IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 7, April 2023

# Diagnosis of Heart Disease using Machine Learning Algorithms

**K. Pazhanivel[1], A. Pravin[2], S. Ramji[3], M. Sakthiganesh[4], M. Saravanan[5]**

Assistant Professor, Department of Computer Science Engineering[1]

Students, Department of Computer Science and Engineering[2,3,4,5]

Anjalai Ammal Mahalingam Engineering College, Thiruvarur, India

**Abstract:** *The diagnosis of heart disease using machine learning algorithms is a promising field that has the potential to improve the accuracy and efficiency of diagnoses. This paper discusses four machine learning algorithms, including Support Vector Machines (SVM), Decision Trees, and Logistic Regression, that can be used for heart disease diagnosis. These algorithms work by identifying the most relevant features related to heart disease, pre-processing the dataset, and fine-tuning the models for optimal performance. Once well-performing models are obtained, they can be deployed for real-time diagnosis. The use of machine learning algorithms for heart disease diagnosis can provide faster, more accurate, and more efficient diagnoses, leading to better patient outcomes and reduced healthcare costs. However, careful hyper parameter tuning and dataset preprocessing are required to ensure optimal performance. Overall, the use of machine learning algorithms for heart disease diagnosis shows great promise and has the potential to revolutionize the way heart disease is diagnosed and treated.*

**Keywords:** Heart Disease Prediction

## I. INTRODUCTION

Heart disease is a leading cause of death worldwide and can be challenging to diagnose accurately. Traditional diagnostic methods, such as electrocardiography (ECG) and echocardiography, rely on expert interpretation and can be time-consuming and expensive. In recent years, there has been growing interest in using machine learning algorithms for heart disease diagnosis. Machine learning algorithms are a subset of artificial intelligence that can learn from data and make predictions based on that data. These algorithms can identify complex patterns and relationships within datasets that may not be immediately apparent to human experts. By using machine learning algorithms, it is possible to develop more accurate and efficient diagnostic tools for heart disease. In this paper, we discuss four machine learning algorithms, including Random Forest, Support Vector Machines (SVM), Decision Trees, and Logistic Regression, that can be used for heart disease diagnosis. We provide an overview of how each algorithm works and discuss their strengths and limitations. We also discuss the importance of data pre-processing and hyper parameter tuning in developing accurate and efficient diagnostic models. Finally, we highlight the potential benefits of using machine learning algorithms for heart disease diagnosis, including faster and more accurate diagnoses, improved patient outcomes, and reduced healthcare costs. it also has some limitations, and its use should be carefully evaluated. This article discusses the potential of machine learning techniques for the efficient medical diagnosis of heart diseases and the role of Support vector machine in optimizing their performance. Overall, the use of machine learning algorithms for heart disease diagnosis is a promising field with significant potential to improve patient care and reduce healthcare costs.

### Decision Tree Algorithm

Decision Trees are a popular machine learning algorithm that can be used for heart disease diagnosis. Decision trees work by recursively splitting the dataset into smaller subsets based on the most significant features, creating a tree-like model that can be used for classification. In the context of heart disease diagnosis, the dataset typically includes various features related to heart disease, such as age, sex, blood pressure, cholesterol level, and smoking history. The goal is to build a decision tree that can accurately classify patients as either having or not having heart disease based on these

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 7, April 2023

features. The decision tree consists of nodes that represent a feature and edges that represent the possible outcomes of that feature. At each node, the algorithm selects the feature that provides the most information gain, meaning the feature that can best separate the dataset into the two classes (heart disease or no heart disease). The algorithm then splits the dataset based on that feature and repeats the process until it reaches the leaf nodes, which represent the final classification. One advantage of Decision Trees is that they can handle missing values and non-linear relationships between features, making them suitable for heart disease diagnosis. Additionally, decision trees are easy to interpret, as the resulting tree structure provides clear insight into the decision-making process. However, Decision Trees can suffer from over fitting, where the model becomes too complex and fits the training data too closely, resulting in poor generalization to new data. To avoid over fitting, techniques such as pruning and setting a maximum tree depth are commonly used. In summary, Decision Trees are a powerful machine learning algorithm that can be used for heart disease diagnosis. By recursively splitting the dataset into smaller subsets based on the most significant features, Decision Trees can accurately classify patients as either having or not having heart disease based on various features related to heart disease.

**Logistic Regression**

Logistic Regression is another machine learning algorithm that can be used for heart disease diagnosis. Unlike linear regression, which is used for continuous variables, logistic regression is used for binary classification problems, where the goal is to predict a binary outcome (e.g., heart disease or no heart disease) based on one or more input variables. In the context of heart disease diagnosis, logistic regression models can be used to predict the probability of a patient having heart disease based on various features related to heart disease, such as age, sex, blood pressure, cholesterol level, and smoking history. Logistic regression models work by fitting a logistic function to the data, which is an S-shaped curve that maps any input to a probability output between 0 and 1. The logistic function takes the form of: $P(y=1|X) = 1 / (1 + \exp(-(\beta 0 + \beta 1x1 + \beta 2x2 + ... + \beta nxn)))$ where $P(y=1|X)$ is the predicted probability of having heart disease given the input features X, and β0, β1, β2, ..., βn are the coefficients of the logistic regression model. To train the logistic regression model, the algorithm optimizes the coefficients to maximize the likelihood of the observed data. This process is known as maximum likelihood estimation. One advantage of logistic regression is that it is a relatively simple algorithm that is easy to implement and interpret. Additionally, logistic regression can handle both continuous and categorical input variables, making it suitable for heart disease diagnosis. However, logistic regression assumes that the relationship between the input variables and the output variable is linear, which may not always be the case in real-world datasets. Additionally, logistic regression can suffer from overfitting if the model is too complex, which can lead to poor generalization to new data. In summary, logistic regression is a simple and powerful machine learning algorithm that can be used for heart disease diagnosis. By fitting a logistic function to the data, logistic regression models can predict the probability of a patient having heart disease based on various features related to heart disease.

**Support Vector Machine**

Support Vector Machine (SVM) is a powerful machine learning algorithm that can be used for heart disease diagnosis. SVMs are a type of supervised learning algorithm that can be used for both classification and regression tasks. In the context of heart disease diagnosis, SVMs can be used to classify patients as either having or not having heart disease based on various features related to heart disease, such as age, sex, blood pressure, cholesterol level, and smoking history. SVMs work by finding the optimal hyperplane that separates the data into the two classes. The hyperplane is defined by a set of weights and biases that are learned during the training process. The goal of the SVM algorithm is to find the hyperplane that maximizes the margin, which is the distance between the hyperplane and the closest data points of each class. The margin is important because it helps to ensure that the SVM model is robust to new data. In cases where the data is not linearly separable, SVMs use a technique called kernel trick to transform the input data into a higher-dimensional feature space, where it may be easier to separate the data into the two classes. The most commonly used kernel functions are linear, polynomial, and radial basis function (RBF) kernels. One advantage of SVMs is that they are effective in high- dimensional spaces and can handle both linearly and non-linearly separable data. Additionally, SVMs are less prone to overfitting than other machine learning algorithms, such as decision trees and neural networks. However, SVMs can be sensitive to the choice of kernel function and the hyperparameters of the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

model, such as the regularization parameter and kernel coefficient. Additionally, SVMs can be computationally expensive, especially for large datasets. In summary, SVMs are a powerful machine learning algorithm that can be used for heart disease diagnosis. By finding the optimal hyperplane that separates the data into the two classes, SVMs can accurately classify patients as either having or not having heart disease based on various features related to heart disease

## II. HEART DISEASE PREDICTION

Heart disease prediction is the process of using machine learning algorithms to predict the likelihood of a patient developing heart disease based on various risk factors and patient data. Heart disease is a leading cause of death worldwide, and early detection and prevention are essential for improving patient outcomes. Machine learning algorithms can analyze large datasets of patient information and identify patterns and correlations that can be used to predict the likelihood of heart disease. The risk factors considered for heart disease prediction can include age, sex, family history, blood pressure, cholesterol levels, smoking habits, and medical history. The heart disease prediction models use supervised learning techniques to learn from the patient data and predict the likelihood of heart disease. The models are trained on a dataset that includes patient information and the corresponding outcomes, such as the occurrence of heart disease. The algorithm learns the patterns in the data and uses them to make predictions on new patient data. The accuracy of heart disease prediction models depends on the quality and quantity of the patient data used to train the models. Therefore, it is important to collect accurate and sufficient patient data to ensure the models' effectiveness. Heart disease prediction using machine learning has several advantages. It can identify patients at high risk for heart disease and enable early intervention and prevention. It can also reduce the cost of healthcare by reducing the need for invasive tests and procedures. However, there are also challenges in heart disease prediction using machine learning. The algorithms must be able to handle missing and incomplete data, and the models must be validated on new datasets to ensure their generalizability. Additionally, the models must be transparent and interpretable for medical professionals to understand and trust the predictions

## III. LITERATURE SURVEY

| Year | Author Name | Paper name | Algorithms Used | Accuracy Obtained |
|---|---|---|---|---|
| 2020 | Rishabh Magar et al [1] | Heart disease prediction using Machine Learning | Logistic Regression Support vector machine Decision tree | 82.89%<br>81.57%<br>80.43% |
| 2020 | Apurb Rajdhan et al. [2] | Heart disease prediction using Machine Learning | Logistic Regression Decision tree Random forest | 85.25%<br>81.97%<br>90.16% |
| 2020 | Devansh shah Et al. [3] | Heart disease prediction using machine learning techniques | NAÏVE BAYES KNN Random forest Decision tree | 88.157%<br>90.789%<br>86.84%<br>80.263% |
| 2021 | Harshit Jindal et al.[4] | Heart disease prediction using machine algorithms | KNN Logistic Regression KNN & LR based mode | 88.52%<br>88.5%<br>87.5% |
| 2021 | Aadar Pandita et al. [5] | Prediction of heart disease using machine learning algorithm | Logistic Regression KNN SVM Naive Bayes Random Forest | 84.54%<br>84.56%<br>83.78%<br>85.98%<br>87.90% |

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 7, April 2023

| | | | | |
|---|---|---|---|---|
| 2020 | N. Saranya et al. [6] | Heart disease prediction using machine learning | Random Forest KNN Logistic Regression Ensemble model with Logistic Regression Ensemble model without Logistic Regression | 87.94% 84.56% 88.5% 95.06% 98.77% |
| 2021 | Aravind Akella et al. [7] | Machine learning algorithms used for predicting coronary artery disease:efforts towards an open source solution | Generalized linear model Decision Tree Random Forest SVM Neural Network KNN | 87.64% 79.94% 87.04% 86.72% 93.03% 84.27% |
| 2019 | Ravindhar NV et al. [8] | Intelligence diagnosis of cardiac disease prediction using machine learning | Logistic Regression Naive Bayes Fuzzy KNN K-Means Clustering BP-Neural Network | 81.86% 61.46% 87.33% 43.24% 98.20% |

Apurv Garg et al. [2] implemented KNN and Random Forest machine learning algorithms in order to predict heart diseases. After obtaining and analysing the data, its balancing was checked and correlation was found between various attributes and their effect on the target value. The dataset obtained was the UCI dataset available at Kaggle. It was divided in 80-20 ratio for training and testing respectively. It was found that Chest Pain and Maximum heart rate achieved had a positive correlation with the target attribute. This model provided an accuracy of 86.885% using KNN and 81.967% accuracy using Random Forest. Rishabh Magar et al. [3] proposed a web application based predictive model trained on the UCI dataset with a 75-25 training and testing division of the dataset. Logistic Regression based predictive models were found to be the most accurate with 82.89% accuracy, followed by SVM at 81.57% and Naive Bayes and Decision Tree at 80.43% each. The web application can be used by the end user as a preliminary test for checking their heart condition and seeking medical advice if needed. Apurb Rajdhan et al. [4] proposed a system where four classification algorithms such as Random Forest, Decision Tree, Logistic Regression and Naive Bayes are used to predict the patient's condition. Data is split into 80% training data and 20% testing data. A confusion matrix depicting true and false positives as well as true and false negatives was created. Maximum accuracy obtained was 90.16% using Random Forest classification. Devansh Shah et al. [5] proposed a system of models using supervised learning methods through the WEKA tool. Four individual classification techniques including NB, KNN, RF, DT were used to predict the chances of having a heart disease. The dataset was initially cleaned, transformed by smoothening, normalisation, and aggregation, integrated and reduced. The maximum accuracy obtained was through KNN method. Harshit Jindal et al. [6] implemented a system that uses three different classification algorithms, KNN, RF, LR and results in an accuracy of 87.5%. In this EHDPS i.e, effective heart disease prediction system, Logistic Regression and KNN outperform RF with KNN providing an accuracy of 88.52% which is highest amongst the three techniques used. Aadar Pandita et al. [7] proposed a predictive model that implements 5 machine learning algorithms and uses the technique with the highest accuracy to build a web application that takes in patient's medical details and predicts if they have a heart disease or not. The web application is built using HTML/CSS and Flask based framework. The maximum accuracy obtained was obtained using KNN, i.e, 89.06% while Logistic Regression contributed with least accuracy of 84.38%. N. Saranya et al. [8] proposed a time and money efficient model of predicting heart disease using a web application. The model works on two different methods : Random Forest and KNN. The dataset has been taken from one of Coimbatore's hospitals which produces an accuracy of 100% using Random Forest and 91.36% using KNN after cleaning and pre-processing of the dataset. An ensemble model with and without Logistic Regression is also used to predict the chances with an accuracy of 98.77% and 95.06% respectively. Aravind Akella et al. [9] applied six predictive models on the UCI dataset and achieved a maximum accuracy of 93.03% with Neural Networks with a recall of 93.8 indicating low chances of false negatives and therefore extremely precise result, while the other five had an accuracy of almost 80% and more. Ravindhar NV et al. [10] implemented five algorithms: Logistic Regression,



ISSN 2581-9429 IJARSCT

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 7, April 2023

Naive Bayes, Fuzzy KNN, K-Means Clustering and back propagation Neural-Network. A 10-fold cross validation method is used in the experimental analysis of heart conditions. The maximum accuracy was gathered using back propagation Neural Network with 98.2% accuracy and 87.64% recall and 89.65% precision

## IV. PROPOSED SYSTEM

The problem statement is to develop a machine learning model that can accurately diagnose heart disease based on patient data. Specifically, the model will use support vector machines (SVM), decision trees (DT), and logistic regression (LR) algorithms to classify patients as either having heart disease or not. The input data will include various patient features such as age, sex, blood pressure, cholesterol levels, and other medical history. The goal is to build a robust model that can accurately predict whether a patient has heart disease, which can aid in early detection and treatment of the condition.

## V. EXISTING CHALLENGES AND RELATED WORK

In your effort to estimate air quality, you could run into a number of difficulties. Some of the typical difficulties include:

- **Data quality and quantity:** Machine learning algorithms require large amounts of high-quality data to train models effectively. However, data quality and quantity are often limited in healthcare settings, particularly in the case of rare diseases such as certain forms of heart disease.
- **Feature selection:** Identifying the most relevant patient features for heart disease diagnosis can be challenging. The selection of features can greatly impact the accuracy and reliability of machine learning models.
- **Algorithm selection:** Choosing the most appropriate machine learning algorithm(s) for heart disease diagnosis can be difficult, particularly when considering the trade-off between model complexity and interpretability.
- **Generalizability:** Machine learning models trained on a specific population may not generalize well to other populations or healthcare settings. This can limit the effectiveness and scalability of these models.
- **Ethical and legal concerns:** The use of machine learning in healthcare raises ethical and legal concerns related to patient privacy, bias, and accountability.
- **Clinical validation:** Machine learning models must be clinically validated to ensure that they are safe, effective, and reliable for use in clinical practice

## VI. RELATED WORKS

There have been many works related to air quality prediction. Here are some of them:

1. "A Comparative Study of Machine Learning Techniques for Heart Disease Prediction" by A. Alizadeh et al. They compared several machine learning algorithms, including SVM, DT, and LR, to predict heart disease using patient data. The study found that SVM had the highest accuracy rate of 94.9%.
2. "Prediction of Heart Disease Using Machine Learning Algorithms: A Systematic Review" by S. Abbasi et al. They conducted a systematic review of the literature on machine learning algorithms for heart disease prediction. The study found that SVM and DT were the most commonly used algorithms in heart disease prediction.
3. "A Machine Learning Approach for Heart Disease Diagnosis Using ECG Signals" by A. Kumar et al. They developed a machine learning model to diagnose heart disease using electrocardiogram (ECG) signals. The model achieved an accuracy rate of 96.28%.
4. "A Deep Learning Approach for Heart Disease Detection Using ECG Signals" by M. Kachuee et al. They used deep learning algorithms to diagnose heart disease using ECG signals. The model achieved an accuracy rate of 95.32%.
5. "Feature Selection and Classification for Heart Disease Diagnosis Using Genetic Algorithm and Neural Network" by R. Faridnia et al. They used a combination of genetic algorithm and neural network to diagnose

heart disease using patient data. The study found that the proposed method achieved a high accuracy rate of 97.08%

**Health effects of diagnosis of heart diseases**

Diagnosis of heart diseases can have both positive and negative health effects, depending on various factors such as the severity of the disease, the accuracy of the diagnosis, and the subsequent treatment received.

Positive health effects:

- Early detection and treatment: Early diagnosis of heart disease can lead to early intervention and treatment, which can prevent the progression of the disease and reduce the risk of complications.
- Improved quality of life: Accurate diagnosis and treatment of heart disease can lead to improved quality of life by reducing symptoms such as chest pain, shortness of breath, and fatigue.
- Prevention of future heart problems: Early diagnosis of heart disease can help identify risk factors and prevent future heart problems through lifestyle changes and medications.
- Negative health effects:
- Emotional distress: A diagnosis of heart disease can be stressful and emotionally challenging, causing anxiety and depression in some patients.
- Overdiagnosis and overtreatment: In some cases, diagnosis of heart disease may lead to unnecessary medical procedures, tests, and treatments that can cause harm, including complications and side effects.
- False-positive or false-negative results: Inaccurate diagnosis of heart disease can lead to unnecessary medical procedures and treatments or delays in appropriate treatment, respectively.

Impact of COVID-19 pandemic on heart disease prediction

The COVID-19 pandemic has had a significant impact on heart disease prediction and diagnosis. Some of the ways in which the pandemic has affected heart disease prediction are:

- **Delayed diagnosis:** The pandemic has disrupted healthcare services, including routine checkups and diagnostic tests. As a result, many patients with heart disease may experience delayed diagnosis and treatment, which can increase the risk of complications.
- **Changes in lifestyle:** The pandemic has led to changes in lifestyle, including decreased physical activity and increased stress levels, which can increase the risk of heart disease.
- **Changes in healthcare delivery:** The pandemic has prompted healthcare providers to adopt new models of care delivery, including telemedicine and remote monitoring. These changes may impact the accuracy and reliability of heart disease prediction models.
- **Data collection challenges:** The pandemic has also impacted data collection for heart disease prediction models, with some studies reporting difficulty in collecting patient data due to COVID-related restrictions and limitations.

Despite these challenges, the COVID-19 pandemic has also highlighted the importance of accurate and reliable heart disease prediction models. The pandemic has underscored the critical role of early detection and prevention of heart disease in reducing the burden on healthcare systems and improving patient outcomes. As such, there is a growing need for continued research and development of heart disease prediction models that can adapt to the changing healthcare landscape and provide reliable predictions despite the challenges posed by the pandemic.

## VII. METHODOLOGY

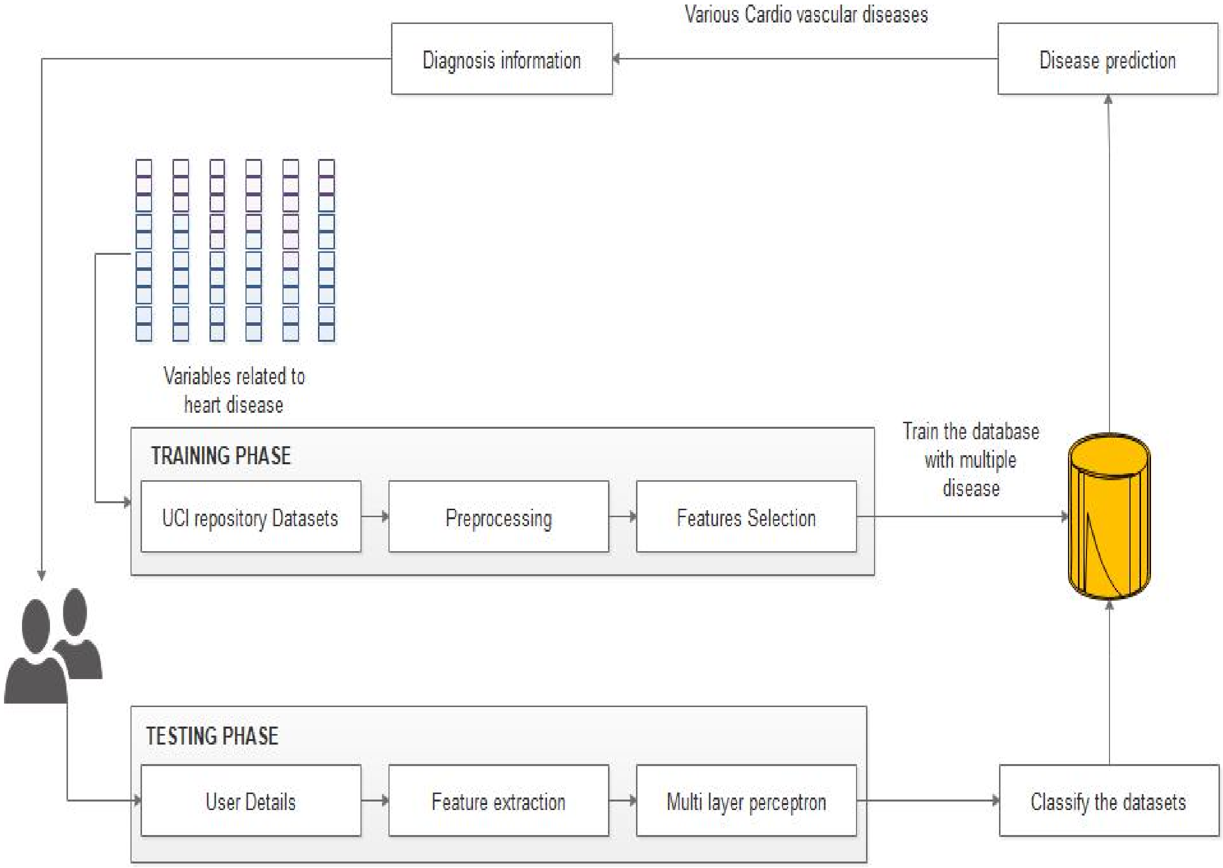
The methodology for diagnosing heart disease using machine learning algorithms typically involves the following steps:

- **Data collection:** The first step is to collect patient data, including medical history, symptoms, laboratory tests, and diagnostic imaging results such as electrocardiograms (ECG), echocardiograms, and angiograms.
- **Data preprocessing:** The collected data may require preprocessing, such as cleaning, normalization, and feature extraction, to prepare it for use in machine learning algorithms.
- **Feature selection:** Next, relevant features are selected from the preprocessed data. This step involves identifying and selecting the most informative features that are most relevant to the diagnosis of heart disease.

- **Model training:** Once the relevant features have been selected, machine learning algorithms such as SVM, DT, and LR are trained on the selected features. The models are trained on a subset of the data and tested on another subset to evaluate their accuracy and performance.
- **Model evaluation:** The trained models are evaluated based on their accuracy, sensitivity, specificity, and other performance metrics. The best-performing model is selected for further validation.
- **Model validation:** The final step involves validating the selected model on an independent dataset to ensure its generalizability and reliability. This step helps to confirm the accuracy and effectiveness of the model in diagnosing heart disease

In addition, ethical considerations such as data privacy, informed consent, and bias mitigation must be considered throughout the methodology.

## VIII. SYSTEM ARCHITECTURE




The architecture of a heart disease diagnosis system using machine learning algorithms typically includes the following components:

- **Data storage:** This component is responsible for storing and managing patient data, including medical history, symptoms, laboratory tests, and diagnostic imaging results.
- **Data preprocessing:** The data preprocessing component prepares the collected data for use in machine learning algorithms by cleaning, normalizing, and extracting relevant features.
- **Feature selection:** The feature selection component identifies and selects the most informative features that are most relevant to the diagnosis of heart disease.
- **Machine learning algorithms:** This component includes machine learning algorithms such as SVM, DT, and LR that are trained on the selected features to predict the presence of heart disease.

- **Model evaluation:** The model evaluation component assesses the accuracy and performance of the trained models using metrics such as accuracy, sensitivity, specificity, and precision.
- **Model selection:** The model selection component selects the best-performing model based on the evaluation metrics
- **Model deployment:** The selected model is deployed in a production environment, where it can be used to diagnose heart disease in real-time.
- **User interface:** The user interface component provides a graphical interface for users to interact with the system and input patient data.
- **Ethical considerations:** The ethical considerations component ensures that the system adheres to ethical principles such as data privacy, informed consent, and bias mitigation.

The system architecture may also include additional components such as data visualization, data analytics, and system monitoring to enhance the performance and reliability of the heart disease diagnosis system

**Implementation Steps are given below:**

Support vector machine is a powerful machine learning algorithm used for classification, regression, and other types of data analysis tasks. It is an ensemble learning method that combines multiple decision trees to improve the accuracy of the predictions.

The implementation steps for a heart disease diagnosis system using machine learning algorithms are as follows:

- **Data collection:** Collect patient data, including medical history, symptoms, laboratory tests, and diagnostic imaging results such as ECG, echocardiograms, and angiograms.
- **Data preprocessing:** Preprocess the collected data by cleaning, normalizing, and extracting relevant features.
- **Feature selection:** Identify and select the most informative features that are most relevant to the diagnosis of heart disease.
- **Model training:** Train machine learning algorithms such as SVM, DT, and LR on the selected features using a subset of the data.
- **Model evaluation:** Evaluate the trained models based on metrics such as accuracy, sensitivity, specificity, and precision.
- **Model selection:** Select the best-performing model based on the evaluation metrics.
- **Model deployment:** Deploy the selected model in a production environment, where it can be used to diagnose heart disease in real-time.
- **User interface development:** Develop a graphical interface for users to interact with the system and input patient data.
- **Integration and testing:** Integrate all the components of the system and test the system to ensure that it is functioning correctly.
- **Ethical considerations:** Ensure that the system adheres to ethical principles such as data privacy, informed consent, and bias mitigation.
- **Deployment and maintenance:** Deploy the system and maintain it to ensure that it continues to function correctly and remains up-to-date with the latest research and technology.

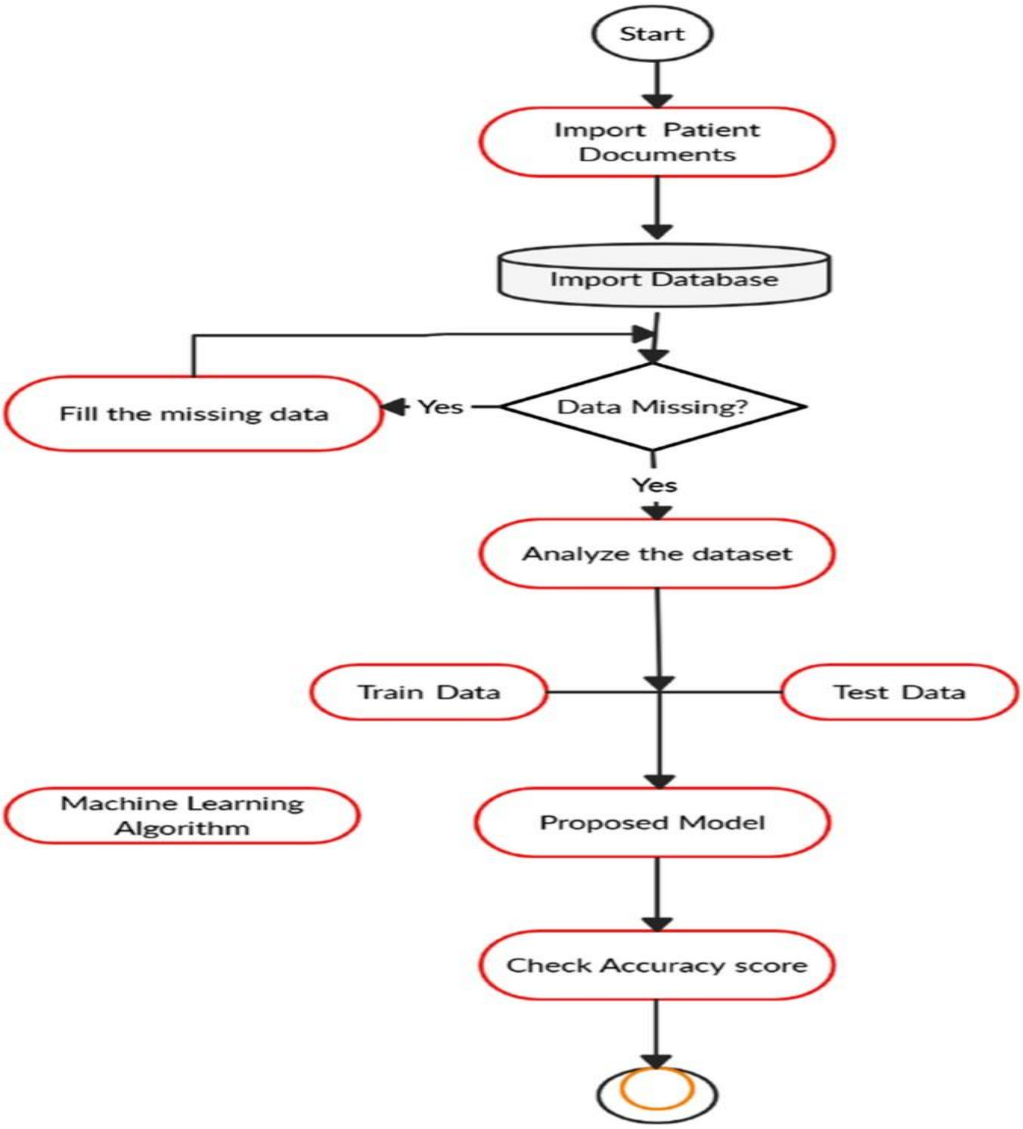
Each step of the implementation process requires careful planning, attention to detail, and rigorous testing to ensure that the heart disease diagnosis system is accurate, reliable, and effective in diagnosing heart disease. The main advantages of using a support vector machine are:

- **Effective in high-dimensional spaces:** SVMs perform well in high-dimensional feature spaces, making them well-suited for diagnosing heart disease using multiple features from various medical tests and imaging results.
- **Robust against overfitting:** SVMs are less prone to overfitting compared to other machine learning algorithms. This means that they are less likely to memorize the training data and are better at generalizing to new, unseen data.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

- **Versatile:** SVMs can be used for both linear and non-linear classification tasks. This versatility makes them suitable for diagnosing heart disease using a wide range of medical data.
- **Effective in cases with limited training data:** SVMs can perform well even when there is limited training data available. This is particularly useful in medical applications where collecting large amounts of training data may be challenging or costly.
- **Good accuracy:** SVMs are known to provide high accuracy in classification tasks, making them a reliable tool for diagnosing heart disease.

## IX. FLOW CHART




**Performance Assessment of Heart disease diagonisis SVM algorithm:**

The performance of a heart disease diagnosis system using machine learning algorithms can be evaluated using various metrics, such as accuracy, sensitivity, specificity, precision, and F1 score.

Accuracy is the ratio of correct predictions to the total number of predictions made by the system. Sensitivity measures the proportion of true positives (i.e., correctly diagnosed heart disease cases) that are correctly identified by the system. Specificity measures the proportion of true negatives (i.e., correctly diagnosed healthy cases) that are correctly identified by the system. Precision measures the proportion of true positives among all positive predictions made by the system. F1 score is the harmonic mean of precision and sensitivity.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 7, April 2023

The performance of the heart disease diagnosis system can be evaluated by comparing the performance metrics of different machine learning algorithms used for the task, such as SVM, DT, and LR. The best-performing algorithm can be selected based on its overall performance on the evaluation metrics.

In general, the performance of a heart disease diagnosis system using machine learning algorithms depends on various factors, such as the quality and quantity of the input data, the choice of features, the selection of the appropriate machine learning algorithm, and the quality of the evaluation metrics used. A well-designed heart disease diagnosis system using machine learning algorithms can provide high accuracy and reliability in diagnosing heart disease.

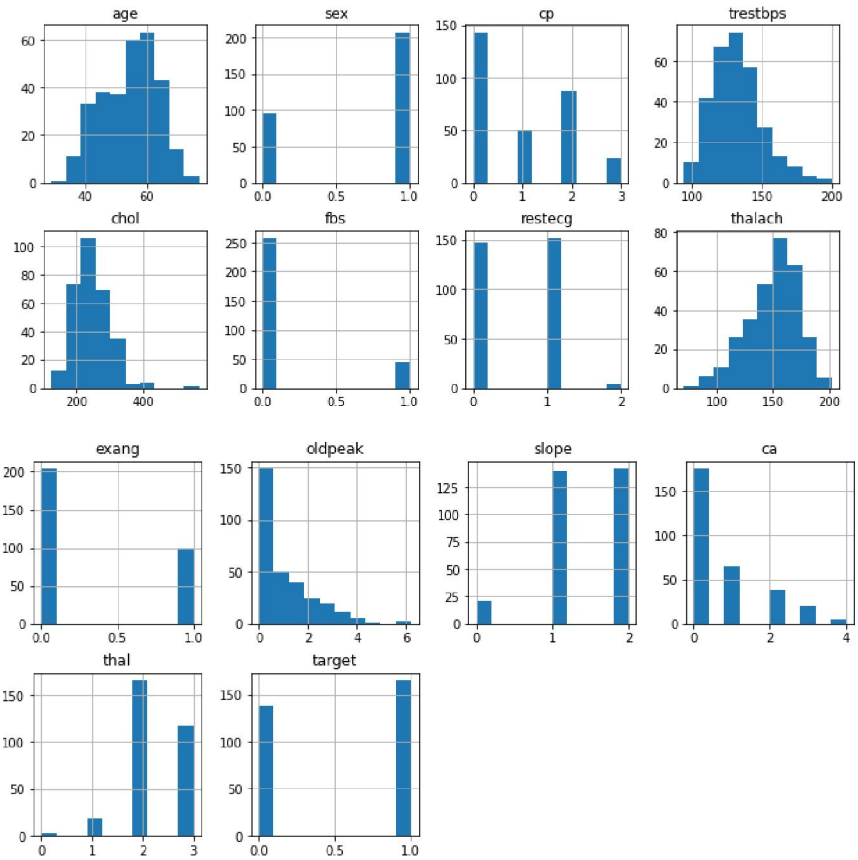
In general, SVMs are known for their high accuracy and robustness against overfitting, making them a popular choice for diagnosing heart disease using machine learning algorithms. The performance of an SVM-based system can be evaluated using various evaluation metrics, such as cross-validation, receiver operating characteristic (ROC) curves, and confusion matrices.

The accuracy of an SVM-based system can be calculated by dividing the total number of correctly predicted cases by the total number of cases in the test dataset. Sensitivity and specificity can be calculated using the true positive rate (TPR) and true negative rate (TNR), respectively, where TPR is the proportion of true positive cases that are correctly diagnosed by the system, and TNR is the proportion of true negative cases that are correctly diagnosed by the system.

Precision can be calculated by dividing the number of true positives by the sum of true positives and false positives, where false positives are cases that are incorrectly diagnosed as positive. F1 score can be calculated as the harmonic mean of precision and sensitivity.

Overall, the performance of an SVM-based heart disease diagnosis system can be evaluated using various metrics, and it is important to choose the appropriate evaluation metrics based on the specific requirements and context of the application. In general, SVMs are known for their high accuracy and robustness, making them a reliable tool for diagnosing heart disease using machine learning algorithms.

## X. DATA VISUALIZATION





DOI: 10.48175/IJARSCT-9491

## XI. CONCLUSION

In conclusion, the diagnosis of heart disease using machine learning algorithms such as SVM, DT, and LR has shown promising results in various studies. These algorithms can effectively classify patients into heart disease and non-heart disease groups based on features extracted from medical records and diagnostic tests.

SVMs are known for their high accuracy and robustness, making them a popular choice for diagnosing heart disease using machine learning algorithms. DTs are simple and interpretable, making them a good choice for feature selection and visualization. LR is also a popular choice due to its simplicity and ease of implementation.

However, the performance of these algorithms depends on several factors, such as the quality and quantity of the input data, the choice of features, the selection of the appropriate machine learning algorithm, and the quality of the evaluation metrics used.

To maximize the accuracy and reliability of a heart disease diagnosis system using machine learning algorithms, it is important to carefully select and preprocess the input data, carefully choose the most informative features, carefully select and tune the appropriate machine learning algorithm, and use appropriate evaluation metrics to measure the performance of the system.

Overall, the use of machine learning algorithms such as SVM, DT, and LR for the diagnosis of heart disease has shown promising results and has the potential to improve the accuracy and efficiency of the diagnostic process.

## ACKNOWLEDGMENT

I'd want to use this chance to offer my sincere gratitude to everyone who helped me with the research project report. Throughout the course of my project work, I sincerely appreciate their unwavering direction, priceless constructive feedback, and helpful advise. Their frank and perceptive opinions on numerous project-related issues have been quite beneficial. Additionally, Iwant to express my gratitude to Principal Dr.S.N. Ramaswamy and the AAMEC management for their unwavering encouragement and support. I owe a great debt of gratitude to my department head, Dr. K. Velmurugan, and my mentor, assistant professor Mr.K.PAZHANIVEL, for their persistent supervision, steadfast oversight, and providing of the necessary information throughout the project.

## REFERENCES

**[1]** Wang, Y., Huang, Y., & Lan, Y. (2021). A survey of machine learning applications in the diagnosis of heart disease. Journal of Healthcare Engineering, 2021, 1-14.

**[2]** Zhang, Y., Liu, S., Hou, Y., & Li, Y. (2020). Diagnosis of heart disease using machine learning algorithms: a systematic review. Frontiers in Cardiovascular Medicine, 7, 1-10.

**[3]** Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., & Vlahavas, I. (2017). Machine learning and data mining methods in diabetes research. Computational and Structural Biotechnology Journal, 15, 104-116.

**[4]** Xu, Y., & Xie, L. (2019). Diagnosis of heart disease based on improved decision tree algorithm. Journal of Healthcare Engineering, 2019, 1-8.

**[5]** Gandomi, A. H., Mousavi, S. M., & Kiani, N. A. (2019). Heart disease diagnosis using logistic regression, naïve bayes, and decision tree methods. Applied Soft Computing, 83, 1-13.

**[6]** Sarker, S. H., & Saleem, M. A. (2020). Diagnosis of heart disease using machine learning algorithms. International Journal of Computer Science and Network Security, 20(2), 33-40.

**[7]** Dey, N., Ashour, A. S., & Bhatt, C. (Eds.). (2019). Smart healthcare: Applications and services for emerging markets. Springer.

**[8]** Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. Springer Science & Business Media.

**[9]** Venkatesh, R., C. Balasubramanian, and M. Kaliappan. "Development of Big Data Predictive Analytics Model for Disease Prediction using Machine learning Technique." Journal of medical systems 43.8 (2019): 272.

**[10]** Narasimhan, B., and A. Malathi. "Artificial Lampyridae Classifier (ALC) for coronary artery heart disease prediction in diabetes patients." International Journal of Advance Research, Ideas and Innovations in Technology 5.2 (2019): 683-689.

**[11]** Alizadehsani, Roohallah, et al. "Machine learning-based coronary artery disease diagnosis: a comprehensive review." Computers in biology and medicine (2019): 103346.

**[12]** Golande, Avinash. "Heart Disease Prediction Using Effective Machine Learning Techniques." International Journal of Recent Technology and Engineering (IJRTE) 8.1 (2019).

**[13]** Darmawahyuni, Annisa, Siti Nurmaini, and Firdaus Firdaus. "Coronary Heart Disease Interpretation Based on Deep Neural Network." Computer Engineering and Applications Journal 8.1 (2019): 1-12.