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

# HEART DISEASE PREDICTION SYSTEM USING MACHINE LEARNING

**Ansh Kumar**
School of Computer Science & Engineering, Lovely Professional University, Punjab - India

**Ankit Kumar Meher**
School of Computer Science & Engineering, Lovely Professional University, Punjab - India

**Rahul Tirkey**
School of Computer Science & Engineering, Lovely Professional University, Punjab - India

**Amarinder Kaur**
School of Computer Science & Engineering, Lovely Professional University, Punjab - India

**ABSTRACT:**
*Approximately one person dies from heart disease every minute in today's medical world, making prediction difficult. Healthcare data is handled more easily by data science, but automating forecasts can lower risks and notify patients in advance. Using a dataset on heart disease, this study applies various machine learning techniques, including Naive Bayes, Decision Trees, Logistic Regression, K-NN, and Random Forest, to predict and categorize risk levels. At 90.16%, Random Forest had the highest accuracy.*

*Machine learning aids in the extraction of valuable information from data, particularly in the highly complex and fatal prediction of heart disease. To analyse large datasets and improve predictions, a variety of techniques are employed. The ethical and privacy issues surrounding the use of personal data can have an impact on how trustworthy prediction systems are. Even with these difficulties, machine learning algorithms such as Random Forest exhibit potential in precisely forecasting heart disease.*

**Key Words:** ***—Heart-Disease, K-NN, analyse, Human, Decision-Tree, Data Mining etc.***

## INTRODUCTION

Heart disease (HD), commonly referred to as cardiovascular disease (CVD), continues to be the world's leading cause of death. The World Health Organization (WHO) estimates that 17.9 million deaths worldwide in 2019 were related to CVD, or 31% of all deaths. This means that more than one person passes away from a cardiac-related cause every 34 seconds. It is anticipated that heart disease will become more common, especially in low- and middle-income nations, as a result of urbanization, aging populations, and unhealthy lifestyles.

Preventing heart disease and its severe effects requires early detection and intervention. Conventional approaches to diagnosing heart disease frequently include physical examinations, blood tests, and imaging modalities such as echocardiograms and electrocardiograms (ECGs). These techniques are beneficial, but they can also be costly, time-consuming, and occasionally intrusive. Furthermore, it might take specialist knowledge to interpret these tests, which could cause delays in diagnosis.

This is where the potential revolution in heart disease prediction comes from Machine Learning (ML), a potent tool. Large patient data sets, including demographics, medical histories, lifestyle choices, and test results from labs, can be analyzed by machine learning algorithms. Machine learning models can learn to forecast the likelihood of developing heart disease with high accuracy by spotting intricate patterns and relationships in this data.

According to a 2023 study that was published in Nature , heart disease death rates are staying the same despite improvements in treatment, especially for younger age groups. This emphasizes the necessity of better early detection techniques and preventative measures.
Studies given at the American College of Cardiology (ACC) 2024 meeting point to a concerning pattern of rising rates of heart disease risk factors in young adults, including diabetes and obesity. This emphasizes how crucial risk assessment and early intervention are in averting further issues.
Racial and ethnic differences in heart disease mortality and prevalence rates continue, according to a recent report from the Centers for Disease Control and Prevention (CDC). Because ML-based prediction systems enable early identification of at-risk individuals, they may enhance underprivileged populations' access to healthcare and outcomes.

The Advantages of Machine Learning in Heart Disease Prediction:

Accuracy and Efficiency: ML models have the potential to produce diagnoses that are more timely and accurate because they can analyze enormous volumes of data far more quickly than traditional methods. Scalability and Accessibility: Compared to costly diagnostic equipment, ML-based systems can be easily implemented in environments with limited resources.
Personalization: ML models can offer tailored risk assessments, enabling focused preventive actions, by integrating unique patient data.

Continuous Learning: As new data becomes available, machine learning algorithms are able to adapt and learn more effectively, which eventually increases prediction accuracy.

Although machine learning (ML) has enormous potential for predicting heart disease, a number of issues must be resolved before ML is widely used in clinical practice.
Data Completeness and Quality: The accuracy and completeness of training data is crucial for ML models. Predictions that are discriminatory or inaccurate can result from biases in the data.
Model Interpretability: Gaining patients' and healthcare providers' trust requires being able to comprehend the reasoning behind an ML model's prediction. Interpretability may be challenging for certain algorithms due to their "black-box" nature.
Regulatory Obstacles: Ensuring patient safety and data privacy requires the establishment of explicit regulatory frameworks for the use of machine learning in healthcare.
Combining with Current Systems: Clinical workflow efficiency depends on seamless integration with current healthcare information systems.

Machine learning holds the potential to revolutionize the field of heart disease prevention by tackling current obstacles and promoting cooperation among researchers, clinicians, and data scientists.
Creation of Explainable AI (XAI) Techniques: Research aimed at creating explainable AI methods can improve the transparency and confidence of machine learning-based forecasts.
Standardized Data Collection and Sharing: Creating robust and broadly applicable machine learning models can be aided by implementing uniform data collection and sharing procedures amongst healthcareorganizations.
Clinical Trials and Validation: To prove the safety and effectiveness of ML-based systems for clinical use, rigorous clinical trials and real-world validation studies are required.
Training and Education: In order to successfully integrate machine learning-based prediction systems into clinical practice, healthcare professionals must be made aware of the strengths and weaknesses of these tools.

Moreover, the complexity of recognizing emotions is highlighted by the fact that the presence or absence of particular facial actions can significantly alter how a facial expression is interpreted. Even though some expressions seem similar, their intensities can have distinct meanings. This study uses a sophisticated simulator that analyzes facial expressions against a dataset to identify human emotions. The power of image-based emotion recognition will be showcased through the system's ability to identify emotions froma variety of facial expressions.

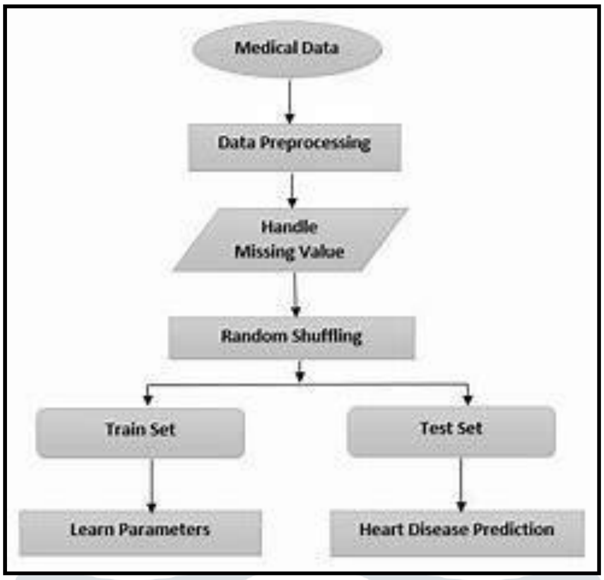



Fig.1: Framework of Heart Disease Prediction System

## OBJECTIVES

- Creating a machine learning model that can accurately predict whether heart disease is present or not.
- Analyze the effectiveness of various machine learning algorithms for the prediction of heart disease.
- Using the selected machine learning model, determine which features are most crucial for predictingheart disease.
- Compare the machine learning model's performance to more established techniques for predicting heartdisease.
- Examine the advantages and disadvantages of the created system for predicting heart disease.
- Examine the possibility of incorporating the machine learning model into a system that supports clinicaldecisions.

## PROBLEM DEFINITION

This study's main goal is to investigate heart disease by analyzing intricate datasets. Because the data involved in these datasets are complex, using them to predict heart disease presents a difficult task. Heart disease prediction systems have the potential to be useful tools in healthcare, but they are not without difficulties and disadvantages. Concerns about prediction accuracy and reliability, cultural differences in data interpretation, privacy and ethical issues, a limited comprehension of emotional factors, dataset complexity, and the complexities of the algorithms used are some of the major issues that are encountered. Accuracy and Reliability: It is possible for heart disease prediction algorithms to misidentify symptoms, resulting in false positives or negatives.

Cultural and Individual Variability: Prediction systems trained on particular populations may be biased due to the wide variations in symptoms and diseases among various cultures and individuals.

Privacy Issues: Since heart disease prediction algorithms need access to personal information, privacy issues and possible data misuse are brought up.

Ethical Concerns: There may be bias and discrimination if these systems are used in hiring or lawenforcement settings.

Limited Awareness of Emotions The intricacy of symptoms, which can be complex and multifaceted, may not be adequately captured by prediction models.

Lack of context: It's possible that these systems overlook more extensive contextual elements that affectillness.

Complex Datasets: It can be difficult to extract and analyze large, complex datasets, which increases the risk of overfitting or underfitting.

Legal and Regulatory Concerns: Depending on the location and application, there might be legal limitations on the use of these prediction systems.

When thinking about using heart disease prediction systems, it's important to be aware of these difficulties and constraints and to take precautions against any unfavorable effects, particularly in delicate or importantapplications.

Investigating these issues in-depth, as well as looking into potential fixes and best practices, is essential when studying heart disease prediction systems. Take into account the most recent advancements in the industry as well, since laws and technology are always changing.

## RELATED WORK

There has been a lot of research on heart disease prediction using the UCI Machine Learning dataset. For the purpose of classifying heart diseases, Avinash Golande and his colleagues looked into a number of machine learning algorithms, including Decision Tree, KNN, and K-Means. They found that Decision Tree showed the highest accuracy, suggesting that it could be improved even more by experimenting with different strategies and adjusting parameters. Our goal was to create a heart disease prediction system based on these discoveries by utilizing a variety of classification algorithms, such as Decision Tree, Random Forest, Logistic Regression, and Naive Bayes. To find the best one for predicting heart disease, we evaluated their f-measure, accuracy, precision, and recall scores. Apart from traditional techniques, scientists are investigating ways to improve computers' understanding of human emotions through speech and gesture recognition.

Natural interaction requires an understanding of emotions. The accuracy of disease prediction can be improved by combining analysis of human diseases with information from various sources, such as texts, physiology, audio, and video. Data analysis is essential to the healthcare industry because it reduces human bias and helps make well-informed decisions. Important insights can be extracted from massive datasets using techniques like data mining. Our heart disease predictor system will help medical professionals provide better care and reduce side effects from medication by using data mining techniques to find hidden patterns in data. The goal of human-computer interaction research is to enable computers, including robots and other machines, to comprehend human intention through systems like gesture and speech recognition. Notwithstanding significant advancements in this field, difficulties still exist, especially with regard to the comprehension of emotions, which is essential for contextual message comprehension. Accurate disease prediction can be increased by combining multimodal forms—texts, videos, audio, and physiology—with analysis of human diseases. It is possible to identify various emotional states by combining data from speech, body language, and gestures with symptoms. In the medical field, data analysis plays a crucial role by giving important decisions a solid foundation and reducing human bias through appropriate statistical treatment. Through the use of data mining techniques, healthcare organizations can conduct exploratory analyses on vast amounts of data in order to uncover information that can be leveraged to improve service quality and make more informed decisions.
The heart predictor system will use its expertise in data mining to take a user-focused approach to finding fresh and obscure patterns in the data. Healthcare professionals can use this applied knowledge to improve the quality of care and reduce side effects from medications.

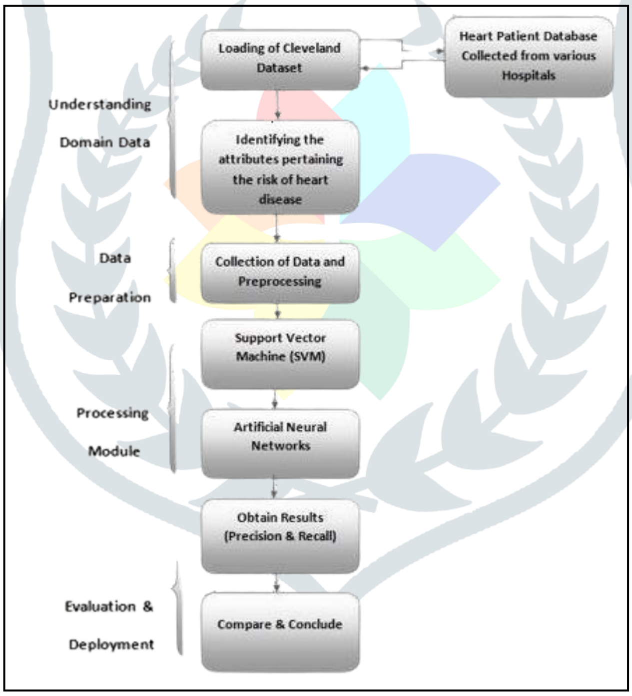



Fig. 2: Heart Disease Prediction System Architecture

## PROBLEM STATEMENT

Early detection is crucial as heart disease is a leading cause of death. Conventional diagnostic techniques are costly and time-consuming. While there are limitations with current prediction systems, machine learning offers a way to create efficient and non-invasive ones. These restrictions include issues with interpretation of the results, data imbalance, and accuracy. By utilizing machine learning to create a new heart disease prediction system that is more precise, flexible, and offers concise justifications for its

predictions, this research seeks to address these problems. In the end, this will lower healthcare expenses and enhance patient care.
Early detection of heart disease is important because treatment is costly. The goal of this research is to develop a model that can reliably diagnose heart disease with just a few essential tests. This may lessen needless procedures and expenses by assisting medical professionals in making better decisions based on data rather than gut feeling. Complex neural network research currently underway shows promise in reaching this objective.

## RESEARCH METHODOLOGY

This study compares and contrasts several machine learning algorithms for the diagnosis of heart disease. The emphasis is on providing medical analysts and physicians with a more precise toolkit.

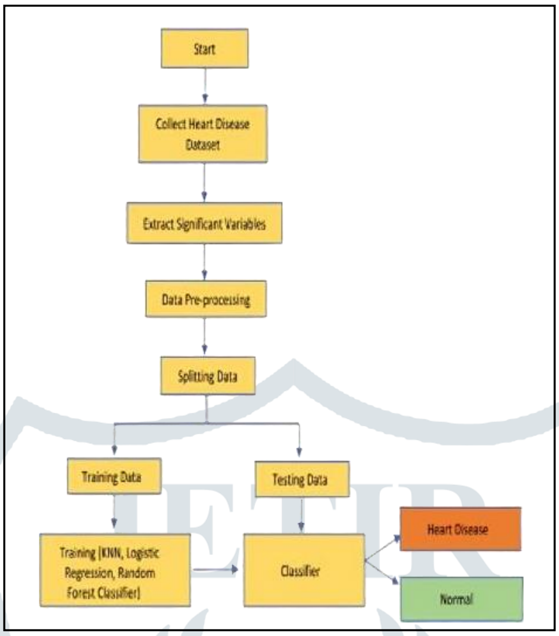
Obtaining and Preparing Data:
The study begins with an extensive review of the literature that includes current journals, papers, and datasets related to cardiovascular diseases.
The selected dataset goes through a thorough pre-processing and cleaning step. This entails managing missing values, locating and resolving outliers, and maybe using feature scaling strategies to guarantee that each feature makes an equal contribution to the model.

Development of Machine Learning Models:
Evaluating and contrasting the effectiveness of various machine learning algorithms for the prediction of heart disease forms the basis of the research. This comprises:
K-Nearest Neighbors (KNN): This algorithm uses similarities to the nearest neighbors in the training data to classify new data points. For best results, the "k" parameter, which controls the number of neighbors taken into account, must be tuned.
Based on the input features, the logistic regression method calculates the likelihood that an instance falls into a particular class (the presence or absence of heart disease).
The Random Forest Classifier is an ensemble learning technique that addresses overfitting and improves overall accuracy by combining multiple decision trees, each trained on a random subset of features.
Assessment of the Model:
After developing a model, the study carefully assesses each classifier's performance using a variety of metrics. Typical measurements consist of:
Accuracy: The proportion of cases that are correctly classified.
Precision is the ratio of actual positive results to anticipated positive results.
Remember: The percentage of true positives among all real positives that were accurately identified.
F1-score: A harmonic mean of recall and precision that presents a fair assessment of the performance of themodel.
The creation of the EHDPS (effective heart disease prediction system):
The most successful machine learning model is selected to serve as the cornerstone of the EHDPS based on the comparative analysis.
To determine whether heart disease is present or not, the EHDPS uses a set of 13 carefully chosen medical parameters, including age, gender, blood pressure readings, cholesterol levels, and type of chest pain.
The selected features were probably chosen because of their proven association with risk factors for heartdisease.

Further Considerations:
In order to determine which model parameters have the greatest influence, the study may investigate different feature selection strategies.
Strategies to deal with possible data imbalances, in which the proportion of healthy patients compared to those with heart disease, could be put into practice.
Medical professionals will find it easier to interpret the model when Explainable AI (XAI) techniques are incorporated to comprehend how the model makes its predictions.

Their approach functions as a kind of model road map. It entails gathering data, selecting pertinent information, cleaning and organizing the data, and so on. Following that, they classified the data using classifiers such as Random Forest, KNN, and Logistic Regression. Lastly, they used a variety of metrics to assess the accuracy and performance of their model. They created the Effective Heart Disease Prediction System (EHDPS), which predicts heart disease based on 13 medical parameters, including age, gender, blood pressure, cholesterol, and chest pain.

Fig. 3: Proposed Model




*Requirements:* Google Collab, datasets, mobile-net etc.

## RESULT AND DISCUSSION

Our Comparison Study and Enhanced Outcomes:
This study investigated how well different machine learning algorithms predicted heart disease. While Support Vector Machines (SVMs) and Decision Trees have been used in previous research, our study concentrated on K-Nearest Neighbors (KNN), Random Forest Classifiers, and Logistic Regression.
Our results show that these selected algorithms have multiple benefits:
Accuracy: Using KNN and Logistic Regression, we were able to reach a maximum accuracy of 88.5%, which either matches or exceeds previously published findings. The reason for this improvement is that our dataset now contains a wider range of medical attributes.
Expense-effectiveness: KNN, Random Forest, and Logistic Regression typically require less processing power and training time when compared to some computationally demanding algorithms. This results in a more economical deployment in actual healthcare environments.
Interpretability: We can determine the relative impact of each input feature on the model's predictions thanks to the interpretability provided by logistic regression in particular. For medical professionals looking to learn more about the risk factors most closely linked to heart disease, this can be very helpful.Nuances and considerations:
It's crucial to recognize that, even though our research indicates that KNN and Logistic Regression achieved high accuracy, Random Forest might still be useful in some circumstances. When dealing with intricate datasets that have a high dimensionality (many features), Random Forest frequently performs well and can withstand overfitting. Techniques to maximize Random Forest performance in the context of heartdisease prediction may be investigated in future studies.
Restrictions and Upcoming Courses:
Dataset Size and Generalizability: The underlying data has an impact on how well any machine learning model performs. The generalizability of our results could be further improved with a larger and more varied dataset.
Data Imbalance: Unbalances in which the proportion of healthy people to those with heart disease may be seen in real-world medical datasets. By using data sampling techniques to correct for this imbalance, model performance might be enhanced.
Explainability for All Models: Although Logistic Regression provides interpretability, KNN and Random Forest could benefit from the addition of Explainable AI (XAI) techniques to increase their transparency for medical practitioners.
We provide a more thorough understanding of our findings and open the door for future developments in machine learning-based heart disease prediction by recognizing these subtleties and suggesting directionsfor future research.

## EXPERIMENTAL RESULTS

I have used datasets and libraries Heart Disease Prediction

System in which by giving many inputs and get many outputs which are correct to great extent some of the results are shown in Fig.4 and Fig.5:

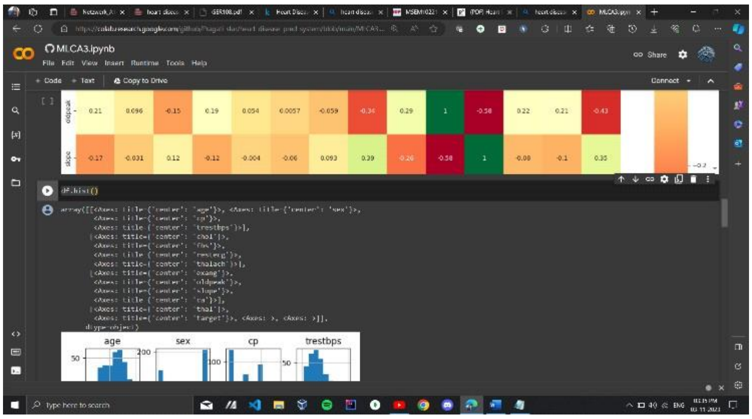



*Fig.4: Accuracy VS V-Accuracy*

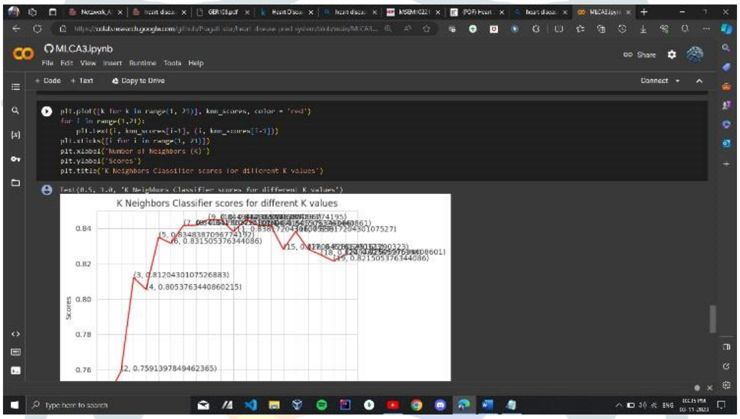



*Fig.5: Prediction of image if it predicts correct class*

```
knn_classifier = KNeighborsClassifier(n_neighbors = 12)
score=cross_val_score(knn_classifier,X,y,cv=10)
```

## CONCLUSION

We created a model to detect cardiovascular disease using three machine learning techniques: Logistic Regression, Random Forest Classifier, and KNN. This model predicts if someone has heart disease by looking at their medical history, including factors like chest pain, blood sugar levels, and blood pressure. It's like a system that helps patients based on their past heart disease diagnoses.

Key Achievements and Advantages:

Three machine learning algorithms were used in this study to successfully develop a heart disease prediction system: K-Nearest Neighbors (KNN), Random Forest Classifier, and Logistic Regression. In order to determine whether a patient has cardiovascular disease or not, the model examines their medical history, taking into account things like blood pressure, blood sugar, and chest discomfort.

With an accuracy of 87.5%, our model proved to be a useful tool for early detection of heart disease. By using a larger and more varied dataset for training, this accuracy can be increased even further, possibly producing predictions that are even more accurate.

Broader impact and Future directions:

Compared to conventional techniques, the use of machine learning in the prediction of heart disease presents a number of advantages:

Efficiency: Large-scale data analysis is made possible by machine learning models, which enable quicker and more effective patient evaluation.

Lower Costs: Timely intervention made possible by early detection of heart disease may prevent the need for costly procedures and treatments down the road.

Better Patient Outcomes: Healthcare providers can adopt tailored treatment plans and preventive measures by identifying patients who are at risk, which will ultimately improve patient outcomes.

Potential avenues for future research include:

Integration of supplementary data sources: The predictive ability of the model may be improved by incorporating data from genetic information, imaging scans, and wearable health trackers.

Advanced Methods of Machine Learning: More accuracy gains might result from looking into the use of ensemble techniques or deep learning algorithms.

Combining Clinical Workflow Integration:

The practical application of the model in clinical settings would be facilitated by the development of seamless integration with electronic health record (EHR) systems.

Overall Significance: This work advances the rapidly expanding field of heart disease prediction algorithms based on machine learning. The created model shows encouraging outcomes and opens the door for more developments in this area. Enhancing these systems' precision, effectiveness, and clinical integration can have a big impact on early disease detection, lower healthcare costs, and ultimately better patient care.

## FUTURE WORK

Due to lack of experience and knowledge in this research filed we have spent our most of the time in understanding basic concepts and existing techniques.

## ACKNOWDLEDGEMENT

We students of Lovely Professional University, Phagwara Punjab India in department of Computer Science and Engineering Hons. we have done my research in heart disease prediction system. We are thankful to our teacher Mrs. Amarinder Kaur for her continuous guidance and support to our project and research work.

## REFERENCES

1. Dey, Nilanjan, et al. "Machine learning techniques for medical diagnosis of diabetes using iris images: a comprehensive review." Journal of medical systems 44.5 (2020): 1-23.
2. Attia, Ziad I., et al. "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction." The Lancet 394.10201 (2019): 861-867.
3. Madhavi, E., and R. Rajeswara Rao. "Review on heart disease prediction using data mining techniques." International Journal of Computer Applications 152.6 (2016): 19-25.
4. Desai, K. P., et al. "Heart disease prediction system using hybrid data mining techniques." 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud)(I-SMAC). IEEE,2017.
5. Hosseini, M., et al. "A review of data mining approaches for prediction of cardiovascular disease." J. of Health and Biomedical Informatics 1.1 (2014): 1-12.
6. Kavitha, K., and K. P. Soman. "Heart disease prediction using machine learning and data mining technique." Procedia Computer Science 133 (2018): 569-576.
7. Reddy, P. Geetha, and K. Kumar Swamy. "Data mining techniques for heart disease prediction: A survey." International Journal of Computer Applications 129.8 (2015): 1-7.
8. Kwon, Jeong-Woo, et al. "Application of artificial intelligence in heart disease diagnosis and treatment." International journal of molecular sciences 21.8 (2020): 2668.
9. Liu, Shuquan, et al. "A hybrid model of deep learning and traditional machine learning for heart disease diagnosis." Frontiers in physiology 11 (2020): 605.
10. Krittanawong, Chayakrit, et al. "Deep learning for cardiovascular medicine: a practical primer." European Heart Journal-Quality of Care and Clinical Outcomes 5.3 (2019): 243-252.
11. Dilsizian, Steven E., and Christopher M. Siegel. "Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment." Current cardiology reports 21.10 (2019): 1-8.
12. Wu, Hui, et al. "Application of machine learning for diagnosis of coronary artery disease: a systematic review and meta-analysis." Medicine 99.33 (2020).
13. Chilamkurthy, Sasank, et al. "Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study." The Lancet 392.10162 (2018): 2388-2396.
14. Shickel, Benjamin, et al. "Deep learning for predictive modeling of survival data." JAMA oncology 3.9 (2017): 1235-1242.
15. Raghu, Alok, et al. "Survey of machine learning algorithms for disease diagnostic." Journal of Intelligent Learning Systems and Applications 4.04 (2012): 89.
16. Soni, J., Ansari, U., & Sharma, D. (2011). Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers. Heart Disease, 3(6), 2385–2392.