IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 11, November 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

© 2024 IJIRCCE | Volume 12, Issue 11, November 2024| DOI: 10.15680/IJIRCCE.2024.1211059

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Heart Disease Prediction Using Machine Learning

Dr.D Sirisha [1], M.Ganesh[2], O.Vamsi[3], M.Jayavardhan[4], P.Narasimha[5], P.Yasodha[6]

SrAssistant Professor, Department of CSE (AI&ML), NSRIT, Visakhapatnam, A.P, India[1]

Department of CSE (AI&ML), NSRIT, Visakhapatnam, A.P, India[2,3,4,5,6]

ABSTRACT:

Background: Heart disease prediction is research in which the prognosis of heart disease through careful treatment takes place. This mainly helps in preventing cardiovascular diseases. Here, by using some of the known Machine Learning algorithms, and classifiers and by using the historical medical data, we can predict the unprecedented and rapid increase heart disease cases. This can help in decrease of mortality rate in our country as well as worldwide. By using electronic health record (EHR) data, this review can summarize the recent findings. From the previous historical data, we have observed the symptoms of heart disease such as decrease in heart rate (or) pulse rate, swollen legs, and in some cases weight gain occurring quite fast, etc.…The aim of this work is the proposal of a dimensionality reduction method and finding the features of heart disease by applying a feature selection.

KEYWORDS:

- Cardiovascular disease (CVD)
- Random Forests Classifier
- Logistic Regression (LR)
- Machine Learning (ML)
- Naive bayes Classifier

I. INTRODUCTION

Cardiovascular Disease (CVD) is one of the leading causing mortalities in the world, has been an important public health concern globally, causing massive socioeconomic burdens on patients, families, countries... every year. In terms of time, accuracy and cost, medical dictation has always remained as high maintenance field. Human beings are susceptible to mistakes and can make errors. At an exponential pace, cases of cardiovascular diseases are that, and that is very troubling.

One in five people develop heart failure (HF), and 50% of heart disease patients die in 5 years. The Heart Failure diagnosis, readmission, and mortality prediction are essential to develop personalized prevention and treatment plans. This review summarizes the recent findings and approaches of machine learning models for Heart Failure diagnostic and outcome prediction using EHR data.

Heart failure accounts for a large portion of this CVD (cardiovascular disease) morbidity and mortality, as well as an equally large portion of related healthcare expense. In the management of this expanding HF patient population, the accurate prediction of HF outcomes is critical to effective prevention and treatment. A set of machine learning models have been developed for HF diagnostic and outcome prediction using diverse variables derived from EHR data, including demographic, medical note, laboratory, and image data, and achieved expert-comparable prediction results. Hence Machine Learning models are being used to predict the heart failure of the patient in the early stage itself in this research.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. RELATED WORK

To predict heart disease by certain attributes using machine learning.

[1]: The algorithms used in this paper are Support Vector Machine (SVM), K- means clustering, K- Nearest Neighbor (KNN), Artificial Neural Network (ANN), Naïve Bayes, Logistic Regression. The dataset used in this paper is Cleveland heart disease raw dataset with 8 attributes. The highest accuracy attained for the algorithm K- means clustering.

[2]: The algorithms used in this paper are Random Forest, Support Vector Machine (SVM), K- Nearest Neighbor (KNN), Artificial Neural Network (ANN), Naïve Bayes. The dataset used in this paper is heart calculation time from UCI machine learning repository with 14 attributes. KNN and Random Forest are the best algorithms for the prediction and classification of heart disease dataset. The highest accuracy attained for the algorithm Random Forest with 99.6% accuracy.

[3]: The algorithms used in this paper are Decision Trees, Support Vector Machine (SVM), Neural Network, Naïve Bayes, Logistic Regression. The dataset used in this paper is UCI dataset with 14 attributes. The highest accuracy attained for the algorithm Support Vector Machine (SVM).

[4]: The algorithms used in this paper are Decision Trees, Support Vector Machine (SVM), K- Nearest Neighbor (KNN), Naïve Bayes, Logistic Regression, Random Forest, AdaBoost, Extra Tree Classifier, Gradient Boosting. The collected dataset is heart disease prediction dataset from Kaggle with 15 features. The highest accuracy attained for the algorithm Support Vector Machine (SVM).

[5]: The algorithms used in this paper are Support Vector Machine (SVM), K- Nearest Neighbor (KNN), Naïve Bayes, Neural Network, Random Forest, AdaBoost, Extra Tree Classifier, Gradient Boosting. The dataset used is heart disease dataset from UCI and among 13 attributes 8 were used due to simple measurements. The highest accuracy attained for the algorithm Random Forest.

[6]: The algorithms used in this paper are K- Nearest Neighbor (KNN) and Random Forest. The dataset used is heart disease dataset from Kaggle and 13 attributes were used. The highest accuracy attained for the algorithm KNN with 86.885% accuracy.

[7]: The algorithms used in this paper are Gaussian NB, Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, K- Nearest Neighbor (KNN), XGB, Extra Trees, Support Vector Machine (SVM), multilayer perceptron. The dataset used is Cleveland heart disease from UCI Repository and 13 medical features were used. The highest accuracy attained for the algorithm Multilayer Perceptron with 84.6% accuracy.

[8]: The algorithms used in this paper are Naïve Bayes and Support Vector Machine (SVM). The dataset used is Cleveland heart disease from UCI Repository and 26 attributes were used. The highest accuracy attained for the algorithm Support Vector Machine (SVM) with 89.11% accuracy.

[9]: The algorithms used in this paper are Naïve Bayes, Support Vector Machine (SVM), K- Nearest Neighbor (KNN), Random Forest, multilayer perceptron, Artificial Neural Network (ANN), Decision Trees. The dataset used is Cleveland heart disease from UCI Repository and 13 features were used. The highest accuracy attained for the algorithm Random Forest.

[10]: The algorithms used in this paper are K- Nearest Neighbor (KNN), Random Forest, AdaBoost, Decision Trees. The dataset used is Cleveland heart disease from UCI Repository and 14 attributes were used. The highest accuracy attained for the algorithm K- Nearest Neighbor (KNN) 100% accuracy.

[11]: The algorithms used in this paper are M5P, Random Tree and Reduced Error Pruning with Random Forest Ensemble Method. The dataset used is Cleveland heart disease from UCI Repository and 14 attributes from UCI repository were used. The highest accuracy attained for the algorithm Random Forest with 99.9% accuracy LRM.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Algorithm	Accuracy
Support vector machine	84.24%
Decision Tree	81.97%
Logistic Regression	85.25%
Random Forest	90.16%
Naïve Bayes	85.25%

Attributes used in datasets:

Number	Attribute	Description	Data type	Domain
1	Age	Patient age in year	Numerical	29 to 77
2	Sex	Gender	Binary	0 = female 1 = male
3	Chp	Chest pain type	Nominal	1 = typical angina, 2 = atypical angina 3 = nonanginal pain, 4 = asymptomatic
4	Bp	Resting blood pressure	Numerical	94 to 200
5	Sch	Serum cholesterol	Numerical	126 to 564
6	Fbs	Fasting blood sugar >120 mg/dL	Binary	0 = false 1 = True
7	Ecg	Resting electrocardiographic result	Nominal	0 = normal 1 = having ST-T wave abnormality 2 = left ventricular hypertrophy
8	Mhrt	Maximum heart rate	Numerical	71 to 200
9	Exian	Exercise induced angina	Binary	0 = no 1 = yes
10	Opk	Old peak	Numerical	Continuous (0 to 6.2)
11	Slope	Slope of peak exercise ST segment	Nominal	1 = upsloping 2 = flat 3 = downsloping
12	Vessel	Number of major vessels	Nominal	0 to 3
13	Thal	Defect type	Nominal	3 = normal, 6 = fixed defect, 7 = reversible defect
14	Class	Heart disease	Binary	0 = absence, 1 = presence

III. COMPARATIVE STUDY

In this paper, we applied machine learning algorithms on heart disease datasets to predict heart disease, with the help of the attribute for each patient. Our main aim was to compare different classification models and to find most efficient classification model. In this research we use different type of classification algorithms like K-nearest neighbour, support vector machine, Random Forest, Naïve Bayes, Logistic Regression.

The motivation for the study was to find the most efficient ML algorithm for detection of heart diseases. This study compares the accuracy score of Decision Tree, Logistic Regression, Random Forest, and Naive Bayes algorithms for predicting heart disease, by training the algorithm using UCI machine learning dataset and statlog dataset.

IV. CONCLUSION

The main aim of this project is to find good accuracy algorithm for the heart disease prediction, the result of this study indicates that the Random Forest algorithm is the most efficient algorithm with accuracy score of 90.16% for prediction of heart disease. Next, Logistic Regression and Naïve Bayes algorithm gives the second highest accuracy score of 85.25%. In the work, we have compared different algorithms individually but did not cover (100%) accuracy. In this research we have tried to test with different features selection method applied on different machine learning tree classifiers algorithms and finally find Random Forest with high accuracy score.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

1. Heart Disease Prediction using Exploratory Data Analysis, R. Indrakumari, T. Poongodi, Soumya Ranjan Jena -2020
2. Heart Disease Prediction and Classification Using Machine Learning Algorithms Optimized by Particle Swarm Optimization and Ant Colony Optimization, Youness Khourdifi(1) , Mohamed Bahaj(2) -- February 2019.
3. Heart Disease Prediction using Machine Learning, Abhijeet Jagtap(1) , Priya Malewadkar(2) , Omkar Baswat(3) , Harshali Rambade(4) -- February-2019.
4. Machine Learning Techniques For Heart Disease Prediction, A Lakshmana Rao (1), Y Swathi(2), Sundareswar Pullela(3). 03 October 2020.
5. classification and feature selection approaches by machine learning techniques: heart disease prediction, N Satish Chandra Reddy. 30 May 2019.
6. Heart disease prediction using machine learning techniques, Apurv Garg(1), Bhartendu Sharma(2), and Rijwan Khan(3). 24 october 2020.
7. MLP-PSO Hybrid Algorithm for Heart Disease Prediction, J. Pers. Med. 25 July 2022.
8.Prediction of Heart Disease Using Machine Learning Algorithms, Nagaraj M. Lutimath(1), Chethan C(2), Basavaraj S Pol(3). 25 September 2019.
9. Multi-Label Active Learning-Based Machine Learning Model for Heart Disease Prediction, Ibrahim M. El-Hasnony (1), Omar M. Elzeki (1), Ali Alshehri (3) and Hanaa Salem (4). 4 February 2022.
10. The Efficacy of Machine-Learning-Supported Smart System for Heart Disease Prediction. Absar, N.; Das, E.K.; Shoma, S.N.; Khandaker, M.U.; Miraz, M.H.; Faruque, M.R.I.; Tamam, N.; Sulieman, A.; Pathan. 18 June 2022
11. Prediction of Heart Disease Using Feature Selection and Random Forest Ensemble Method. Dhyan Chandra Yadav(1), Saurabh Pal(2). 6 June 2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details