JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EXPLORING MACHINE LEARNING ALGORITHMS FOR HEART DISEASE PREDICTION

Jenifer Shylaja M [1], A Thanam[2] , Lokayya Deepthi [3] , Madireddy Khyathi [4] , Rakshana k [5]

[1,2]Assistant Professor & [3,4,5]UG Scholar Department of Computer Science and Engineering,
Vel Tech High Tech Dr. Rangarajan & Dr. Sakunthala Engineering College, Avadi, Chennai, India

ABSTRACT:

Heart disease is a prevalent and intricate health issue affecting numerous individuals globally, necessitating timely and effective identification, especially in cardiology. Our study employs Fisher Score and Chi-Square as feature selection methods to rank and eliminate irrelevant or redundant attributes. Following feature selection, we employ Support Vector Machine (SVM) and Logistic Regression algorithms for classification. Chi-Square, a mathematical procedure, condenses correlated attributes into fewer, correlated variables, termed principal components, offering a simple means of determining relevant variables. Fisher Score, a widely utilized supervised feature selection method, ranks variables based on Fisher's score in descending order, facilitating variable selection based on case requirements. Chi-Square is particularly useful when dealing with continuous variables. Assuming a target variable is chosen, we assess each parameter to determine if the Chi-Square technique identifies any relationship with the target. Our project proposes an efficient and accurate system for heart disease diagnosis based on machine learning techniques.

KEYWORDS:

Heart disease, Fisher Score, Chi-Square, Feature selection, Support Vector Machine (SVM), Logistic Regression, Principal components, Correlated attributes, Supervised feature selection, Continuous variables, Target variable, Machine learning techniques, Diagnosis, Efficiency, Accuracy

I. INTRODUCTION

Heart disease, the leading cause of global mortality, accounts for around 30% of all deaths. Common symptoms include shortness of breath, weakness, and swollen feet. With approximately 26 million cases diagnosed by the European Society of Cardiology, heart disease affects many, particularly in the United States. Existing diagnostic methods are often ineffective in early detection, leading researchers to seek more efficient techniques. Traditional approaches rely on medical history, physical exams, and symptom analysis but yield inaccurate results and are costly. To address these challenges, a non-invasive machine learning-based diagnostic system is proposed, aiming to improve accuracy and reduce computational complexity. By employing classification algorithms and advanced feature selection techniques like chi-square, the system aims to enhance classification accuracy while minimizing costs and computational burden in clinical settings. The human heart, a vital organ composed of muscles and nerves, pumps approximately 5 liters of blood throughout the body, providing essential nutrients. Given the complexity and criticality of heart function, implementing machine learning, particularly classification algorithms, in medical devices has garnered significant interest among researchers. However, diagnosing heart conditions presents challenges, especially in areas lacking advanced medical equipment and expertise. Physician intuition alone may not suffice, leading to medical errors and adverse outcomes. To address this, there's a growing need for state-of-the-art computer-based diagnosis systems to reduce errors, enhance patient safety, and ultimately save lives. This research focuses on understanding the causes, complications, and treatments of heart diseases, culminating in the development of an intelligent diagnostic system. Utilizing the Statlog Heart Disease dataset from UCI Machine Learning, this system aims to prevent misdiagnosis by accurately identifying the presence or absence of heart disease in patients. The dataset is divided into training, validation, and testing sets to train and evaluate the system's performance. Machine learning is a branch of artificial intelligence focused on developing algorithms that improve with experience. These algorithms analyze sample data, known as "training data," to make predictions or decisions without explicit programming. They find applications in email filtering, computer vision, and other tasks where conventional algorithms are impractical. Machine learning intersects with computational statistics and mathematical optimization, with data mining being an associated field focusing on exploratory data analysis.

II. LITERATURE REVIEW

[1] *Norma Latif Fitriyani , Muhammad Syafrudin, Ganjar Alifian and Jongtae Rhee*, "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System ", Open Access Journal, Volume 8, July 20, 2020.This project was developed a heart disease prediction model that were implemented the clinical decision support system and could be used to help clinicians assess the risheart illness and offer suitable therapies to further control the risk. The authors proposed an effective HDPM for a Clinical Decision Support System which consists of DBSCAN-based to detect and eliminate the outliers, SMOTE- ENN to balance the training data distribution and XGBoost to predict heart disease. The results showed that LR performed better than the other models by achieving up to 85%, 89%, 81%, and 85 for the accuracy, sensitivity, specificity, and precision.

[2] *IEEE member Simanta Shekhar Sarmah's article,* "An Efficient IoT-Based Patient Monitoring and Heart Disease Prediction System Using Deep Learning," Modified Neural Network "Open Access Journal, Volume 8, July 6, 2020". This paper focused on health monitoring with IOT. The body sensors networks (BSN) were formed of disparate wearable or implantable devices, say accelerometer, cardioverter-defibrillator, and pacemaker, which sense as well as monitor the breathing rates, blood pressures, pulse, together with body temperature of users.

[3] *Gamal G. N. Geweid*, (Member, IEEE), and Mahmoud A. Abdallah,(Member, IEEE) , "A New Automatic Identification Method of Heart Failure Using Improved Support Vector Machine Based on Duality Optimization Technique", Open Access Journal, Volume 7, October 4, 2019. This project specifically focused on functions include an initialization function which were used to acquire an ECG signal using a bio-potential amplifier and then displayed it using ECG instrumentation, a preprocessing function, an analyzing function to analyse ECG signal, and a classifier function to classify the heart failure disease.

[4] *SenthilKumar Mohan, ChandraSegar Thirumalai ,And Gautum Srivastava*, (Member, IEEE)**,** "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", Open Access Journal, June 19, 2019. The authors worked on various techniques in data mining and neural networks has been employed to find out the severity of heart disease among humans Numerous techniques, including the Genetic Algorithm, Decision Trees, K-Nearest Neighbor Algorithm, and Naive Bayes, were used to classify the disease's severity. We introduced neural networks using heart rate time series.

[5] *Muhammad Hammad Memon, Amin Ul Haq, Jian Ping Li, Shah Nazir, and Ruinan Sun***,** "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Hindawi -Mobile Information Systems Volume 2018, ,2 December 2018. This project designed a diagnostic system for heart disease and used machine learning classifier multilayer perceptron ANN-driven back propagation learning algorithm and feature selection algorithm. In order to evaluate the performance of classifier, various performance evaluation metrics such as classification accuracy, classification error, specificity, sensitivity, Matthews' correlation coefficient (MCC), and receiver optimistic curves (ROC) were used.

[6] *Sellappan Palaniappan and Rafiah Awang*, "Data Mining Techniques for Intelligent Heart Disease Prediction System," IEEE Journal, 2008. The authors developed a prototype Intelligent Heart Disease Prediction System using three data mining modeling techniques, namely, Decision Trees, Naïve Bayes and Neural Network. Intelligent Heart Disease Prediction System can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. It could answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot.

[7] Amin Ul Haq, Amin Ul Haq, Jianping Li, Jalaiuddin Khan, Muhammad Hammad Memon, Shadma Parveen, Mordecai Folain Raji, Wasfi Akbar, Tanvir Ahmad, Sana Ullah, Latipova Shosista, Happy N Monday, the IEEE Journal, 2019 article titled "Identifying The Predictive Capability Of Machine Learning Classifiers For Designing Heart Disease Detection System."'IEEE Journal,2019.This project enhanced the diagnosis of heart diseases through invoice based techniques as well as ordinary medical based methods were not reliable. The algorithm for feature selection Important features were chosen using relief, and classifier performances were then calculated based on these features. Ensemble machine learning techniques (boosting, bagging, stacking) were utilized to further raise the classifiers performance. According to criteria used to evaluate model performance, the SVM (RBF) did a great job on all features, achieving an accuracy rate of 86%.

III. PROPOSED SYSTEM

The proposed system utilizes supervised classification algorithms like Support Vector Machine (SVM) and Logistic Regression. These algorithms aim to classify data points into different categories based on their features. Support Vector Machine (SVM) works by representing different classes in a multidimensional space, aiming to create a hyperplane that maximizes the margin between different classes. This hyperplane is generated iteratively by the SVM algorithm to minimize errors and accurately classify data points. Logistic Regression, on the other hand, is also a classification algorithm primarily used when the dependent variable is binary (i.e., having two classes). It predicts the probability of an event's success or failure based on input features, fitting a logistic curve to the data to make predictions. In simpler terms, both SVM and Logistic Regression are tools used to separate data into distinct categories based on their characteristics. SVM focuses on finding the optimal hyperplane to maximize the margin between classes, while Logistic Regression estimates the probability of belonging to a certain class.

3.1 ARCHITECHTURE DIAGRAM:

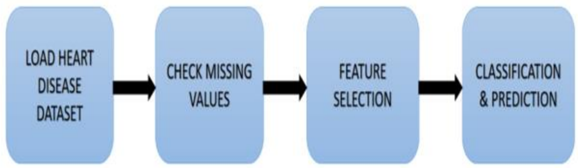


Fig 1. Architecture diagram of dataset

In the above (fig 1) shows system design, the dataset heart disease is loaded as input. The input dataset is taken from UC Irvine (UCI) Machine Learning Repository. The collected input data are subjected to check if there is any missing values. Then the data are split into Train set (80%) and Test set (20%) for decision and Apply fisher score and chi square as feature selection algorithm. Support Vector Machine (SVM) and Logistic Regression algorithms are implemented. Then analyze the performance metrics like Accuracy, Recall, F-Measure and Precision. Finally we predict whether the patient has heart disease or not. Heart Patients count displayed as bar graph in the tkinter window along with classification accuracy comparison chat. Finally an alert dialog box displayed which shows that the result of prediction is done successfully.

3.2 FLOW DIAGRAM:

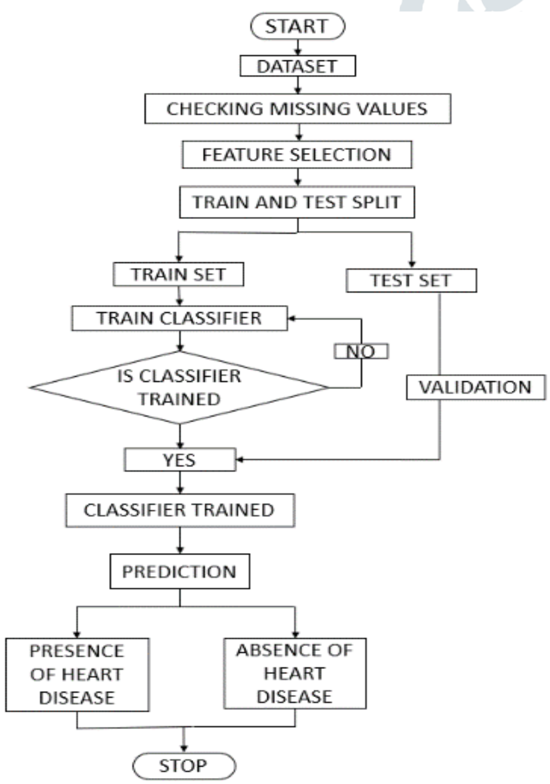


Fig 2. Flow diagram of proposed system

In the above (fig:2) the working and function of the proposed system is discussed. Initially the dataset has to be loaded and check if there is any missing values. The data has to split into the ratio of 8:2 as train set and test set respectively and then feature selection algorithm, fisher score and chi square has been applied. If the classifier is trained, then predict whether the patient has heart disease or absence of heart disease.

IV. MODULE DESCRIPTION

4.1 Heart Disease Dataset:

- This dataset contains information about the transactions age, sex, cp, trestbps, chol, fbs, restecg, thalach, exang, old peak, slope, target attributes.
- There are 304 rows and 14 columns

4.2 Checking missing values:

The gathered dataset is examined to see whether any values are missing.

There are no missing values in our robust dataset.

4.3 Feature selection:

We used fisher score and chi square as feature selection metrics. The qualities were rated using this metric, and any repetitive or superfluous information was eliminated. Support vector machines (SVM) and logistic regression will be utilized for classification after feature selection. to be applied to the datasets. The Fisher score is one of the most often used supervised feature selection methods. By using the fisher's score as the basis, our method generates the ranks of the variables in decreasing order.

The variables can then be selected in accordance with the situation. A set of related attributes is divided into 26 principle components, or smaller sets of correlated variables, using a mathematical technique called chi square. The easiest way to find relevant variables is to use the chi square technique ($\chi 2$). The chi square approach is used when every variable is continuous. Assume that a goal variable is selected. Next, every parameter is scrutinized to determine whether the chi square method pinpoints a connection between it and the objective.

$$\chi^{2}=\sum \frac{\left({O}_{i}-E_{i}\right)^{2}}{E_{i}}$$

χ^2 = chi squared

${O}_i$ = observed value

E_{i} = expected value

V. CLASSIFICATION

5.1 Support Vector Machine (SVM):

In the above (fig 3) supervised machine learning model called a support vector machine (SVM) applies classification methods to two-group classification issues. An SVM model can classify new text after being given sets of labeled training data for every category.

❖ Linear SVM:

This classifier is appropriate for datasets that can be split into two classes by a single straight line, or linearly separable data. Data of this kind are referred to as linearly separable data.

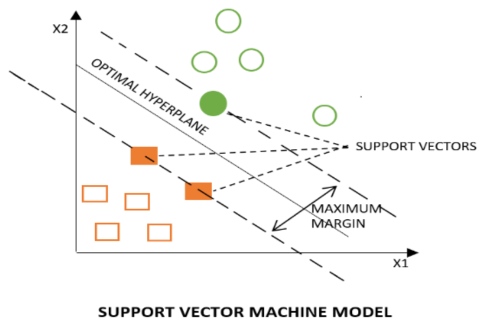


Fig 3. Classification of SVM model

❖ **Non-linear SVM:**

This type of SVM is applied to data that has been non-linearly separated. In other words, if a dataset cannot be classified using a straight line, it is referred to as non-linear

❖ **Hyperplane:**

To divide the classes in n-dimensional space, there can be a number of lines or decision boundaries; nevertheless, we must determine which decision boundary is most useful for classification the informational points. The SVM hyperplane is the name given to this ideal boundary. The features in the dataset determine the hyperplane's dimensions; if there are two characteristics (as in the image), the hyperplane will have a straight line as its shape.

❖ **Support Vectors**

The closest data points or vectors near the hyperplane that have an impact on its position are referred to as support vectors. Given that these vectors a support vector because it provides assistance to the hyperplane.

5.2 Logistic regression:

The given (fig 4) shows, In its most basic form, logistic regression is a statistical model that models a binary dependent variable using a logistic function, though there are many more intricate variations available. Logistic regression in regression analysis is the process of calculating a logistic model, a binary regression type of analysis.

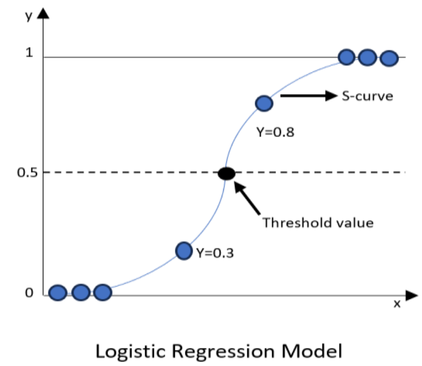


Fig 4. Analysis of logistic regression model

In logistic regression, we fit a "S" shaped regression model rather than a regression line. two maximum values are predicted by the logistic function (0 or 1). With logistic regression, the result of a categorical dependent variable is predicted. The outcome must therefore be a discrete or category value. Instead of providing the exact values, which are 0 and 1, it provides the probabilistic values, which fall between 0 and 1 It can be expressed as True or False, as well as as Yes or No.

IV. HARDWARE AND SOFTWARE COMPONENTS

6.1 Hardware requirements:

The system is implemented using the following hardware components.

- RAM : 4 GB or above
- Hard Disk : 1000 GB or more 3.2

6.2 Software requirements:

The system is implemented using the following software components

- Operating System : Windows 7 or above
- Language : python
- IDE : Jupyter Notebook

6.3 Software description:

6.3.1 Python:

High-level, general-purpose, interpreted programming is done with Python. Guido Van Rossum created Python, which was first made accessible in 1991. Its design philosophy heavily utilizes whitespace and places a high priority on readability of the code. It provides clearly-defined building elements for both small- and large-scale programming. The programming language Python is multi-paradigm and fully supports both object-oriented and structured programming.

6.3.1.1 Features of python:

• Simple to read: Python code is easier to understand and more readable for the human eye.

• Easy to maintain - The source code for Python is not too difficult to maintain.

• Interactive Mode: Python includes an interactive mode that enables interactive debugging and testing of short code segments.

• Portable: Python has an identical user interface across a broad range of hardware systems.

• Databases: All of the main commercial databases have interfaces available in Python.

• Scalable: Compared to shell programming, Python offers larger projects more structure and assistance.

6.3.3 Anaconda:

The most widely used Python data science 9platform is called Anaconda. It is the quickest the simplest method for performing data science and machine learning with Python and R on Windows, Linux, and Mac OS X. For designing, testing, and training on a single machine.

VII. DATASET

The given (Table 1) shows, The UC Irvine Machine Learning Repository contains a dataset on cardiac illness that is utilized for detection. The transactions' age, sex, cp, trestbps, chol, fbs, restecg, and thalach, exang, slope, oldpeak, and target characteristics are all included in this dataset. 14 columns and 304 rows are present.

S.No	Attributes	Description	Domain of Values
1	Age	Age in Years	29-77
2	Sex	Sex	Male (1) Female (0)
3	Cp	Chest Pain Type	Typical Angina (0) Atypical Angina (1) Non-Anginal (2) Asymptomatic (3)
4	Trestbps	Resting Blood Sugar	94 to 200 mm Hg
5	Chol	Serum Cholesterol	126 to 564 mg/dl
6	Fbs	Fasting Blood Sugar	Greater than 120 mg/dl True (1) False(0)
7	Restecg	Resting ECG Result	Normal (0) ST-T Wave abnormality (1) LV Hypertrophy (2)
8	Thalach	Maximum Heart Rate Achieved	71 to 202
9	Exang	Exercise Induced Angina	Yes (1) No (0)
10	Oldpeak	ST Depression induced by exercise relative in rest	0 to 6.2
11	Slope	Slope of peak exercise ST Segment	Unsloping (0) Flat (1) Downsloping (2)
12	Ca	Number of Major Vessels Coloured by Fluoroscopy	0 - 3
13	Thal	Defect Type	Normal (1) Fixed Defect (2) Reversible Defect(3)
14	Target	Heart Disease	0 = Absence 1 = Presence

Table 1. Attributes Information of Heart Disease Dataset

Table 1. Dataset of attributes information

VII. PREDICTION ALALYSIS

7.1 Confusion Matrix:

In the given (fig 5) shows, It's the simplest system for assessing how well a bracket task performs when the result can belong to two or further different class types. All that exists is a table with two confines," factual" and" prognosticated," and also, both the As demonstrated below, the confines include" True Cons(TP),"" True Negatives(TN),"" False Cons(FP)," and" False Negatives(FN)".

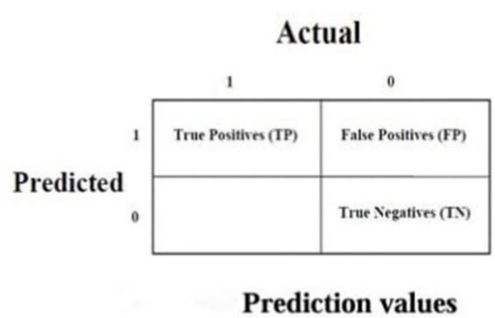


Fig 5. Predicted values of dataset

The following is an explanation of the terms used in the confusion matrix.

• True Cons(TP) This occurs when a data point's factual and anticipated classes are both 1.

• True Negatives(TN) − This occurs when a data point's factual and awaited classes are both 0.

• False Cons(FP) This occurs when a data point's projected class is 1 but its factual class is 0.

• False Negatives(FN) This occurs when a data point's factual class is 1 and its projected class is 0.

Confusion Matrix of our bracket model may be reckoned using the confusion matrix function ofsklearn.metrics

7.2 Classification Accuracy:

It's the most extensively used performance index for algorithms that classify data. It can be characterized as the rate of all the prognostications made to the number of accurate prognostications. The following formula can be used to calculate it by confusion matrix with ease. The delicacy of our bracket model may be calculated using the delicacy score function ofsklearn.metrics.

7.3 Classification Report:

The precision, Recall, F1, and Support scores are included in this report. They're explained as follows

7.3.1 Precision:

The number of accurate documents our ML model returns can be employed to measure perfection when it comes to document reclamations. The following (fig 6) formula can let us calculate it by confusion matrix with ease.

$$Precision = \frac{TP}{TP + FP}$$

Fig 6. Formula for precision

7.3.2 Recall or Sensitivity:

The volume of positive results our ML model delivers can be appertained to as recall. With the aid of the confusion matrix and the following (fig 7) formula, we can snappily cipher it

$$Recall = \frac{TP}{TP + FN}$$

Fig 7. Formula for recall

7.3.3 Support:

Support may be defined as the number of samples of the true response that lies in each class of target values.

7.3.4 F1 Score:

F1 Score We'll gain the harmonious mean of recall and perfection from this score. The weighted normal of recall and perfection is the fine representation of the F1 score. F1's optimal value is 1, and its worst value is 0. We may use to determine the F1 score Recall and perfection have an equal commensurate donation to the following formula F1 score.

F1 = 2 ∗(perfection ∗ RECALL)/(perfection RECALL)

We're suitable to To gain the bracket report for our bracket model, use the sklearn. criteria bracket, report function.

VIII. STEPS INVOLVED IN IMPLEMENTATION

8.1 Import Required Package:

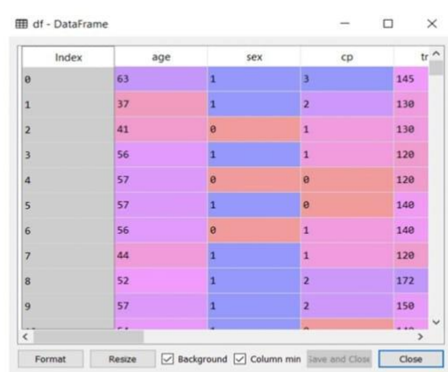
From (fig 8) import needed Package Import Pandas, matplotlib(pyplot), seaborn and warnings libraries.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
import warnings
```

Fig 8. packages to import

8.2 Read the csv file:

From (fig 9) Select and cargo the dataset from the UCI Repository.



Index	age	sex	cp	tr
0	63	1	3	145
1	37	1	2	130
2	41	0	1	130
3	56	1	1	120
4	57	0	0	120
5	57	1	0	140
6	56	0	1	140
7	44	1	1	120
8	52	1	2	172
9	57	1	2	150

Heart Disease Dataset

Fig 9. Dataset of heart disease

8.3 Check if there is missing values:

From (fig 10) This phase requires us to display the dataframe with its type and determine whether any values are missing from it.

```
print(df.isna().sum())
print("\n")

"""Checking the Dataframe type"""
print(df.info())
```

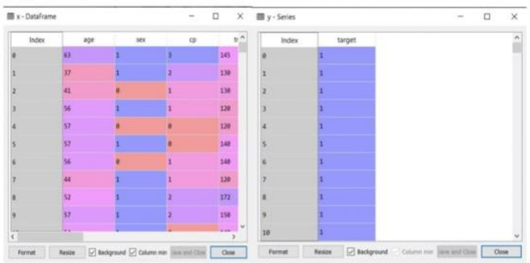
Fig 10. Code to display dataframe

8.4 Define X and Y:

In (fig 11) the input features we wish to take into account are stored in X(Marker). The affair value(Target) is stored in Y.

```
x = df.drop("target",axis = 1)
y = df["target"]
```

Fig 11. To print input values of X and Y



X Dataframe Y Series

Fig 12. Dataframe of X and series of Y

8.5 Divide Data into a Training and Testing Dataset and apply feature selection algorithm:

As shown in (fig 13) Divided the data into an 8:2 train and test set, which was then fed to the SVM model for model fitting, prediction, and scoring. The chi square algorithm and fisher score were then used. For the logistic regression algorithm, the same process was done.

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.2,random_state = 42)

f_score = chi2(x_train, y_train)
f_score

p_values = pd.Series(f_score[1], index = x_train.columns)
p_values.sort_values(ascending = True, inplace = True)
p_values
p_values.plot.bar()
plt.title('pvalues with respect to features')
```

Fig 13. code to divide the data in to an 8:2 train and test

8.6 Train the model:

In (fig 14) Make the model. Using 80% of the data, train our model. In order for our data to have trained in accordance with that model, fit the data into it.

Prediction values

Fig 14. Predicted values for input

8.7 Prediction Analysis:

As shown in (fig 15), The total classification and forecast will be the basis for generating the ultimate outcome. A few of the metrics are used to assess how well this suggested method performs.

- Precision
- Recall
- F1 Measure
- Support

```
classification_report(y_pred,y_test)

confusion_matrix(y_pred,y_test)

(accuracy_score(y_pred,y_test)*100)
```

Fig 15. Classification and forecast

8.8 TKinter

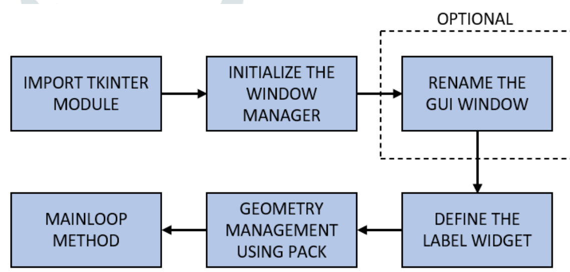


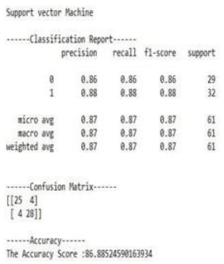
Fig 16. Workflow diagram of Tkinter

A popular Python package used to construct GUI apps is called TKinter. With Tkinter, creating a graphical user interface is incredibly simple and quicker. Several widgets from Tkinter can be used to create graphical user interfaces. Among them are radio buttons, buttons, and boxes, etc. Heart Disease In Tkinter, the patient count is displayed as a bar graph. window and Pie Chart together. Finally, the Tkinter Window will show the accuracy comparison bar graph chart. It is frequently included with Python and uses Tk, which is the default GUI framework for Python. It is well-known for having an easy-to-use graphical user interface. It may be obtained under the Python License and is open-source. Which is given in the below (fig 16).

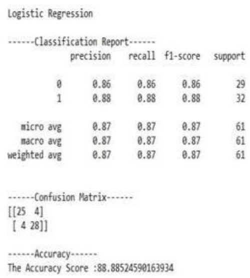
IX. RESULTS AND DISCUSSION

With an accuracy of 86.88% and 88.88%, respectively, this system can determine if a patient has heart disease or not using the Support Vector Machine and Logistic Regression Classification Algorithm. Lastly, it shows the people with cardiac problems. total and categorization The Tkinter window displays an accuracy comparison chart as a bar graph. At the conclusion, an alert dialog box appears indicating that the prediction was successful. With the help of this method, accuracy is increased, patients are protected from heart disease at an earlier stage, and heart disease can be somewhat treated.

9.1 Snapshot:



Classification Report, Confusion Matrix and Accuracy for SVM



Classification Report, Confusion Matrix and Accuracy for LR

Fig 17. Classification report, confusion matrix and accuracy for SVM and LR

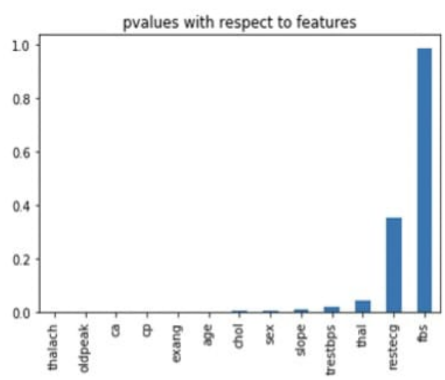
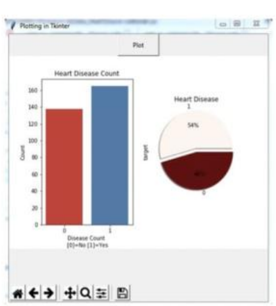
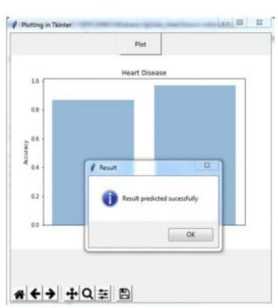


Fig 18. Pvalues with respect to featues



Heart Disease Count



Comparison Chart of SVM and LR

Fig 19. Count of heart disease and comparison chart of SVM and LR

X. CONCLUSION AND FUTURE WORK

In order to rank the features toward the classification of heart illness, we employed various characteristics and support vector machines (SVM and logistic regression) in this research. This indirectly lowers the number of diagnostic procedures that a patient must undergo. According to our experimental findings, feature subset selection and Chi Square generally result in improved classification accuracy and dimensionality reduction. We offer an approach that removes erroneous and misleading data. This will support the development of a quick and accurate automated system for diagnosing cardiac problems, which might potentially save lives.

To further improve performance, it may be possible to use intelligent agents to contribute extensions or revisions to the suggested clustering and classification methods in the future. In addition to the tried-and-true mix of data mining methods, To increase the accuracy, additional combinations like soft computing, artificial intelligence, and other clustering methods can be applied.

XI. REFERENCES

[1] Norma Latif Fitriyani ,Muhammad Syafrudin, Ganjar Alifian and Jongtae Rhee, "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System ", Open Access Journal, Volume 8, July 20, 2020.

[2] Simanta Shekhar Sarmah, (member,IEEE), "An Efficient IoT-Based Patient Monitoring and Heart Disease Prediction System Using Deep Learning Modified Neural Network" Open Access Journal, Volume 8, July 6, 2020.

[3] Gamal G. N. Geweid, (Member, IEEE), and Mahmoud A. Abdallah, (Member, IEEE) , "A New Automatic Identification Method of Heart Failure Using Improved Support Vector Machine Based on Duality Optimization Technique", Open Access Journal, Volume 7, October 4, 2019.

[4]Senthil Kumar Mohan, Chandra Segar Thirumalai ,And Gautum Srivastava, (Member, IEEE), "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques", Open Access Journal, June 19, 2019.

[5] Amin Ul Haq, Amin Ul Haq, Jianping Li, Jalaiuddin Khan, Muhammad Hammad Memon, Shadma Parveen, Mordecai Folain Raji, Wasfi Akbar, Tanvir Ahmad, Sana Ullah, Latipova Shosista, Happy N. . Monday, the IEEE Journal, 2019 article titled "Identifying The Predictive Capability Of Machine Learning Classifiers For Designing Heart Disease Detection System."

[6] Muhammad Hammad Memon, Amin Ul Haq, Jian Ping Li, Shah Nazir, and Ruinan Sun, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms", Hindawi -Mobile Information Systems Volume 2018, ,2 December 2018.

[7] Sellappan Palaniappan and Rafiah Awang, "Data Mining Techniques for Intelligent Heart Disease Prediction System," IEEE Journal, 2008.

[8] Gouthami Kokkula, Isha Pandya, Aditi Gavhane, Kailas Devadkar presents his work on "Prediction" in the Proceedings

of the 2nd International Conference on Electronics, Communication and Aerospace Technology (ICECA 2018) of Heart Disease Using Machine Learning."

[9] A. U. Haq, J. P. Li, M. H. Memon, J. Khan, A. Malik, T. Ahmad, A. Ali, S. Nazir, I. Ahad, and M. " "Feature selection based on L1-norm support vector machine and efficient Parkinson's disease recognition system using voice recordings," stated Shahid.IEEE Access, vol. 7, pp.37718–37734, 2019.

[10] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig said, "The average severity of Parkinson's disease symptoms can be clinically usefully quantified through nonlinear speech analysis algorithms mapped to a standard metric." J.Roy. Soc. Interface, vol. 8, no. 59, pp. 842–855, 2011.

[11] E. O. Olaniyi, O. K. Oyedotun, and K. Adnan, ''Heart diseases diagnosis using neural networks arbitration,'' Int. J. Intell. Syst. Appl., vol. 7, no. 12, pp. 72, 2015.

[12] "Quick binary feature selection using conditional mutual information," Fleuret, J.J. Mach. Learn. Res., vol. 5, pp. 1531–1555, Nov. 2004.

[13] J. Lopez-Sendon, ''The heart failure epidemic,'' Medicographia, vol. 33, no. 4, pp. 363–369, 2011.

[14] J. H. Gennari, P. Langley, and D. Fisher, ''Models of incremental concept formation,'' Artif. Intell., vol. 40, nos. 1–3, pp. 11–61, Sep. 1989.

[15] L. Zhu, J. Shen, L. Xie, and Z. Cheng, ''Unsupervised topic hypergraph hashing for efficient mobile image retrieval,'' IEEE Trans. Cybern., vol. 47, no. 11, pp. 3941–3954, Nov. 2017.