iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52839

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: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Comparative Analysis of Classification Algorithms for Heart Disease

Shradha Solapure[1], Ruchika Jadhav[2], Yash Surana[3], Prajit Thube[4]
[1, 2, 3, 4]*Student Department of Computer Engineering, Zeal College of Engineering and Research, Pune*

Abstract: ***Heart is the main component of the human body and without it the body can't function. It provides the flow of blood of different organs and body parts. It purifies the blood by removing the carbon dioxide(co2). It is also known as cardiovascular disease, it creates many risk factors for a human, including death. Heart disease is one of the most common causes of death around the world nowadays. Often, the enormous amount of information is gathered to detect diseases in medical science. All of the information is not useful but vital in taking the correct decision. Thus, it is not always easy to detect the heart disease because it required skilled knowledge or experiences about heart failures symptoms for an early prediction. Most of the medical dataset are dispersed, widespread and assorted. However data mining is a robust technique for extracting invisible, predictive and actionable information from the extensive databases. for identifying the possibility of heart disease in a patient. This work is justified by performing a comparative study and analysis using classification algorithms namely Logistic Regression, KNN, SVM, Decision Tree, and Random Forest are used at different levels of evaluation. Further, this research work is aimed towards identifying the best classification algorithm for identifying the possibility of heart disease in a patient. This work is justified by performing a comparative study and analysis using three classification algorithms namely Logistic Regression, KNN, SVM, Decision Tree and Random Forest are used at different levels of evaluation.***

Key Phrases: Heart Disease, design section, KNN, Decision tree, SVM, Logistic Regression, Random forest

I. INTRODUCTION

The amount of data in the medical industry is increasing day by day. It is a challenging task to handle a large amount of data and extracting productive information for effective decision making. For this reason, medical industry demands to apply a special technique which will provide fruitful decision from a vast database. Data mining is an exciting field of machine learning and thus capable of solving this type of problem very well. For solving various kinds of real-world problems, data mining is a novel field for discovering hidden patterns and the valuable knowledge from a large dataset. Because it is very strenuous to extract any useful information without mining large database. In brief, it is an essential procedure for analyzing data from various perspectives and gathering knowledge. However, heath care industry is another field where a substantial amount of data collected using different clinical reports and patients manifestations. Nowadays, people can face any heart failure symptoms at any stage of a lifetime. But old people face this type of problem rather than the young people. Data mining classification techniques can discover the hidden relationship along correlated features which plays a consequential role in predicting the class label from a large dataset. By using those hidden patterns along with the correlated features, it is straightforward to detect heart disease patients without any support of medical practitioners. Then, it will act as an expert system for separating patients with heart disease and patients with no heart disease more accurately with lower cost and less diagnosis time. According to the World Health Organization, heart disease claims the lives of 17.7 million people per year, accounting for 31% of all deaths worldwide. Heart disorders have been a leading cause of death in India as well . According to the Global Burden of Disease study released on the 15th of September 2017, heart disease claimed the lives of 1.7 million Indians in 2016. According to WHO estimates, India lost $237 billion between 2005 and 2015 due to heart-related or cardiovascular disease. Therefore an accurate and feasible prediction of heart diseases is necessary.

II. LITERATURE SURVEY

Sr.No.	Page Title	Author	Publisher and year	Approach Used	Strengths	Limitations
1	Prediction of Diabetes using Machine	Naveen Kishor, G.V.Rajesh,	International Journal of Science	SVM, Deci- sion Tree, LR, KNN,	Random Forest shows more	Misclassification is observed to be

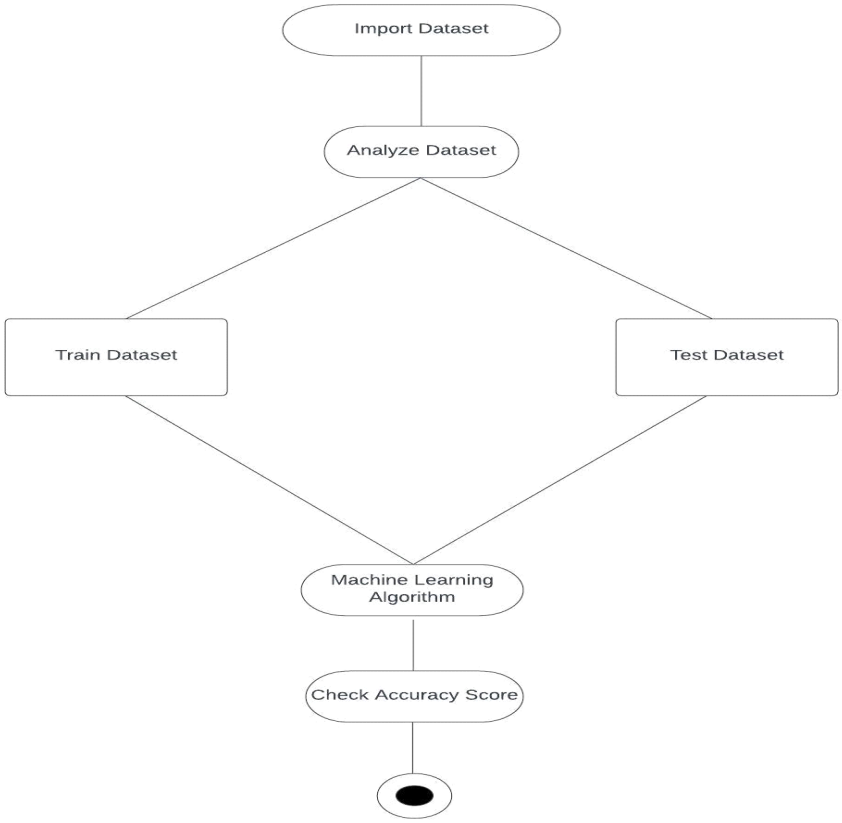
©IJRASET: All Rights are Reserved

	Learning Classification Algorithms	A.Vamsi Akki Reddy, K.Sumedh, T.Rajesh Sai Reddy	and Technology Research Volume 9,January2020	Random Forest	accuracy in prediction fol- lowed by SVM and Logistic Regression	more in KNN and less in Random Forest
2	Heart Disease Prediction using Machine Learning	Apurb Ra- jdhan, Avi Agarwal, Milan Sai, Dundi- Galla Ravi, Dr.Poonam Ghuli	International Journal of Engineering Research	Technology (IJERT)04, April-2020	Random For- est, Decision Tree, Logistic Regression and Naive Bayes	Random Forest algorithm is the most efficient algorithm with more accuracy.
	Prediction of heart disease and classifiers sensitivity analysis	Khaled Mohamad Almustafa	Bioinformatics (BMC) 2020	KNN, Naive Bayes, Decision Tree, SVM ,Ad boost.	Algorithms are giving very promising re- sults in terms of classification accuracy.	Not every time Naive Bayes gives an accurate result.
4	Predicting Heart Disease using Machine Learning Algorithm	Anagha Sridhar, Anagha S Kapardhi	International Research Journal of Engineering and Technology (IRJET) 04 Apr 2019	Decision Tree, Naive Bayes, decision Tree, Naive Bayes	Decision tree was more precise in calculation of heart disease	
5	Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques	C. Beulah Christalin Latha, S. Carolin Jeeva	Karunya Institute of Technology and Sciences, India 2019	Bayes Net, Naive Bayes, Random forest, C4.5, Multilayer perceptron Ensemble techniques - Bagging, Boosting, Stacking, Majority vote.	In order to improve the performance of the weak classifiers, ensemble algorithms are used.	using feature selection tools, the results could be improved even further.
6	Heart Disease Prediction using Machine Learning	S.Nandhini, Monojit Debnath, Anurag Sharma, Pushkar	International Journal of Recent Engineering Research and Development ,10 October 2018	SVM, Decision Trees, LR, Random Forest Naive Bayes, KNN	Random Forest shows more accuracy in prediction followed by SVM and KNN	Algorithms do not perform well according to the popularity of data
7	Prediction and Analysis the occurrence of Heart Disease using data mining techniques	Chala Bayen	2018	J48,Na¨ıve Bayes, SVM	It provides quick outcomes, allowing individuals to receive important analytic while saving money.	
8	Machine Learning Algorithms based Skin Disease Detect	Shuchi Bhadula, Sachin Sharma, Piyush Juyal, CHitransh Kulshrestha	IJITEE 2019	Random Forest, Naive Bayes, LR, CNN and SVM.	CNN gives more accurate results ,followed by logistic regression and random forest.	Error rate in Naive Bayes is comparatively more than other

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

III. SYSTEM ARCHITECTURE



IV. METHODOLOGY

A. Data Preparation

The dataset is gathered from a net understand dataset known as the heart assault analysis and prediction dataset. due to the fact the records is tough to come back by means of, the most effective manner to run the model and make a forecast turned into to apply information from a reliable source. The dataset contains numerous attributes along with Age, gender, fasting blood sugar, serum cholestrol , most coronary heart charge achieved, rating blood strain , exercising listed angina, oldpeak , wide variety of vessels etc

B. Data Preprocessing

Actual world things includes errors in it so does our records so for this the preprocessing is a superb step to enhance it. the speed of the approach is determined on whether or now not the records has been preprocessed. better the preprocessing achieved better may be the result of the version which will one use. first off the author exams for all the null values after which remove the id.

C. Feature Selection

Capabilities are critical factor for purchasing proper effects from the algorithm used. Visualization facilitates us to look the special functions and the way they could make an effect at the effects.

V. RESULT ANALYSIS

Each classification algorithm has its personal characteristics. For various traits, the output of each type algorithm became distinct inside the prediction of heart ailment patients. motives, why these type algorithms behaved like this for this precise dataset like this, is given below:

1) *K-Nearest Neighbors:* On this method, okay- Nearest buddies confirmed bad performance because KNN classifies take a look at records at once from the dataset, no education became carried out before trying out.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

2) *Decision Tree (ID3):* At schooling level, it converted the non-stop value's records into express values and given a range. whilst test statistics pattern contained values out of this given variety, the classifier overall performance became affected and therefore predicts incorrect class label.
3) *Support Vector Machine(SVM):* Support Vector Machine is the one of the most popular supervised learning algorithms, which is used for classification as well as Regression problems. However, primarily, it is used for classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future.
4) *Logistic Regression:* Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No,0 or 1,true or false etc. but instead of giving the exact value as 0 or 1, it gives the probabilistic values which lies between 0 or 1. Logistic Regression is a significant machine learning algorithms because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.
5) *Random Forest:* Random wooded area is an ensemble classification technique that's primarily based on choice Tree set of rules. This set of rules takes a portion of the dataset after which builds a tree, repeat this step for developing a forest by way of combining the generated timber. at the test level, every tree predicts a category label for each check records and majority values of the magnificence label is assigned to the take a look at facts. therefore, it confirmed reasonable performance than conventional decision tree set of rules for this records.

VI. CONCLUSION

This paper compares the performances of the classification algorithms in the prediction of heart disease. It tries to find out the best classifier for this task. In the experimental dataset, 13 attributes were used. But all the attributes are not equally emphasized for detecting heart disease. For this reason, a feature selection method was presented that removes the irrelevant attributes which are not highly correlated with the other features used for classification. Each classification algorithm gives a noticeable performance while using the selected 13 attributes in the prediction of heart disease. Among the studied classifiers, Logistic Regression performs better than other classification algorithms. Binary class problem is solved to identify whether the patient has heart disease or not. It is recommended to solve the multiclass problem for detecting heart disease by dividing heart disease patients into various classes.

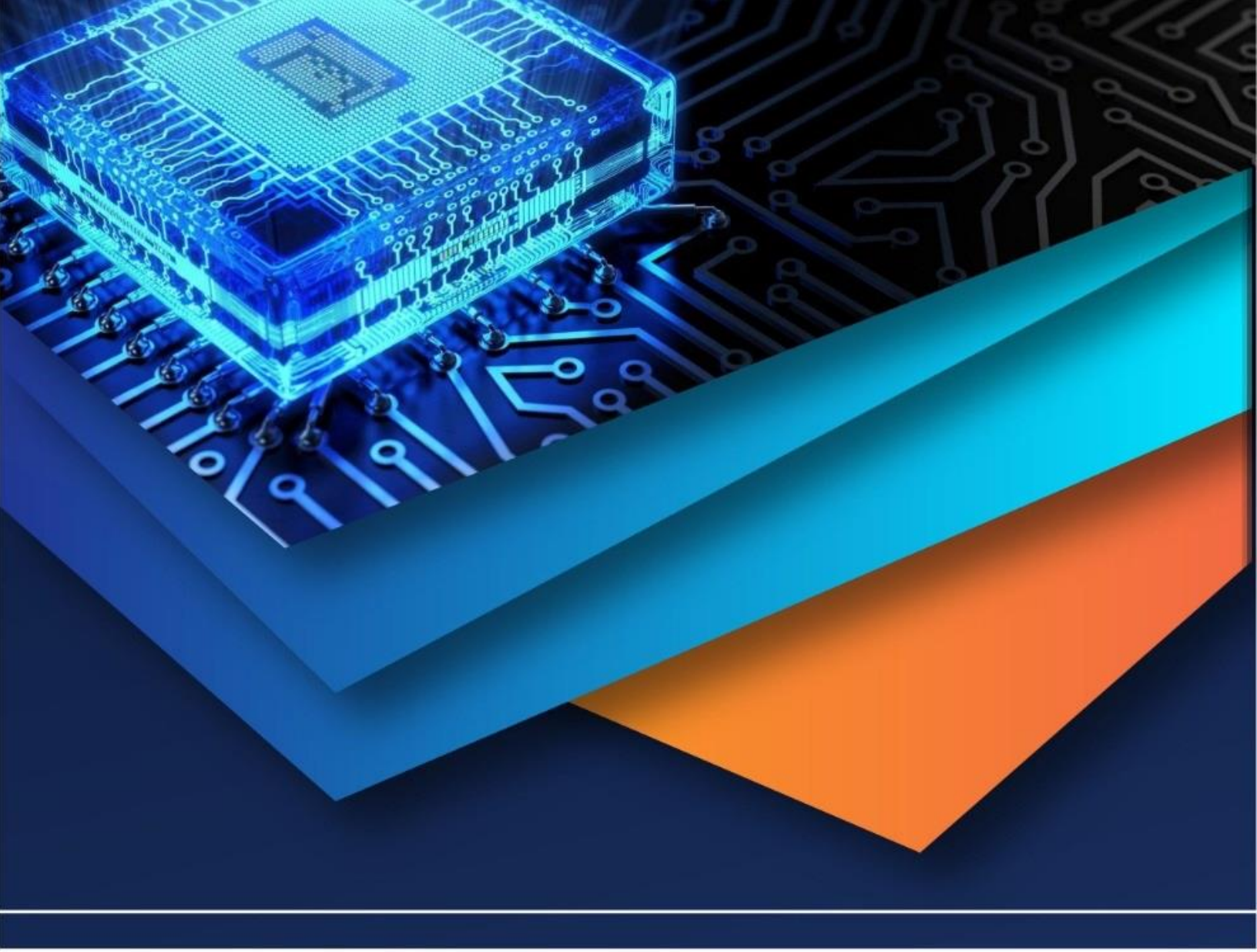
VII. FUTURE SCOPE

Furthermore, we can use the following dataset to test out different regression models and neural networks and see how does it perform. Secondly, we can try this algorithm on a different dataset to know how does it perform and what problems it faces during the testing of the model. This system will be customized to predict not only the presence or absence of heart disease but also to predict the risk factor of heart failure to take extra care of those patients at an early stage and avoid heart failure. Real-time data from different hospitals may be collected for detecting heart disease patients and compute the effectiveness of classifiers for more consistent diagnosis of heart disease patients.

REFERENCES

[1] Apurb Rajdhan, Avi Agarwal, Milan Sai, Dundigalla Ravi ,Dr.Poonam Ghuli, Heart Disease Prediction using Machine Learning, International Journal of Engineering Research Technology (IJERT) 04, April-2020.
[2] Anagha Sridhar, Anagha S Kapardhi, Predicting Heart Disease using Machine Learning Algorithm, International Research Journal of Engineering and Technology (IR- JET) 04 Apr 2019.
[3] Nawal Soliman ALKolifi ALEnezi, A Method Of Skin Disease Detection Using Image Processing And Machine Learning, 16th International Learning Technology Conference 2019.
[4] Neha Prerna Tigga,Shruti Garg, Prediction of Diabetes using Machine Learning,ICCIDS 2019.
[5] M. Kamber and P. J. Han, Data Mining Concepts, and Techniques, 3rd ed., 2012
[6] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Computational Intelligence Technique for Early Diagnosis of Heart Disease," 2015 IEEE International Conference on Engineering and Technology (ICETECH), 20th March 2015.
[7] H. M. Islam, Y. Elgendy, R. Segal, A. A. Bavry and J. Bian, "Risk prediction model for in-hospital mortality in women with ST-elevation myocardial infarction: A machine learning approach," Journal of Heart & Lung, pp. 1-7, 2017..
[8] P. C. Austin, J. V. Tu, J. E. Ho, D. Levy, D. S. Lee, "Using Methods from Data Mining and Machine Learning Literature for Disease Classification and Prediction: a Case Study Examining Classification of Heart Failure Subtypes," Journal of Clinical Epidemiology 66 (2013) pp. 398-407, 2013

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)