iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 10 **Issue:** VII **Month of publication:** July 2022

**DOI:** https://doi.org/10.22214/ijraset.2022.45923

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 10 Issue VII July 2022- Available at www.ijraset.com*

# Heart Disease Prediction Using Machine Learning Technique

Dr. Haressh K[1], Ganesh N[2], Chethan Kumar V P[3], Pavan Naik[4], Gowtham M[5]
[1]*Associate Professor & HOD Dept of CS & E, Govt. Engineering College KR Pet, Mandya- 571426 INDIA*
[2, 3, 4, 5]*UG Student, Dept of CS & E, Govt. Engineering College KR Pet, Mandya, -571426 INDIA*

***Abstract:*** ***Heart sickness has been a significant dangerous issue in a person. Because of a few reasons like dietary patterns, absence of activity, becoming overweight and smoking, and unfortunate way of life propensities cause heart sicknesses. This propels us to find out about the put factors in our life in extreme danger in regards to the reason for coronary illness. The proposed strategy recommends the degrees of chance elements agreeing with the patient's information, with the goal that one can deal with their wellbeing appropriately to forestall the coronary illness. In this examination work, the proposed procedure is to isolate coronary illness patients in light of their gamble factors. As a matter of some importance applying the half assurance is to diminish the records of knowledge and to line up the model speedier. The 5 classifiers area unit strategic relapse, support vector machine, K-closest neighbor, selection tree, and irregular backwoods area unit applied. beyond applying the classifiers, a projected model can disengage the coronary illness patients considering their bet factors in line with the Age of the patients. this may assist the consultants with researching the card-playing parts of the patients.***

## I. INTRODUCTION

Heart jumble is presently an intense disease. Various causes add to coronary illness, that as elevated cholesterol, high BP, harm to blood supply routes, on the off chance that a heart that doesn't work however much it could that prompts coronary illness. Heart harm may likewise cause by an undesirable eating routine, like smoking, drinking a lot of liquor, or eating a lot of sugar, any of which might prompt hypertension. There are different types of coronary illness, like coronary corridor infection, angina, valve sickness, and so on.

The fundamental test in the present way of life in regards to medical care is to get exact medical services and administration. AI and man-made consciousness articulation a principal part in medical services. In light of the upgrades in the field of medical care framework will have on such countless people. What's more, their capacity to save a great many lives and assets, medical services has turned into a center industry for a venture.

In our examination, first, remove the information and undertake the element determination to that dataset. While making a prescient model, including choice is the technique that lessens the number of information factors. We will want to get six distinct elements. For the resultant component determination by utilizing the five managed methods.

## II. EXISTING SYSTEM

The data nuances are gained from the patient. Then from the client inputs, it is analyzed to use ML procedures coronary sickness. The data on coronary sickness patients assembled from the UCI research focus is used to track down plans with NN, DT, LR, and Naive Bayes. The results are differentiated for execution and accuracy and these estimations. The proposed combination methodology returns 87% of the Results.

### *A. Disadvantage*

*1)* Expectation of cardiovascular disease results isn't precise.
*2)* Information mining methods don't assist with giving powerful direction.
*3)* Can't deal with tremendous datasets for patient records.

## III. PROPOSED METHODOLOGY

Execution of python and pandas exercises to perform coronary sickness plan for the data. It gives an easy to-use visual depiction of the dataset.

*A. Advantage*

*1)* Increased exactness for compelling coronary illness determination.
*2)* Handles the most unpleasant measure of information utilizing SVM, KNN calculation, and component determination.
*3)* Reduce the time intricacy of specialists.
*4)* Cost-powerful for patients.

## IV. SYSTEM ARCHITECTURE

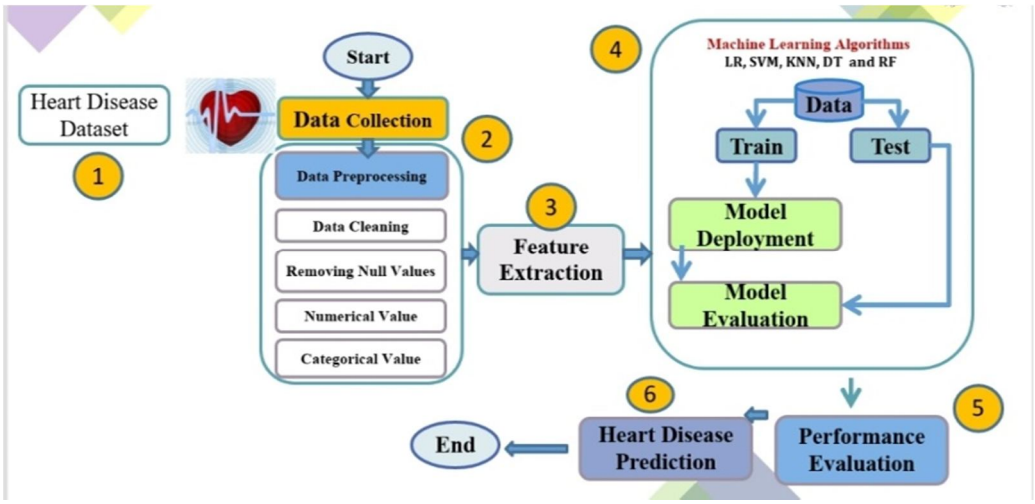
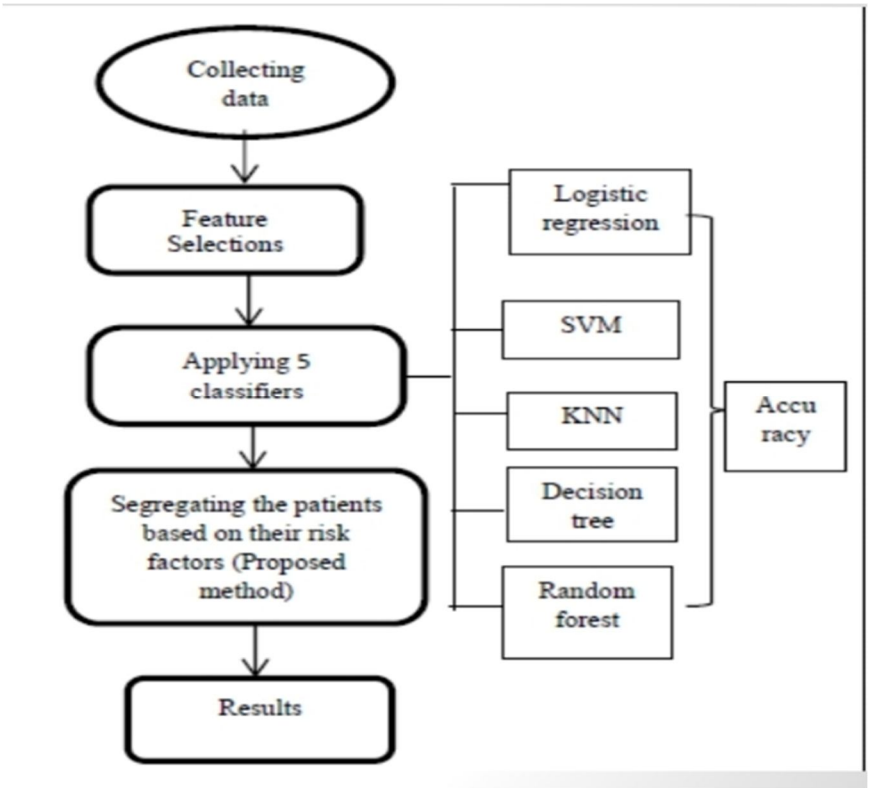



Fig: System Architecture

## V. FLOW CHART



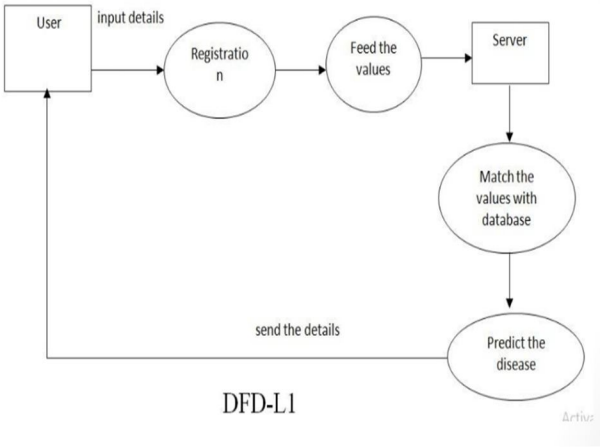
## VI. DATAFLOW DIAGRAM

A data flow chart (DFD) could be a graphical depiction of the "stream" of knowledge through associate degree information structure, showing its cycle viewpoints demonstrating its cycle perspectives. A DFD is much of the time utilized as a primer move toward outlining the framework without carefully describing the situation, which can be utilized for the representation of information handling.

1) *Level 0:* Describes the general course of this undertaking. We are passing quiet subtleties as information; the framework will proficiently process and foresee the outcome utilizing AI calculations

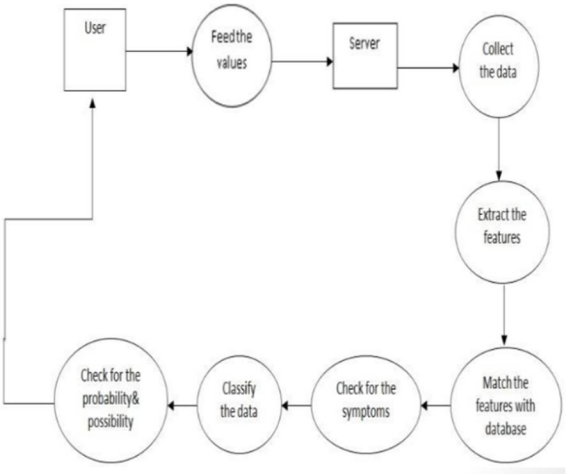

Dataflow Diagram Level-0

2) *Level 1:* Describes the primary stage interaction of this task, we are passing understanding subtleties as a contribution to the framework will processes, remove the elements, and coordinate the qualities with the information base.




DFD-L1

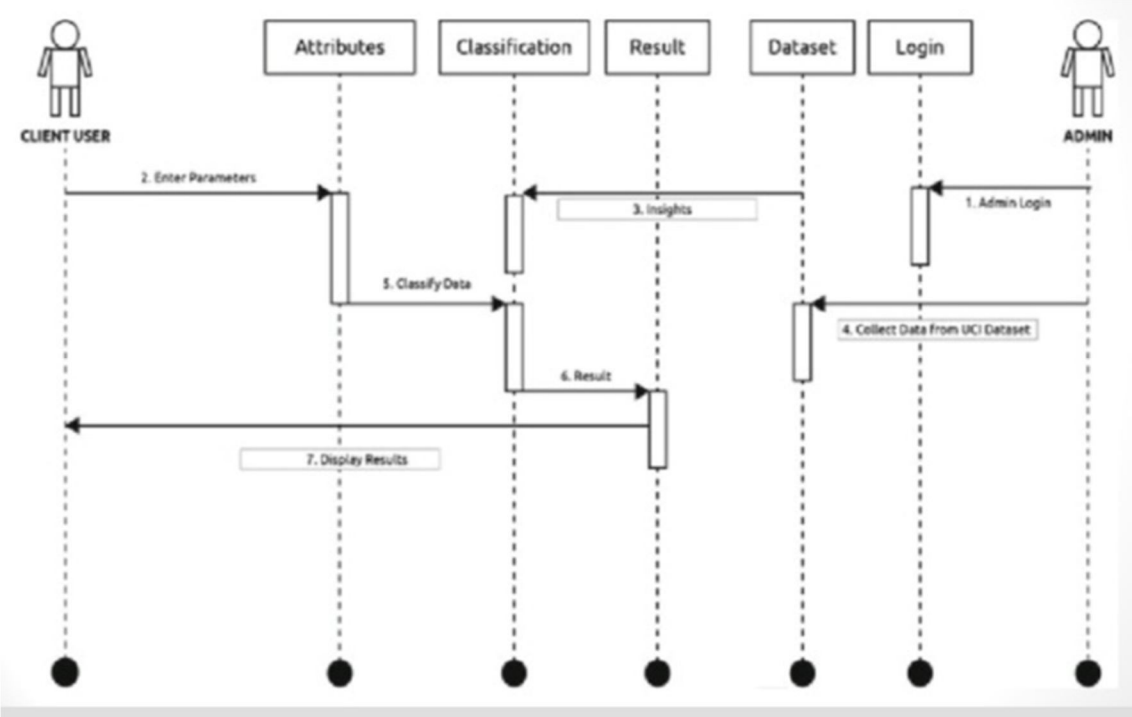
3) *Level 2:* Describe the last stage interaction of this undertaking. We are passing extricated elements to check for the side effects and order the information will anticipate the outcome.



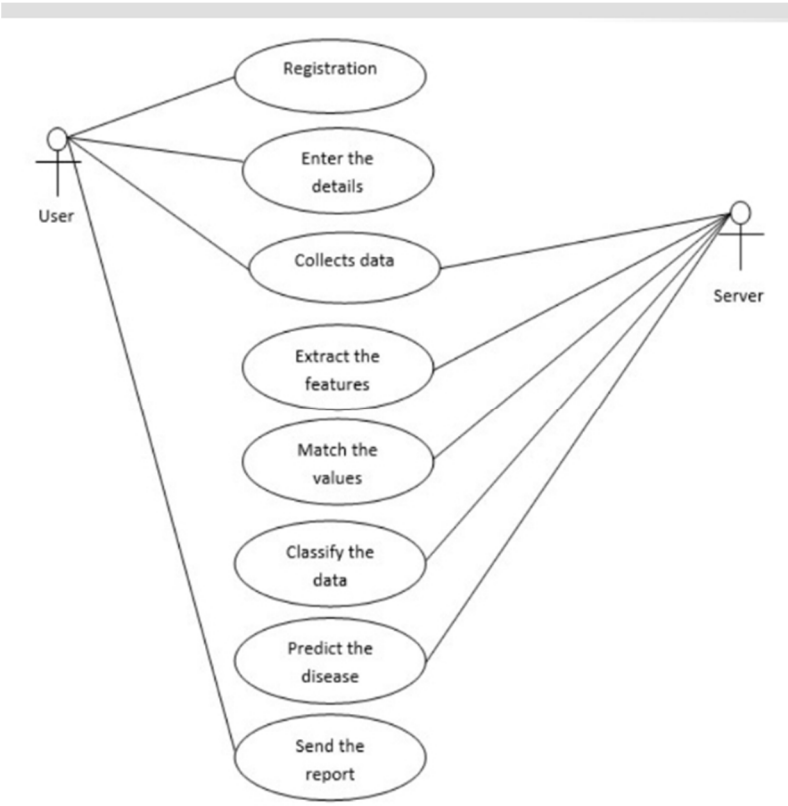

 

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue VII July 2022- Available at www.ijraset.com*

## VII. SEQUENCE DIAGRAM




## VIII. USE CASE DIAGRAM



*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue VII July 2022- Available at www.ijraset.com*

## IX. MODULES

### *A. Module 1: Data Collection*

At first, we gather a dataset for our coronary illness expectation framework. After the range of the dataset, we have a tendency to split the knowledge set into designing knowledge and testing data. The readiness knowledge set is employed for assumption model learning and testing data is employed for evaluating the belief model. For this trip, seventieth of preparing knowledge is employed and half-hour of information is employed for testing. The dataset used for this trip is cardiovascular disease UCI. The dataset contains seventy six properties; out of that, fourteen credits area unit utilised for the framework.

Input Dataset Attributes

| S. No. | Attribute | Description | Type |
|---|---|---|---|
| 1 | Age | Patient's age (29 to 77) | Numerical |
| 2 | Sex | Gender of patient(male-0 female-1) | Nominal |
| 3 | Cp | Chest pain type | Nominal |
| 4 | Trestbps | Resting blood pressure( in mm Hg on admission to hospital ,values from 94 to 200) | Numerical |
| 5 | Chol | Serum cholesterol in mg/dl, values from 126 to 564) | Numerical |
| 6 | Fbs | Fasting blood sugar>120 mg/dl, true-1 false-0) | Nominal |
| 7 | Resting | Resting electrocardiographics result (0 to 1) | Nominal |
| 8 | Thali | Maximum heart rate achieved(71 to 202) | Numerical |
| 9 | Exang | Exercise included agina(1-yes 0-no) | Nominal |
| 10 | Oldpeak | ST depression introduced by exercise relative to rest (0 to .2) | Numerical |
| 11 | Slope | The slop of the peak exercise ST segment (0 to 1) | Nominal |
| 12 | Ca | Number of major vessels (0-3) | Numerical |
| 13 | Thal | 3-normal | Nominal |
| 14 | Targets | 1 or 0 | Nominal |

### *B. Module 2: Data Preprocessing*

Information pre-handling is a significant stage for the production of an AI model. At first, information may not be great or in the normal course of action for the model which can cause deceiving results. In the pre-treatment of data, we change data into our normal design. Overseeing commotions is used, copies, and missing upsides of the dataset. Information pre-handling has the exercises like bringing in datasets, parting datasets, quality scaling, and so on. Pre-handling of information is expected for working on the exactness of the model.

*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue VII July 2022- Available at www.ijraset.com*

### *C. Module 3: Feature Extraction*

Unessential elements frequently influence the model's preparation cycle, and some clamor includes even making the model digress from the right track. Highlight determination picks a subset of factors that can successfully depict the information and guarantee great forecast results.

Some component choice strategies are applied to lessen the impact of commotion or immaterial factors generally summed up into channel techniques, covering techniques, and inserted strategies.

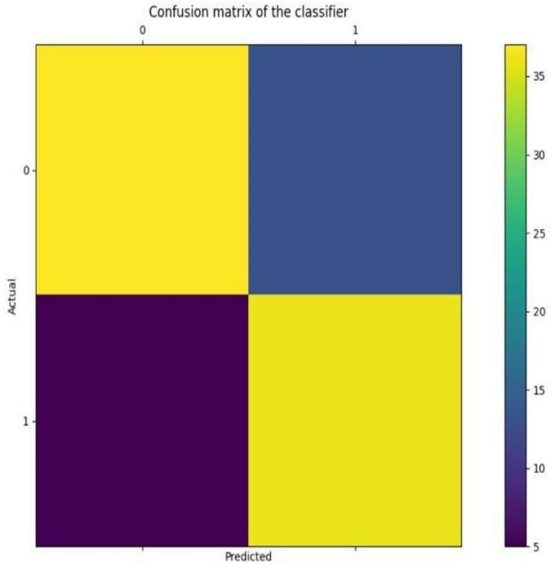
### *D. Module 4: Machine Learning Model Deployment*

Various Supervised Machine Learning Algorithms are used for classifications for example:

*1)* Logistic Regression
*2)* Support Vector Machines
*3)* K-Nearest Neighbor
*4)* Decision Tree
*5)* Random Forest

### *E. Module 5: Performance Evaluation of Models*

Different traits of the patient like orientation, chest torment type, fasting pulse, serum cholesterol, and so on are considered for this task. The precision for individual calculations needs to gauge and whichever calculation gives the best exactness is considered for the coronary illness expectation. For assessing the analysis, different assessment measurements like exactness, disarray framework, accuracy, review, and f1-score are thought of.




Where
TP: True positive
FP: False Positive
FN: False Negative
TN: True Negative

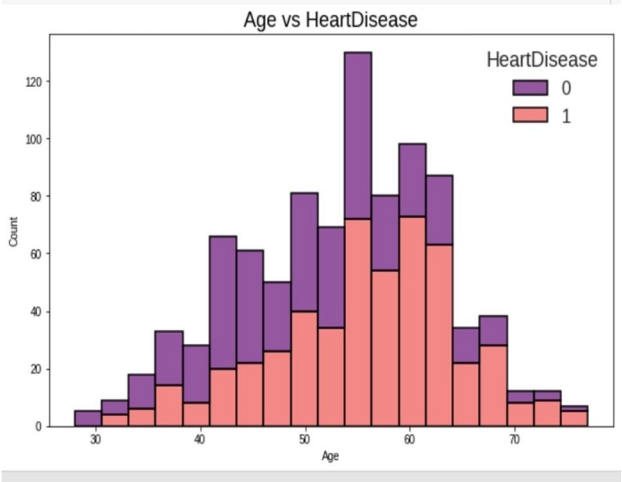
### *F. Module 6: Predictions*

Forecast of Disease

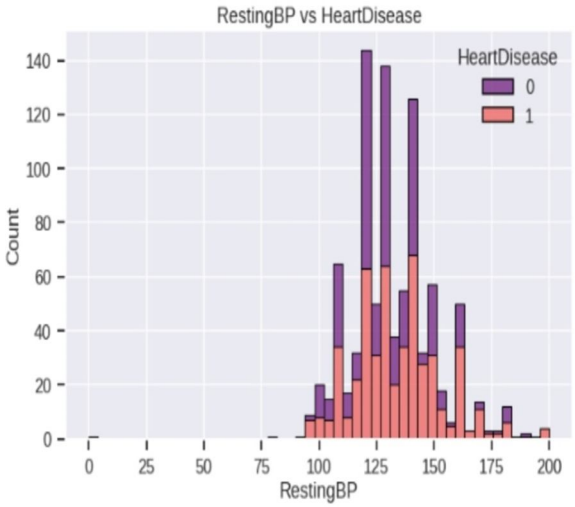
Different AI calculations are utilized for grouping. Near examination is performed among calculations and the calculation that gives the most noteworthy exactness is utilized for coronary illness expectation.

*1) Exploratory Data Analysis (EDA)*

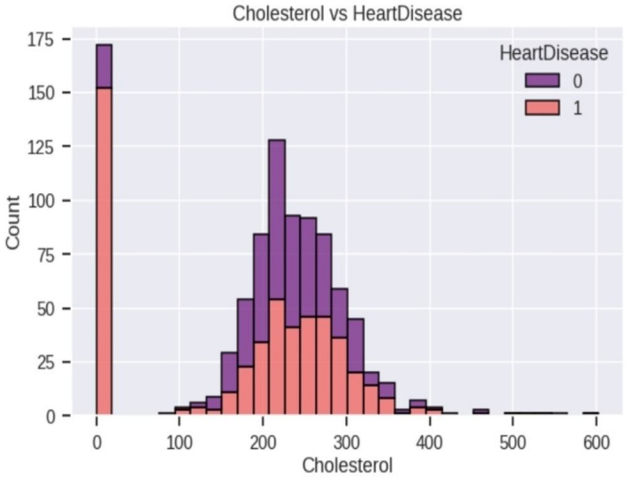
*a)* Age vs. Heart Disease




*b)* RestingBP vs. Heart Disease

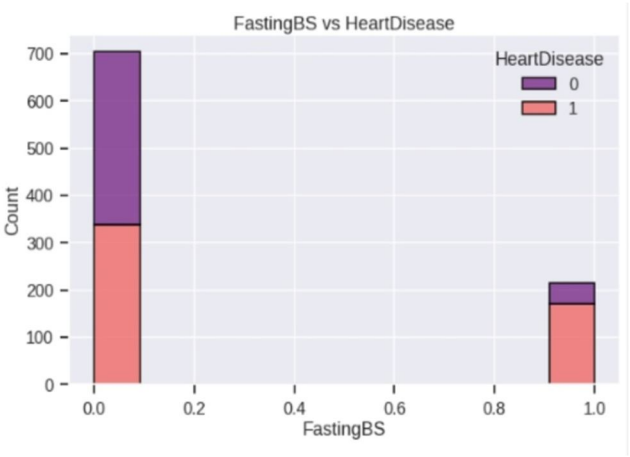



*c)* Cholesterol vs. Heart Disease

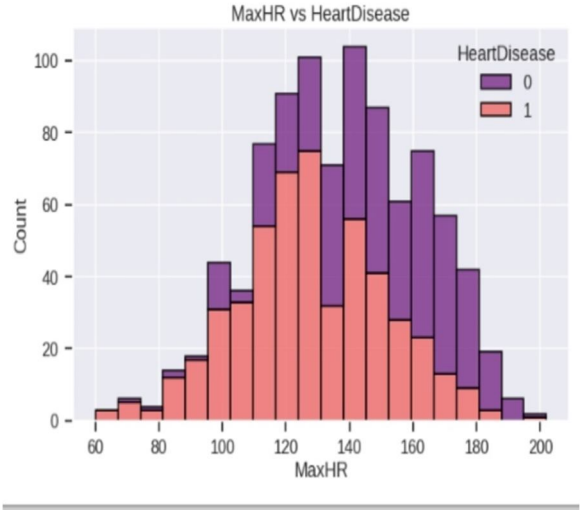


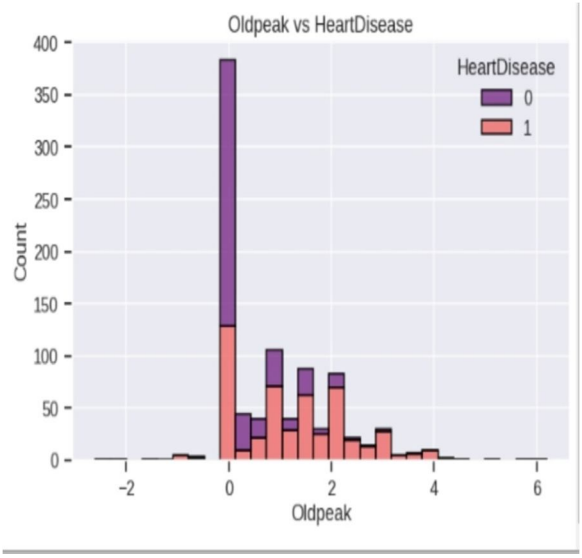
*d)* FastingBS vs. Heart Disease



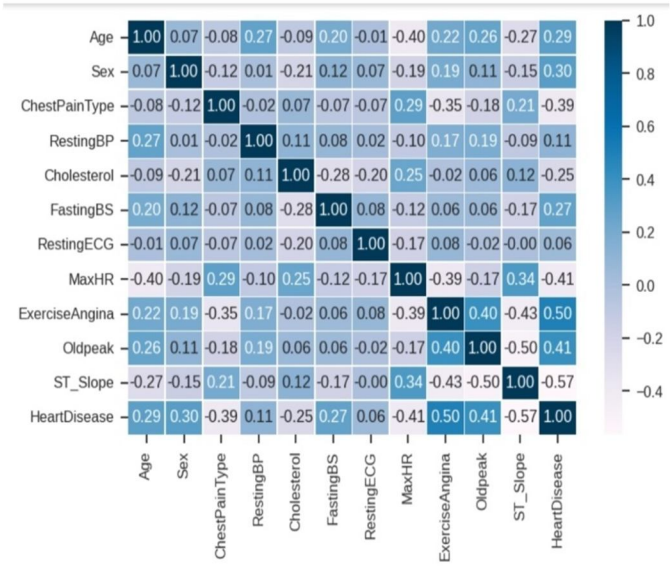

*e)* MaxHR vs. Heart Disease



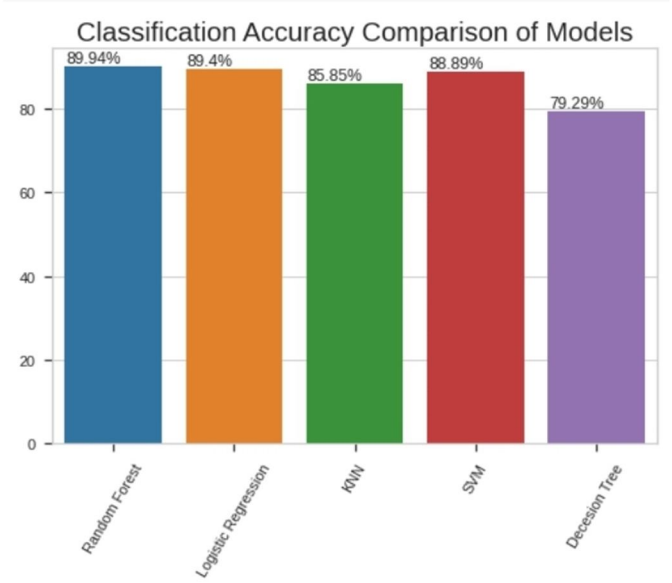

*f)* Oldpeak vs. Heart Disease



*2) Detect Outliers*




*3) Accuracy Comparison Models*




## X. CONCLUSION

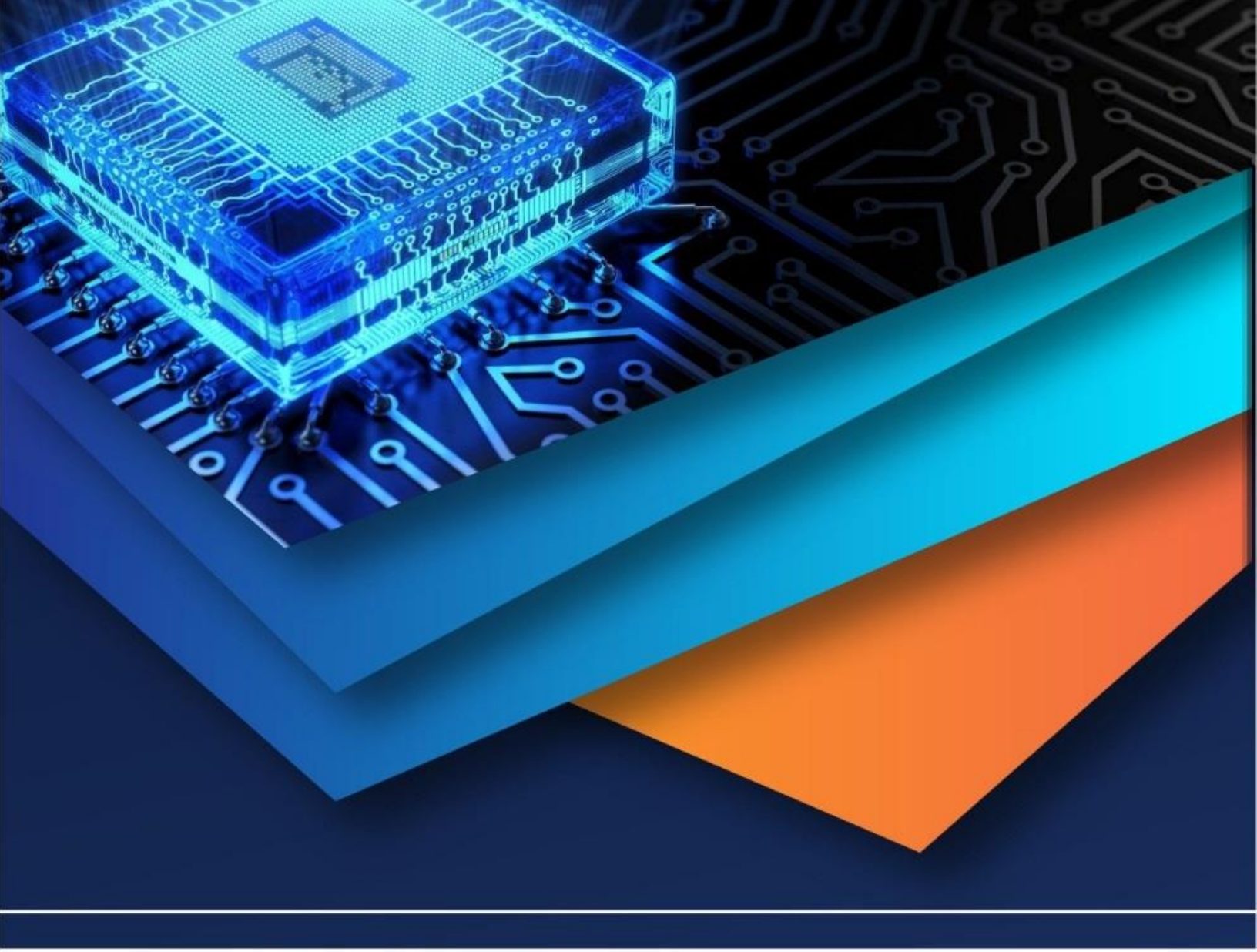
By utilizing SVM and KNN classifiers the precision is estimated and analyzed.

The degrees of chance variables assist the specialists with isolating the patients and treating them.

A few benefits of these strategies are it assists with expanding the exactness. It can deal with both downright and persistent information. This mechanizes the most common way of filling in missing qualities in information.

## REFERENCES

[1] S Hameetha begum, S N Nisharani (24- 26 AUG 2021) | Added to IEEE Xplore on 17 SEP 2021
[2] Machine Learning Basics: Support Vector Machine | K-Nearest Neighbor | by Gurucharan M K | Towards Data Science
[3] M D Amzad Hossein, Taliya Tazin | Hindawi, Mathematical problem of Engineering(10 DEC 2021) | Article ID:1792201
[4] Pooja Gupta, Pritesh Jain, Upendra Singh(June 2018) | International Journal of Scientific Development and Research (IJSDR1806036) | ISSN:24552631
[5] Amayo Mordecai(29 APR 2020) | Towards Data Science.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY