iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** IV **Month of publication:** April 2020

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue IV Apr 2020- Available at www.ijraset.com

Prediction of Heart Disease Symptoms using Machine Learning Algorithm

Sunil Rajput[1], Ashish Singh[2], Mayank Patil[3], Ashish Yadav[4], Prof. Bharti Patil[5]
[1, 2, 3, 4, 5]*Department Of Computer Engineering, SITRC, Nashik*

Abstract: ***Heart is the next major organ comparing to brain which has more priority in Human body. It pumps the blood and supplies to all organs of the whole body. Heart related diseases or Cardiovascular Diseases (CVDs) are the main reason for a huge number of death in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. Machine learning isan application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Now a days, Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. The primary aim is to allow the computers learn automatically. Many doctors suggest healthy diet and regular exercise to keep the heart healthy. Following are the parameters which are considered for the study in designing the system. Age , Sex , Blood Pressure , Heart Rate , Diabetes , Hyper cholesterol, Body Mass Index (obesity). We propose to develop an application which can predict the vulnerability of a heart disease given basic symptoms like age, sex, pulse rate etc. The machine learning algorithm neural networks has proven to be the most accurate and reliable algorithm and hence used inthe proposed system. This System predicts the arising possibilities of Heart Disease. The outcomes of this system provide the chances of occurring heart disease in terms of percentage. The datasets used are classified in terms of medical parameters. This project proposes a prediction model to predict whether a people have a heart disease or not and to provide an awareness or diagnosis on that.***
Index Term: Decision Tree classification, Prediction, Healthcare, Machine learning, Heart, Cardiovascular disease, Symptoms, Deep learning.

I. INTRODUCTION

The main topic is prediction using machine learning technics. Machine learning is widely used now a days in many business applications like e commerce and many more. Prediction is one of area where this machine learning used, our topic is about predic- tion of heart disease by processing patient's dataset and a data of patients to whom we need to predict the chance of occurrence of a heart disease. The performance of the diagnosis model is obtained by using methods like classification, accuracy, sensitivity and specificity analysis. It is impractical for a common man to frequently undergo costly tests like the ECG and thus there needs to be a system in place which is handy and at the same time reliable, in predicting the chances of a heart disease. Intelligent Heart Disease Prediction System Using Data Mining Techniques: The healthcare industry collects huge amounts of healthcare data which, unfortunately, are not "mined" to discover hidden information for effective decision making. Discovery of hidden patterns and relationships often goes unexploited. Advanced data mining techniques can help remedy this situation. This research has developed a prototype Intelligent Heart Disease Prediction System (IHDPS) using data mining techniques, namely, Decision Trees, Naive Bayes and Neural Network.

II. LITERATURE REVIEW

A. Mister Mohammed Abdul Khaleel has given paper in the Survey of Techniques for mining of data on Medical Data for Finding Fre- quent Diseases locally. This paper focus on dissect information mining procedures which are required for medicinal information mining particularly to find locally visit illnesses, for example, heart infirmities, lung malignancy, bosom disease et cetera. In- formation mining is the way toward extricating information for finding inactive examples which Vembandasamy et al. performed a work, to analyze and detect heart disease. In this the algorithm used was Naive Bayes algorithm. In Naïve Bayes algorithm they used Bayes theorem. Hence Naive Bayes has a very power to make assumption independently. The used data-set is obtained from a diabetic research institutes of Chennai, Tamilnadu which is leading institute. There are more than 500 patients in the dataset. The tool used is Weka and classification is executed by using 70% of Percentage Split. The accuracy offered by Naive Bayes is 86.419%.

B. L.Sathish Kumar and A. Padmapriya has given a paper named Prediction for similarities of disease by using ID3 algorithm in television and mobile phone. This paper gives a programmed and concealed way to deal with recognize designs that are covered up of coronary illness. The given framework utilize information min- ing methods, for example, ID3 algorithm.

C. M.A.Nishara Banu and B.Gomathy has given a paper named Disease Predicting system using data mining techniques. In this paper they talk about MAFIA (Maximal Frequent Item set algo- rithm) and K-Means clustering. As classification is important for prediction of a disease. The classification based on MAFIA and K-Means results in accuracy.

D. Analysis of Data Mining Techniques for Heart Disease Prediction: Heart disease is considered as one of the major causes of death throughout the world. It is hard to predict for the medical practitioners as it is a difficult task which demands expertise and higher knowledge for prediction. This paper addresses the issue of prediction of heart disease according to input attributes based on data mining techniques. We have investigated the heart disease prediction using KStar, J48, SMO, Bayes Net and Multilayer Perceptron through Weka software. The performance of these data mining techniques is measured by combining the results of predictive accuracy, ROC curve and AUC value using a 6 standard data set as well as a collected data set. Based on performance factor SMO and Bayes Net techniques show optimum performances than the performances of K-Star and J48 techniques.

E. Malay Shah and Rupal Kapadi, "Object Detection using Deep Neural Network", International Conference on Intelligent Computing and Control Systems ICICCS, 2017. Object detection has been a topic for challenge and many methodologies are applied. Object detection is detecting a specific object from an image of multiple and complex lines and shapes. Object detection is used in face detection, object tracking, image retrieval, automated parking systems

F. Wang Zhiqiang, Liu Jun, "A Review Of Object Detection based On Convolutional Neural Network", 36th Chinese Control Conference, 2017.The paper focused on the object detection based on CNN, the structure of CNN, the framework of object detection based on CNN and the methods of improving detection performance are introduced. CNN has strong ability in feature extraction, it can compensate for the drawback existing in hand-crafted features. CNN also has better advantage than conventional methods on real-time, accuracy, adaptability, but it still has lots of room for improvement

G. Meera M. K. , Shajee Mohan B. S., "Object Recognition in Images", International Conference on Information Science(ICIS), 2018. A SIFT feature based k-NN classifier and a GIST feature based SVM classifier are implemented. GIST feature based SVM classifier using Gaussian kernel showed better classifi- cation accuracy than SIFT feature based k-NN. Thus SVM classifier with Gaussian kernel is finalised for query image classification. In the second phase of this work, we will implement a k-NN classifier which will make use of SIFT feature based similarity matrix to retrieve the images containing the query image.

H. Christian Szegedy, Alexander Toshev and Dumitru Erhan," Deep Neural Networks for object detection", advances in Neural Information Processing Systems, 2017.

I. Zhong-Qiu Zhao, Peng Zheng and Shou-tao Xu, "Object Detection With Deep\Learning : A Review", 2014.Due to its powerful learning ability and advantages in dealing with occlusion, scale transformation and background switches, deep learning based object detection has been a research hotspot in recent years. This paper provides a detailed review on deep learning based object detection frameworks which handle different subproblems, such as occlusion, clutter and low resolution, with different degrees of modifications on R-CNN. The review starts on generic object detection pipelines which provide base architectures for other related tasks. Then, three other common tasks, namely salient object detection, face detection and pedestrian detection, are also briefly reviewed. Finally, we propose several promising future directions to gain a thorough understanding of the object detection landscape. This review is also meaningful for the developments in neural networks and related learning systems, which provides valuable insights and guidelines for future progress.

III. PROBLEM DEFINATION

The main objective of this project is to develop a platform which will be simple and easy to use, as here one must provide the patient's medical details and based on the features extracted the algorithm will then detect the heart disease and spot its type. As here algorithm does the task hence a welltrained model is less bound to make errors in predicting the heart disease and its type hence, in short accuracy is improved and thereby it also saves time and makes easier for doctors as well as patients to predict whether they are prone to any type of heart disease or not, which is otherwise we difficult to do without doctor's involvement.

IV. PROPOSED WORK

1) *Cleaning*: Data that we want to process will not be clean that is it may contain noise or it may contain values missing of we process we cant get good results so to obtain good and perfect results we need to eliminate all this, the process to eliminate all this is data cleaning. We will fill missing values and can remove noise by using some techniques like filling with most common value in missing place.
2) *Transformation*: This involves changing data format to one form to other that is making them most understandable by doing normalization, smoothing, and generalization, aggregation techniques on data.
3) *Integration*: Data that we need not process may not be from a single source sometimes it can be from different sources we do not integrate them it may be a problem while processing so integration is one of important phase in data pre-processing and different issues are considered here to integrate.
4) *Reduction*: When we work on data it may be complex and it may be difficult to understand sometimes so to make them understand- able to system we will reduce them to required format so that we can achieve good results.

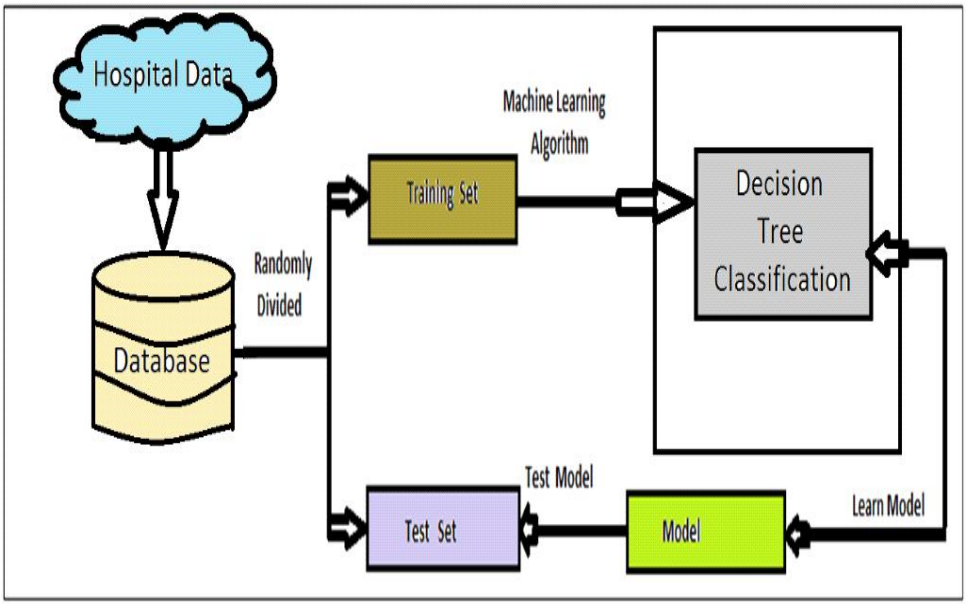
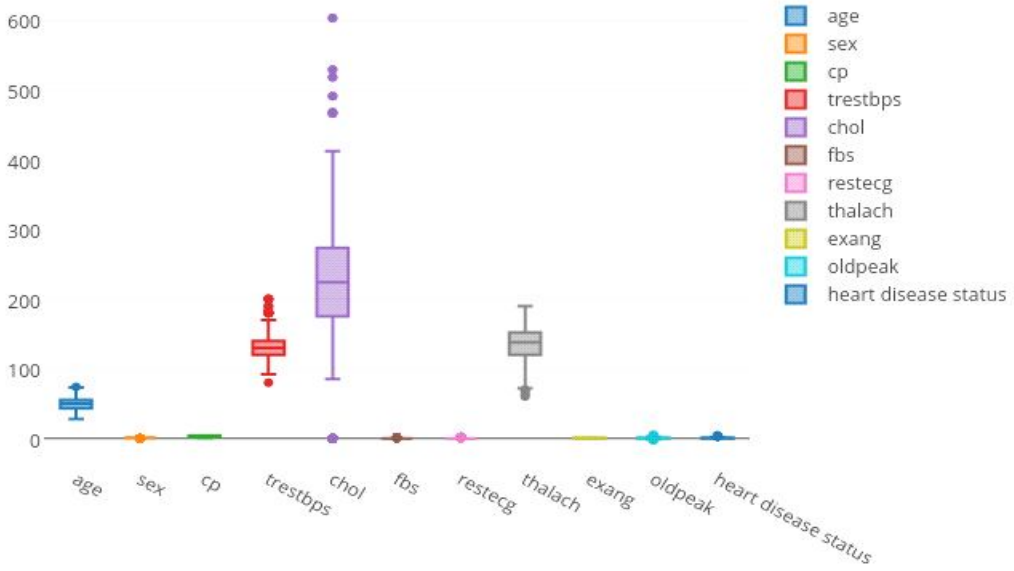


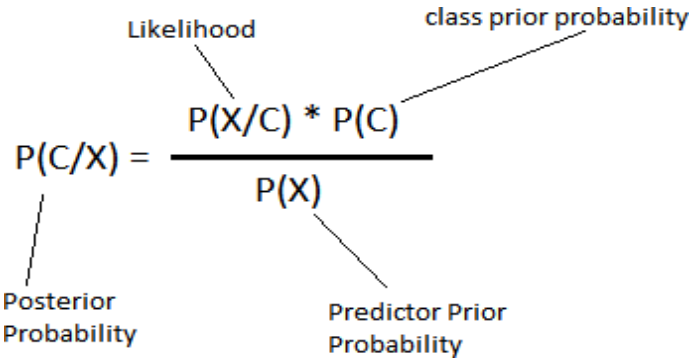
Fig: Architecture Diagram

5) *Website:* The system will consist of a website, where users will register themselves for getting the report of health of their heart in terms of predictive analysis about their heart disease. User will have to fill a form initially for registration. Then user will get redirected to the profile page the patient will be able to have look at the report where they will be knowing the status or risk of their heart in terms of percentage. If the user will have risk greater than 60% then user will be redirected ib Heart Disease Prediction using Machine Learning Algorithms.
6) *Database:* The server will be using MySQL database. The systems database consists of following tables.
7) *User Table:* This table will consist of all the user information which includes user's name, e-mail id, phone number, address, etc.
8) *Medical History Table :* This table will consist of all the health related information of users which is related to heart that includes attributes such as age, gender, resting blood pressure, cholesterol, fasting blood sugar, old peak, etc.
9) Machine Learning Algorithm: The machine learning algorithm will be used to predict the risk of heart disease in terms of percentage.
10) Heart diseases prediction is a web-based machine learning application, trained by a UCI dataset. The user inputs its specific medical details to get the prediction of heart disease for that user. The algorithm will calculate the probability of presence of heart disease. The result will be displayed on the webpage itself. Thus, minimizing the cost and time required to predict the disease. Format of data plays crucial part in this application. At the time of uploading the user data application will check its proper file format and if it not as per need then ERROR dialog box will be prompted. Our system will be implementing the following algorithms:

11) *Decision Tree Algorithm :* The decision tree approach is more powerful for classification problems. There are two steps in this techniques building a tree & applying the tree to the dataset. There are many popular decision tree algorithms CART, ID3, C4.5, CHAID, and J48. From these J48 algorithm is used for this system. J48 algorithm uses pruning method to build a tree. Pruning is a technique that reduces size of tree by removing over fitting data, which leads to poor accuracy in predications. The J48 algorithm recursively classifies data until it has been categorized as perfectly as possible. This technique gives maximum accuracy on training data. The overall concept is to build a tree that provides balance of flexibility & accuracy.



12) *Naïve Bayes Algorithm:* The Naïve-Bayesian classifier relies upon Bayes' speculation with autonomy suppositions among attributes [7-13]. A Naïve- Bayesian output is definitely not hard to run, with no entrapped repetitive parameter estimation which makes it particularly sup- portive for broad datasets in spite of its effortlessness, the Naive Bayesian classifier generally completes its job shockingly good and is broadly used in light of the fact that it frequently outflanks high order techniques which are complex. The Naïve Bayes treats every variable as independent which helps it to predict even if variables don't have proper relation [1].



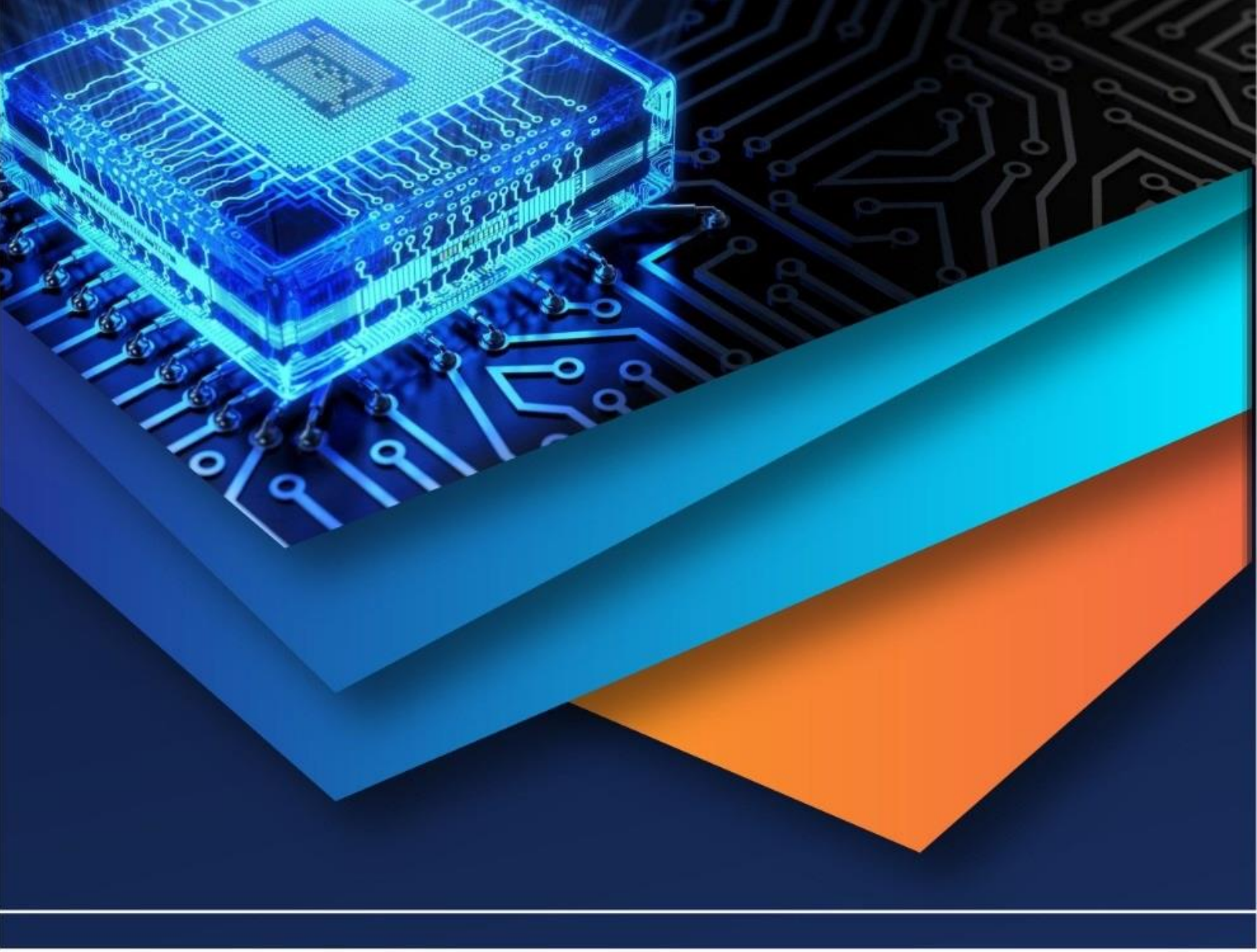
P(c|x) is the posterior probability of class (target) giv- en predictor (attribute)
P(c) is the prior probability of class.
P(x|c) is the likelihood which is the probability of predictor given class.
P(x) is the prior predictor.

V. CONCLUSION

In this way we found is during small datasets in some other cases most of time decision trees direct us to a solution which is not accurate, but when we look at Naïve Bayes results we are getting more accurate results with probabilities of all other possibilities but due to guidance to only one solution decision trees may miss lead. Finally we can say by this experiment that Naïve Bayes is more accurate if the input data is cleaned and well maintained even though ID3 can clean it self it cannot give accurate results every time, and in this same way Naïve Bayes also will not give accurate results every time we need to consider results of different algorithms and by all its results if a prediction is made it will be accurate. But we can use Naïve Bayes consider variables as individual we can use combination of algorithms like Naïve Bayes and K-means to get accuracy. Heart attack is crucial health problem in human society. This paper has summarised state of art techniques and available methods for predication of this disease. Deep learning an emerging area of artificial intelligence showed some promising result in other field of medical diagnose with high accuracy. An analytical comparison has been done for finding out best available algorithm for medical dataset. In future our aim is to carry forward the work of temporal medical dataset, where dataset varies with time and retraining of dataset is required.

REFERENCES

[1] Sellappan Palaniappan, Rafiah Awang “Intelligent Heart Disease Prediction System Using Data Mining Techniques”, IEEE, July 2015
[2] M. Raihan, Saikat Mondal, Arun More, Md. Omar Faruqe Sagor, Gopal Sikder, Mahbub Arab Majumder, Mohammad Abdullah Al Manjur and Kushal Ghosh “Smartphone Based Ischemic Heart Disease (Heart Attack) Risk Prediction using Clinical Data and Data Mining Approaches, a Prototype Design”, September 2014.
[3] Amudhavel, J., Kodeeshwari, C., Premkumar, K., Jaiganesh, S., Rajaguru, D., Vengattatraman, T., Haripriya, R., "Comprehensive analysis on information dissemination protocols in vehicular ad hoc networks", (2015) International Journal of Applied Engineering Re- search, 10 (3), pp. 2058-2061.
[4] Soodeh Nikan, Femida Gwadry-Sridhar, and Michael Bauer “Machine Learning Application to Predict the Risk of Coronary Artery Atherosclerosis”, IEEE, August 2016
[5] Sanjay Kumar Sen Asst. Professor, Computer Science & Engg. Orissa Engineering College, Bhubaneswar, Odisha – India.” Predicting and Diagnosing of Heart Disease Using Machine Learning Algorithms” International Journal of Engineering and Computer Science. Volume 6 Issue 6, June 2017
[6] V.V. Ramalingam, Ayantan Dandapath, M Karthik Raja “Heart disease prediction using machine learning tech : A survey” International Journal of Engineering & Technology, 7 (2.8), April 2018.
[7] K. Srinivas, B. Kavihta Rani, A. Govrdhan “Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attack” IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010.
[8] Sonam Nikhar, A.M. Karandikar “Prediction of Heart Disease Us- ing Machine Learning Algorithms” in International Journal of Ad- vanced Engineering, Management and Science (IJAEMS) June- 2016 vol-2.
[9] Deeanna Kelley “Heart Disease: Causes, Prevention, and Current Research” in JCCC Honors Journal.
[10] Marjia Sultana, Afrin Haider and Mohammad Shorif Uddin “Analysis of Data Mining Techniques for Heart Disease Prediction”, May 2015.
[11] Nabil Alshurafa, Costas Sideris, Mohammad Pourhomayoun, Haik Kalantarian, Majid Sarrafzadeh "Remote Health Monitoring Out- come Success Prediction using Baseline and First Month Interven- tion Data" in IEEE Journal of Biomedical and Health Informatics.
[12] Heart Disease Dataset - https://www.kaggle.com/c/heart-disease dated: Sept 2018
[13] https://archive.ics.uci.edu/ml/datasets/heart+Disease.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY