IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111107

Heart Disease Prediction Using Machine Learning

Prof.Waghmode P.S[1], Wayase Dhanashri[2] , Kharat Dipali[3]

Professor ,Computer Department ,SPCOET Someshwarnagar College , Baramati , India[1]

Student ,Computer Department Someshwarnagar College , Baramati , India[2]

Student ,Computer Department Someshwarnagar College , Baramati , India[3]

Abstract: In recent times, Heart Disease prediction is one of the most complicated tasks in medical field. In the modern era, approximately one person dies per minute due to heart disease. Data science plays acrucial role in processing huge amount of data in the field of healthcare. As heart disease prediction is a complex task, there is a need to automate the prediction process to avoid risks associated with it and alert the patient well in advance. This paper makes use of heart disease dataset available in UCI machine learning repository. The proposed work predicts the chances of Heart Disease and classifies patient's risk level by implementing different data mining techniques such as Naive Bayes , Decision Tree, Logistic Regression and Random Forest. Thus, this paper presents a comparative study by analysing the performance of different machine learning algorithms. The trial results verify that Random Forest algorithm has achieved the highest accuracy of 90.16% compared to other ML algorithms implemented.

Keywords: Machine Learning Decision Tree, Naive Bayes, Logistic Regression,Random Forest, Heart Disease Prediction.

I. INTRODUCTION

The work proposed in this paper focus mainly on various data mining practices that are employed in heart disease prediction. Human heart is the principal part of the human body. Basically, it regulates blood flow throughout our body. Any irregularity to heart can cause distress in other parts of body. Any sort of disturbance to normal functioning of the heart can be classified as a Heart disease. In today's contemporary world, heart disease is one of the primary reasons for occurrence of most deaths. Heart disease may occur due to unhealthy lifestyle, smoking, alcohol and high in take of fat which may cause hypertension . According to the World Health Organization more than 10 million die due to Heart diseases every single year around the world. A healthy lifestyle and earliest detection are only ways to prevent the heart related diseases. The main challenge in today's healthcare is provision of best quality services and effective accurate diagnosis .

Even if heart diseases are found as the prime source of death in the world in recent years, they are also the ones that can be controlled and managed effectively. The whole accuracy in management of a disease lies on the proper time of detection of that disease. The proposed work makes an attempt todetect these heart diseases at early stage to avoid disastrous consequences.Records of large set of medical data created by medical experts are available for analysing and extracting valuable knowledge from it.

Data mining techniques are the means of extracting valuable and hidden information from the large amount of data available. Mostly the medical database consists of discrete information. Hence, decision making using discrete data becomes complex and tough task. Machine Learning (ML) which is subfield of data mining handles large scale well-formatted dataset efficiently.

In the medical field, machine learning can be used for diagnosis,detection and prediction of various diseases. The main goal of this paper is to provide a tool for doctors to detect heart disease as early stage . This in turn will help to provide effective treatment to patients and avoid severe consequences. ML plays a very important role to detect the hidden discrete patterns and thereby analyse the given data. After analysis of data ML techniques help in heart diseaseprediction and early diagnosis. This paper presents performance analysis of various ML techniques such as Naive Bayes, Decision Tree, Logistic Regression and Random Forest for predicting heart disease at an early stage

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022
DOI: 10.17148/IJARCCE.2022.111107

II. LITERATURE SURVEY

A literature review is a text of a scholarly paper, which includes the current knowledge including substantive findings, as well as theoretical and methodological contributions to a particular topic.Literature reviews use secondary sources, and do not report new or original experimental work.

1.Paper name: "Predicting the Risk of Heart Failure With EHR Sequential Data Modeling"
Author: Bo Jin, Chao Che et al.

model designed by applying neural network. This paper used the electronic health record (EHR) data from real-world datasets related to congestive heart disease to perform the experiment and predict the heart disease before itself. We tend to used one-hot encryption and word vectors to model the diagnosing events and foretold coronary failure events victimization the essential principles of an extended memory network model. By analyzing the results, we tend to reveal the importance of respecting the sequential nature of clinical records.

2.paper name: "Heart Disease Prediction using Evolutionary Rule Learning"
Author.- Aakash Chauhan et al.

This study eliminates the manual task that additionally helps in extracting the information (data) directly from the electronic records. To generate strong association rules, we have applied frequent pattern growth association mining on patient's dataset. This will facilitate (help) in decreasing the amount of services and shown that overwhelming majority of the rules helps within the best prediction of coronary sickness .

3.paper name: "An Intelligent Learning System based on Random Search Algorithm and Optimized Random Forest Model for Improved Heart Disease Detection"
Author: Ashir Javeed, Shijie Zhou et al

This paper uses random search algorithm (RSA) for factor selection and random forest model for diagnosing the cardiovascular disease. This model is principally optimized for using grid search algorithmic program. Two forms of experiments are used for cardiovascular disease prediction. In the first form, only random forest model is developed and within the second experiment the proposed Random Search Algorithm based random forest model is developed.

This methodology is efficient and less complex than conventional random forest model. Comparing to conventional random forest it produces 3.3% higher accuracy. The proposed learning system can help the physicians to improve the quality of heart failure detection

4.paper name: "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques"

Author: Senthilkumar Mohan, Chandrasegar Thirumalai et al.
efficient technique using hybrid machine learning methodology. The hybrid approach is combination of random forest and linear method. The dataset and subsets of attributes were collected for prediction.

The subset of some attributes were chosen from the pre-processed knowledge(data) set of cardiovascular disease .After prep-processing , the hybrid techniques were applied and disgnosis the cardiovascular disease

5.paper name: "Fast Rule-Based Heart Disease Prediction using Associative Classification Mining".

Author : K.Prasanna Lakshmi, Dr. C.R.K.Reddy

In the proposed Stream Associative Classification Heart Disease Prediction (SACHDP), we used associative classification mining over landmark window of data streams. This paper contains two phases: one is generating rules from associative classification mining and next one is pruning the rules using chi-square testing and arranging the rules in an order to form a classifier. Using these phase to predict the heart disease easily

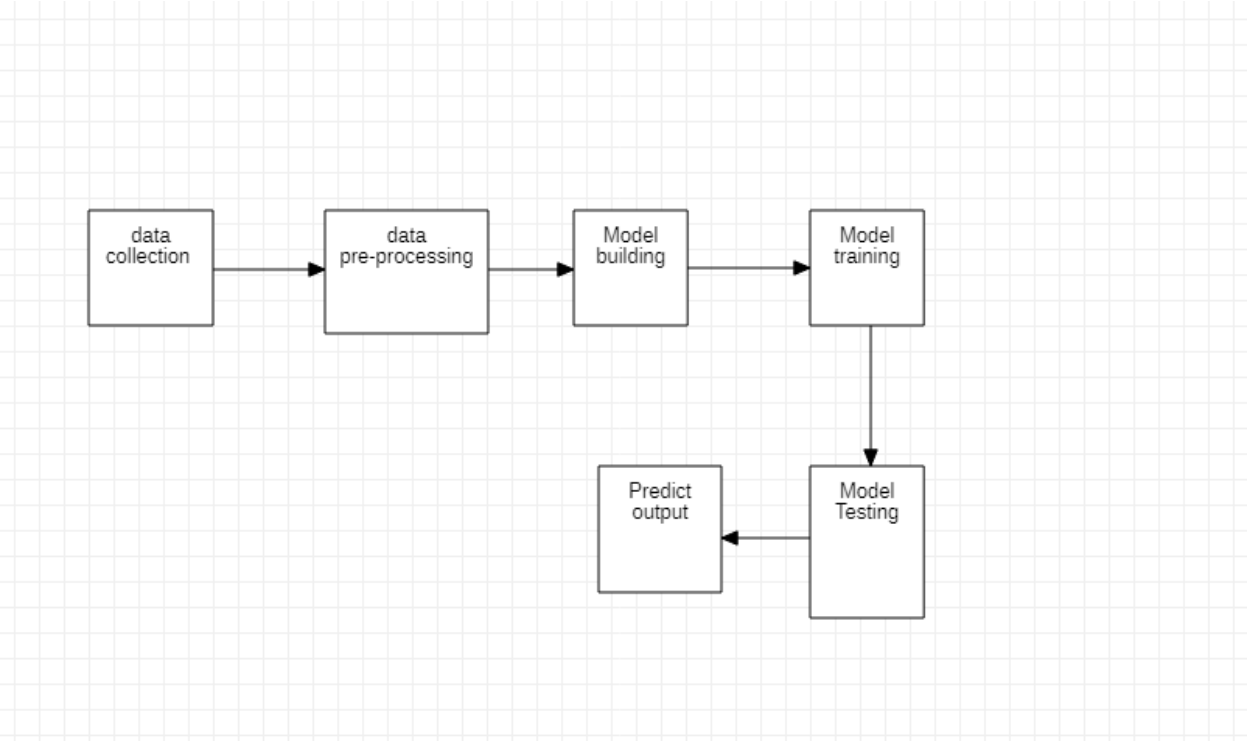
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022

DOI: 10.17148/IJARCCE.2022.111107

III. PROPOSED SYSTEM



The proposed work predicts heart disease by exploring the above mentioned four classification algorithms and does performance analysis. The objective of this study is to effectively predict if the patient suffers from heart disease.
The health professional enters the input values from the patient's health report. The data is fed into model which predicts the probability of having heart disease.

IV. ALGORITHM

1 DECISION TREES

For training samples of data D, the trees are constructed based on entropy inputs. These trees are simply constructed in a top down recursive divide and conquer (DAC) approach. Tree pruning is performed to remove the irrelevant samples on D.
Entropy = −Xm j=1 pij log2 pij

Algorithm for Decision Tree-Based Partition Require:

Input: D dataset – features with a target class for ∀features do
for Each sample do Execute the Decision Tree algorithm
end for Identify the feature space f1, f2, . . . , fx of dataset UCI.
end for Obtain the total number of leaf nodes l1,l2,l3,. . . ,
ln with its constraints Split the dataset D into d1, d2, d3,. . . , dn
based on the leaf nodes constraints. Output: Partition datasets d1, d2, d3

2 LANGUAGE MODEL

For given input features (xi, yi) with input vector xi of data D the linear form of solution
f (x) = mx+b equation is solved by subsequent parameters:
m = P

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 11, Issue 11, November 2022
DOI: 10.17148/IJARCCE.2022.111107

3 RANDOM FOREST

This ensemble classifier builds several decision trees and incorporates them to get the best result.
For tree learning, it mainly applies bootstrap aggregating or bagging. For a given data, X = {x1, x2, x3, ..., xn} with responses Y = {x1, x2, x3, ..., xn} which repeats the bagging from b = 1 to B

4 SUPPORT VECTOR MACHINE

Let the training samples having dataset Data = {yi, xi}; i = 1, 2, . . . , n where xi ∈ R n represent the i th vector and yi ∈ R n represent the target International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072 © 2020, IRJET | Impact Factor value: 7.529 | ISO 9001:2008 Certified Journal | Page 5275 item.

The linear SVM finds the optimal hyperplane of the form f (x) = w T x + b where w is a dimensional coefficient vector and b is a offset. This is done by solving the subsequent optimization problem:

Minw,b, ξi 1 2 w 2 + C Xn i=1 ξi s.t. yi , w T xi + b ≥ 1 − ξi, ξi ≥ 0, ∀i ∈ {1, 2, . . . ,}

V. CONCLUSION

In this paper, we proposed a method for heart disease prediction using machine learning techniques, these results showed a great accuracy standard for producing a better estimation result. By introducing new proposed Random forest classification, we find the problem of prediction rate without equipment and propose an approach to estimate the heart rate and condition. Sample results of heart rate are to be taken at different stages of the same subjects, we find the information from the above input via ML Techniques. Firstly, we introduced a support vector classifier based on datasets

REFERENCES

[1] A. S. ABDULLAH AND R. R. RAJALAXMI, ''A DATA MINING MODEL FOR PREDICTING THE CORONARY HEART DISEASE USING RANDOM FOREST CLASSIFIER,'' IN PROC. INT. CONF. RECENT TRENDS COMPUT. METHODS, COMMUN. CONTROLS, APR. 2012, PP. 22–25.
[2] A. H. Alkeshuosh, M. Z. Moghadam, I. Al Mansoori, and M. Abdar, ''Using PSO algorithm for producing best rules in diagnosis of heart disease,'' in Proc. Int. Conf. Comput. Appl. (ICCA), Sep. 2017, pp. 306– 311.
[3] N. Al-milli, ''Backpropogation neural network for prediction of heart disease,'' J. Theor. Appl.Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.
[4] C. A. Devi, S. P. Rajamhoana, K. Umamaheswari, R. Kiruba, K. Karunya, and R. Deepika, ''Analysis of neural networks based heart disease prediction system,'' in Proc. 11th Int. Conf. Hum. Syst. Interact. (HSI), Gdansk, Poland, Jul. 2018, pp. 233–239.
[5] P. K. Anooj, ''Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules,'' J. King Saud Univ.- Comput. Inf. Sci.,vol.24,no.1,pp.27– 40,Jan.2012.doi:10.1016/j.jksuci.2011.09.002.
[6] L. Baccour, ''Amended fused TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets,'' Expert Syst. Appl., vol. 99, pp. 115–125, Jun. 2018. doi: 10.1016/j.eswa.2018.01.025.
[7] C.-A. Cheng and H.-W. Chiu, ''An artificial neural network model for the evaluation of carotid artery stenting prognosis using a national-wide database,'' in Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Jul. 2017, pp. 2566–2569.
[8] H. A. Esfahani and M. Ghazanfari, ''Cardiovascular disease detection using a new ensemble classifier,'' in Proc. IEEE 4th Int. Conf. Knowl.Based Eng. Innov. (KBEI), Dec. 2017, pp. 1011–1014.
[9] F. Dammak, L. Baccour, and A. M. Alimi, ''The impact ofcriterion weights techniques in TOPSIS method of multi-criteria decision making in crisp and intuitionistic fuzzy domains,'' in Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE), vol. 9, Aug. 2015, pp. 1–8.
[10] R. Das, I. Turkoglu, and A. Sengur, ''Effective diagnosis of heart disease through neural networks ensembles,'' Expert Syst. Appl., vol. 36, no. 4, pp. 7675–7680, May 2009. doi: 10.1016/j.eswa.2008.09.013.