithm in the tree-based methods to predict the data. It turns out that the random forest model fits better than the decision tree model which can predict the data much more accurate.

Reference Link :

https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12259/1225935/Heart-disease-prediction-model-using-tree-based-method/10.1117/12.2639449.short?SSO=1

Paper 3: Data Analytics for Cardiac Diseases

- **Publication Year :** March 2022
- **Author :** Martti Juhola, Henry Joutsijoki, Kirsi Penttinen, Disheet Shah
- **Journal Name :** ScienceDirect

Summary :

The research was based on calcium transient sigmals measured from induced pluripotent stem cell-derived cardiomyocytes. All in all, 55 different machine learning alternatives were used to model eight classes by applying the principle of 10-fold crossvalidation with the peak data of 1626 signals. The best classification accuracy of approximately 69% was given by random forests, which can be seen high enough here to show machine learning to be potential for the differentiation of the eight disease classes.

Reference Link :

https://www.sciencedirect.com/science/article/pii/S0010482522000105

Paper 4 : Prediction of heart disease at early stage using data mining and big data analytics

- **Publication Year:** December 2016
- **Author:** N. K. Salma Banu, Suma Swamy
- **Journal Name:** IEEE

Summary :

In this paper, the various technologies of data mining (DM) models for forecast of heart disease are discussed. Data mining plays an important role in building an intelligent model for medical systems to detect heart disease (HD) using data sets of the patients, which involves risk factor associated with heart disease. Medical practitioners can help the patients by predicting the heart disease before occurring. The large data available from medical diagnosis is analyzed by using data mining tools and useful information known as knowledge is extracted. Mining is a method of exploring massive sets of data to take out patterns which are hidden and previously unknown relationships and knowledge detection to

help the better understanding of medical data to prevent heart disease. Several studies have been carried out for developing prediction model using individual technique and also by combining two or more techniques. This paper provides a quick and easy review and understanding of available prediction models using data mining from 2004 to 2016. The comparison shows the accuracy level of each model given by different researchers.

Reference link :

https://ieeexplore.ieee.org/document/7955226/authors#authors

Paper 5: Predictive analytics to prevent and control chronic diseases

- **Publication Year:** July 2016
- **Author:** Kumari Deepika, S. Seema
- **Journal Name:** IEEE

Summary :

In these days, chronic diseases are the imperative reason for death in the world. Therefore, there is a noteworthy increment in consideration being paid to individual wellness as a preventative methodology in healthcare. However, creating and building a prediction model for chronic diseases is an extraordinary change to healthcare technology on the premise of data-analysis and decision-making level. In this paper, effective mechanisms have been used for chronic disease prediction by mining the data containing historical health records. Here, we used Naïve Bayes, Decision tree, Support Vector Machine (SVM) and Artificial Neural Networks (ANN) classifiers for the diagnosis of diabetes and heart disease.

Reference link :

https://ieeexplore.ieee.org/document/7912028

Paper 6:   Prediction of mortality in patients with cardiovascular disease using data mining methods

- **Publication Year:** March 2020
- **Author:** Damir Imamovic, Elmir Babovic, Nina Bijedic
- **Journal Name:** IEEE

Summary :

Healthcare information systems store a huge amount of patient data, so the trend of the use of data mining in healthcare is on the rise. Heart and blood vessel diseases are a leading cause of mortality both worldwide and here in Bosnia and Herzegovina, and prevention, surveillance and treatment are of great public health importance. Based on data on patients with cardiovascular disease, collected from 2011 to 2017 at Mostar Hospital, models for mortality prediction using techniques for data tree mining, neural network and logistic regression are presented. The aim of this research is to compare the effectiveness of these methods in modeling the effectiveness of predicting mortality in patients with cardiovascular disease.

Reference link :

https://ieeexplore.ieee.org/document/9066297/authors#authors

Paper 7:   Real-time machine learning for early detection of heart disease using big data approach

- **Publication Year:** April 2019
- **Author:** Khalil Maalmi
- **Journal Name:** IEEE

Summary :

This paper propose a real-time heart disease prediction system based on apache Spark which stand as a strong large scale distributed computing platform that can be used successfully for streaming data event against machine learning through in-memory computations. The system consists of two main sub parts, namely streaming processing and data storage and visualization. The first uses

Spark MLlib with Spark streaming and applies classification model on data events to predict heart disease. The seconds uses Apache Cassandra for storing the large volume of generated data.

Reference link :

https://ieeexplore.ieee.org/document/8723839/authors#authors

Paper 8: Predicting Heart Diseases through Feature Selection and Ensemble Classifiers

- **Publication Year:** June 2022
- **Author:** Mridu Sahu, N Kumar Swamy
- **Journal Name:** Research gate

Summary :

In this paper, they proposed a novel machine learning model for heart disease prediction. The proposed method was tested on two different datasets from Kaggle and UCI. We applied sampling techniques to the unbalanced dataset and feature selection techniques are used to find the best features. Later several classifier models were applied and achieved good accuracy with ensemble classifier. The experimentations on two datasets shown that the proposed model is effective for heart disease prediction. Python was used for all implementations.

Reference Link :

https://www.researchgate.net/publication/361128846_Predicting_Heart_Diseases_through_Feature_Selection_and_Ensemble_Classifiers

Paper 9: Predicting Heart Disease at Early Stages using Machine Learning: A Survey

- **Publication Year:** September 2022
- **Author:** Mr. Sumit Hawal, Sandeep Pande
- **Journal Name:** Research gate

Summary :

Heart disease is the largest cause of death worldwide, and early detection of heart disease is critical. Machine learning has evolved as one of the most progressive, dependable, and supportive tools in the medical field in recent years, providing the greatest assistance for disease prediction when properly trained and tested. The primary objective of this research is to evaluate several algorithms for heart disease prediction.

Reference Link :

https://www.researchgate.net/publication/354819509_Predicting_Heart_Disease _at_Early_Stages_using_Machine_Learning_A_Survey

Paper 10:   Predicting Heart Disease with Multiple Classifiers

- **Publication Year: January 2022**
- **Author:** Charly Gnoguem, Jules Degila, Carlyna Bondiombouy
- **Journal Name:** Research gate

Summary :

A novel hybrid technique is proposed to improve the prediction of heart disease in this document. This technique focuses on the reduction of false negatives for the betterment of patient care. The proposed technique assigns weights to four classifiers each built with one of the four reputable algorithms— decision tree, random forest, K-nearest neighbor, and logistic regression. The final class of a new instance is that predicted by the maximum weighted sum of predictions from the classifiers. This technique is compared with already existing methods, and an improvement in accuracy (92.10%) and sensitivity (94.59%) and a drastic reduction in false negatives are observed with the Cleveland dataset

Reference Link :

https://www.researchgate.net/publication/358219223_Predicting_Heart_Disease _with_Multiple_Classifiers

Paper 11:   Predicting the chances of occurrence of Cardio Vascular Disease (CVD) in people using classification techniques within fifty years of age

- **Publication Year :** January 2018
- **Author :** D. Karthick, B. Priyadharshini
- **Journal Name :** IEEE

Summary :

This paper aims in predicting the likelihood of cardiac disease in people based on the attributes such as the age, sex, cp, cholesterol, thallium etc. The data of existing patients within fifty years of age has to be collected and trained. Classification algorithms like Naive Bayes algorithms is applied to develop a model. This model will form a basis for predicting the chances of occurrence of CVD. If a person's data are given the classifier will predict if he is liable to CVD or not. The output is a binary classification which will give 1 or 0; 1 being the chances to occur and 0 being the chances not to occur.

Reference link :

https://ieeexplore.ieee.org/document/8398990/metrics#metrics

Paper 12:   Prediction of Cardiopathy Using Exploratory Data Analysis

- **Publication Year:** December 2021
- **Author:** Anum Manzoor Malik, Anil Kumar Sagar, Subrata Sahana
- **Journal Name:** IEEE

Summary :

EDA is a data analytics method which excludes mathematical modelling and inferences. Data analytics is a lowcost technology and has a vital role in health care industries, various sources include emergency situations, biomedical research, epidemics, pandemics etc. In this research they took Cleveland

cardiopathy dataset and then utilized K-means method to identify risk variables that cause cardiopathy. Age, hypertension, sugarlevel, chest discomfort, Electrocardiogram at relaxation, heart palpitation, and three forms of angina are among the 209 records in the collection and it was found that Kmeans method is the most effective one because of its speed and the proficiency of its output, it gives output in about 8sec so, for the prediction of cardiopathy K-means clustering method is employed by analyzing data in association with a visualization dashboard, and the data that is visualized in tableau demonstrates that the forecast is correct.

<u>Reference link :</u>

<u>https://ieeexplore.ieee.org/document/9666241/authors#authors</u>