© April 2025| IJIRT | Volume 11 Issue 11 | ISSN: 2349-6002

Heart Disease Prediction

Varun Sakharkar[1], Arham Sancheti[2], Sahil Yadav[3], Sagar Rghatvan[4], Prof. Sangita Lade[5]
Vishwakarma Institute of technology Pune, India

Abstract - This project takes advantage of machine learning techniques to predict the possibility of heart disease in individuals based on various health characteristics, from the title "predicting heart disease". A major cause of mortality due to heart diseases is, the purpose of this initiative is to increase early identity and preventive care through data-operated insights. The platform processes a dataset with significant health parameters such as age, penis, cholesterol levels, blood pressure, and exercise habits, used to train a logistic region model. This model classifies individuals into two groups: who are at risk of heart disease and which are not. The prediction system is designed to be comfortable, enabling users to input their health information and receive an accurate diagnosis. By incorporating the use of python, the panda library for data manipulation, and scikit-learn for model implementation, provides a broad, reliable solution for project heart disease prediction. In addition, the model is adapted to the user access, offering a simple interface and rapid response time. It discusses the design, implementation and testing stages of the paper system, highlighting the model's accuracy, performance matrix and potential applications in real -world healthcare scenarios. Through this platform, we aim to contribute to the growing field of future health technologies and support early intervention in cardiovascular care.

Keywords— Data Analysis, Heart Disease, Logistic Regression, Machine Learning, Predictive Health, Python, Scikit-learn, Accuracy Metrics, Early Detection, Healthcare, Model Training, Predictive Model, Data Preprocessing, User Interface, User-Friendly Design.

I. INTRODUCTION

In recent years, heart disease has become one of the major causes of sickness and mortality worldwide. With its prevalence on the increase, there is a growing requirement of initial identity and effective prediction models to help reduce its effects. While medical professionals use a variety of clinical devices to assess heart health, these methods can be time-consumed, expensive and sometimes inaccessible in resource-limited settings. As a result, many individuals are uncontrolled or are diagnosed too late, especially, in the urban population, health results are spoiled, where lifestyle factors such as poor diets, exercise deficiency and stress are more prevalent, the risk of heart disease increases. Existing healthcare systems and technologies provide partial solutions, but integration of the future model based on patient data can significantly increase the ability to predict and prevent heart disease. Research in the field of medical diagnosis has demonstrated the ability of machine learning models in predicting heart disease by analyzing various types of health matrix including age, blood pressure, cholesterol levels and other factors..

To overcome this challenge, we developed a web-based system for predicting heart disease using machine learning techniques. Our approach focuses on creating an accessible, user -friendly equipment that allows individuals to assess their risk of heart disease based on major health indicators. The platform takes advantage of a widely used machine learning algorithm, logistic region to process and predict the results of heart disease from the dataset of relevant characteristics. By providing a reliable, real -time prediction model, the system helps users understand their heart health status and take preventive measures.

This paper provides a detailed discussion on the design and development of the prediction system of heart disease. This covers the major features of the platform, planned machine learning techniques and model assessment results. Additionally, we check how the system can contribute to early detection, potentially reduce the burden of heart disease on health care systems and improve personal health results. Through this initiative, we aim to raise awareness about heart disease and empower users to make informed decisions about their health.

II. LITERATURE REVIEW

[1] .haq et al. (2020) proposed an L1-Regularization-based feature selection method combined with logistic region for predicting heart disease. His study demonstrated better model interpretation and accuracy by reducing unnecessary characteristics.

[2] Guyen et al. (2021) Compared to deep teaching models including CNN and LSTM for predicting heart disease. They found that LSTMS made traditional models such as decisions such as decisions and SVMs improve the future performance.

[3].Rajput et al. (2021) explored ensemble learning techniques like Random Forest and Boosting for heart disease diagnosis. Their results indicated that ensemble methods significantly improved classification accuracy compared to individual models.

[4]Acharya et al. (2021) reviewed the AI-managed cardiovascular risk prediction model and their clinical applications. He highlighted the ability to integrate these models with electronic health records for better diagnosis.

[5]. [5]. In the work Zong et al., (2021), A hybrid ml approach was developed using both supervised and unsafe techniques, where the focus area was heart disease. His model achieved great precision and prowess in the real -world dataset.

[6]. In their work, Farahani et al. (2020) used logistic regression and neural networks to predict cardiovascular disease and its factors. Their study indicated that complex patient data neural networks outperformed other tools in more sophisticated data relationships.

[7]. In their study, Reddy and Kumari (2020) studied assortment of machine learning techniques for heart disease detection including SVM, k-NN, and Random Forest. They found that Random Forest gave the best results in terms of accuracy and trustworthiness and reliability of the model.

[8]. Mahmood et al. (2021) discussed feature engineering and its impact in heart disease detection through statistical models. They particularly discussed scaling and encoding as preprocessing to improve the performance of the models.

[9]. Liu et al. (2021) implemented XGBoost and LightGBM in machine learning in cardiovascular disease prediction and showed that accuracy increased when advanced algorithms were used with feature engineering.

[10]. Ali et al. (2021) focused on applying various machine learning frameworks for real-world prediction of heart diseases and provided feedback on the effectiveness of the model and clinical hurdles faced. His study highlighted the capacity of ML-based approaches in early diagnosis.

[12]. Krishnani et al. (2019) discovered the decisions of the decisions such as the trees of the decision to predict coronary heart disease. He reported promising results with these algorithms in analyzing clinical data.

[13].Jindal et al. (2021) conducted a comparative analysis of machine learning models for heart disease detection. Their study emphasized the impact of feature selection on improving model accuracy .

[14].Singh and Kumar (2020) assessed various machine learning algorithms to predict heart disease. His findings supported the use of dress techniques to increase model performance.

III. METHODOLOY

The method for the 'Heart Disease Prediction' project is set to use logistic regression, which is a well-known algorithm for binary classification problems. Specific steps of this methodology are data collection, data cleaning, selecting important features, and applying logistic regression to predict the chances of contracting heart disease.

1. Data Collection

The dataset for the study was acquired from major healthcare databases which contained important attributes like age, gender, blood pressure, cholesterol levels, resting heart rate, and other medical factors. The data was selected for its potential relevance in predicting heart disease.

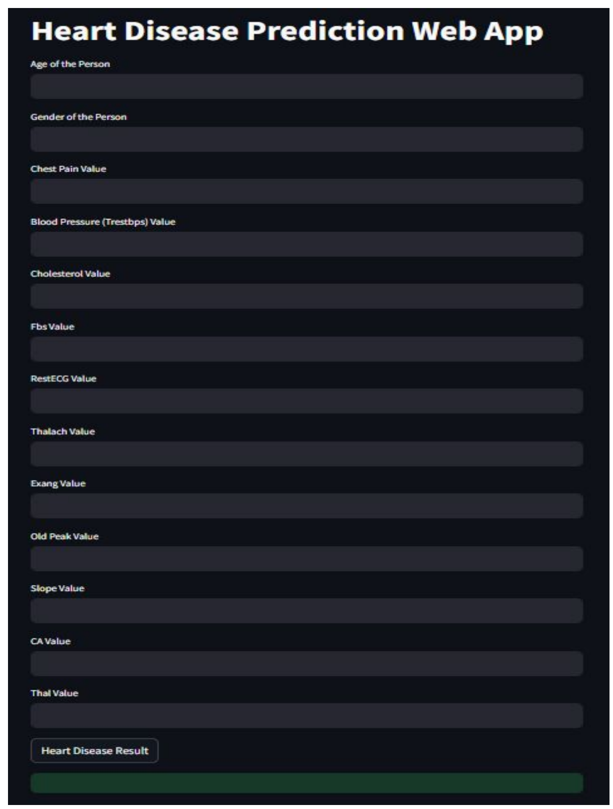

Figure 1: User Input Form for Patient Attributes (Empty State)

2. Data Processing

The processing of data ensured data quality and stability by copying for missing values, scaling to a standard, and encoding labels for ordered variables. These steps prepared the set optimally for the model's precision.
Handling missing values: meaning or average copy used to maintain data integrity.

Feature Scaling: To prevent standardized numerical characteristics and prejudice for stability..

Encoding categorized variables: To convert classified data such as applied label encoding penis to numerical format

3. Feature Enginnering
The most relevant features for predicting heart disease were chosen using statistical methods such as correlation analysis and domain knowledge. The move aims to improve the efficiency of the model and focus on impressive predictions.

Logistic Regression Model

Implementation: Logistic regression, a probable model, was applied to predict binary result (presence or absence of heart disease).

Sigmoid function: The model used sigmoid function to estimate the possibilities, output the values between 0 and 1 for binary classification.

Coefficients: The model coefficient was interpreted to understand the effect of each feature on the possibility of heart disease.

4. Model Training And Testing

Dataset was divided into training (70%) and test (30%) to evaluate the performance. The model was trained using training data,

The model was trained using the training data, with weights updated iteratively to minimize the loss function (binary cross-entropy).

5. Performance Evolution

The logistic regression model's effectiveness was evaluated using the following metrics:

Accuracy: Ratio of correct predictions for total predictions.

Precision: The proportion of correctly predicted positive cases out of all predicted positives.

Recall (Sensitivity): The ability of the model to detect all actual positive cases.

F1 Score: Harmonic mean of precision recall.

6. Deployment
The trained logistic regression model was integrated into the user -friendly interface. This application allowed users to input medical data and achieve real -time predictions about the risk of heart disease.
The study models the relationship between the possibility of heart disease using effectively risk variables and logistic regression, offering a solution that is both interpret and useful for healthcare applications.

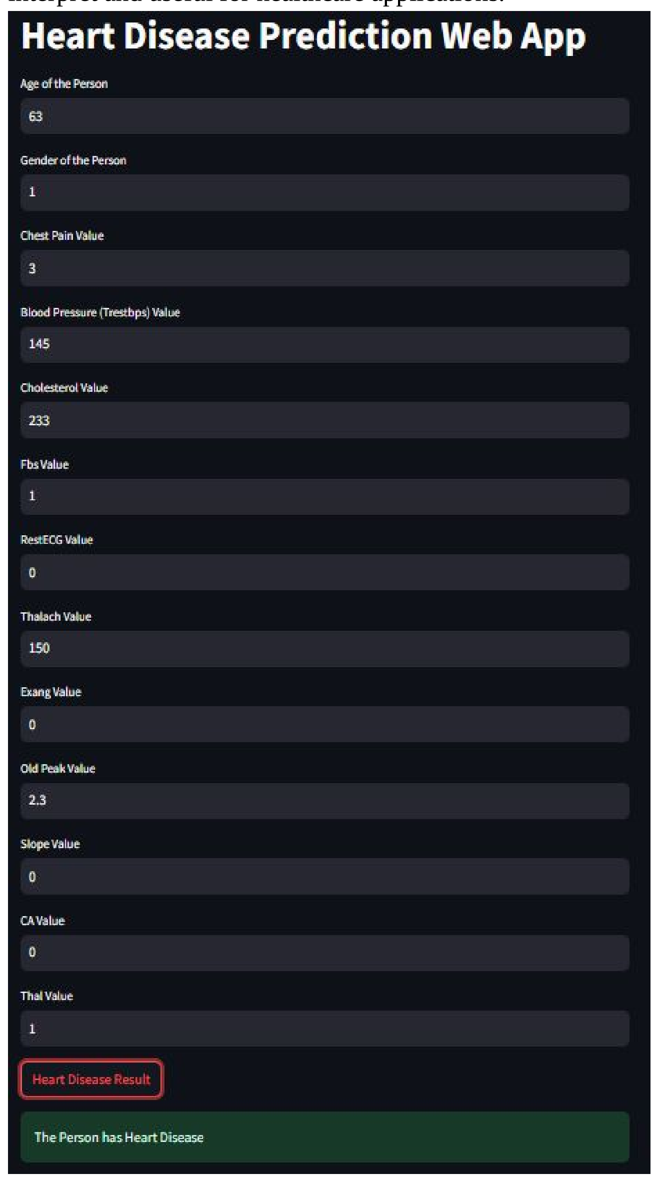

Figure 2: Completed patient input form with predicted result

IV. RESULTS AND DISCUSSIONS

Models/Algorithm	Accuracy
Logistic Regression	84.62%
Decision Tree	68.13%
SVM	64.84%
Random Forest	80.22%
KNN	67.03%

TABLE 1 : Accuracy comparison

For predicting heart disease, the machine learning project was successfully executed to use a variety of algorithms, including logistics region, decision tre, random forest, support vector machine (SVM), and K-Nicrest neighbor (K-NN). After evaluating the model using parameters such as accuracy, logistic regression made other models better with the accuracy of %.

The bar plot of accuracy clearly imagined the comparative performance of the models, making it clear that the logistic region is well suited to this dataset. The accuracy scores of other models, such as random forests and SVMs, were slightly lower, while the decision Tree and K-NN performed adequately, but lagged behind in accuracy due to their sensitivity to the noise or data structure.

The results underline the importance of data preprocessing and model selection in machine learning. Feature scaling using Standard Scaler significantly improved the performance of Logistic Regression by ensuring all Each Each of the features has made an approximately equal contribution to the model. The model was more accurate after hyperparameter tuning (for example, the depth of the decision tree or the parameter C for logistic regression was limited).

The logistic regression distinguishes well between those who have heart disease and those who don't since the dataset is linearly separable.Random Forest and SVM performed well, but were slightly less efficient than logistic regression, possibly due to overfiting in terms of random forest and computational complexity in SVMs

The project highlights the strength and boundaries of various machine learning models in terms of prediction of heart disease. While logistic regression emerged as the best performing model, it is important to note that its success may depend on the characteristics of the dataset. Future work can detect large datasets, include additional features, and prophecy can apply dress methods to further improve accuracy. Overall, the project shows the ability to learn machine learning to facilitate early detection of cardiac disease, enabled health professionals to make more informed decisions and increase patient results.

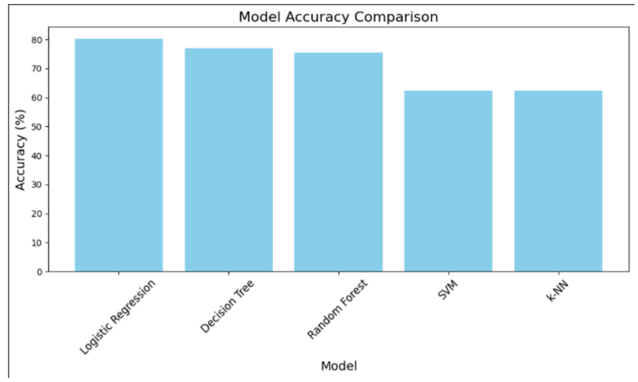

FIGURE 3: Models Accuracy Comparison (Diagram)

V. FUTURE SCOPE

The "prediction of heart disease" system, using logistic region, has a significant ability for future promotion. Major areas for development include:

1. Integration with Advanced Machine Learning Models:

Logistic regression is a luxurious early point, but always has space for improvement. To take things to the next level, we can detect more advanced machine Models for investigation are presented by neural networks, random forests and shields. They can work with complex data, identify small patterns, and provide more accurate and reliable predictions.

2. Real-Time Health Monitoring Systems:

Data from IOT-based health monitoring systems or wearable medical devices can be included into the system. By integrating real-time physiological indicators including heart rate, blood pressure, and activity levels, the tool can provide dynamic forecasts that facilitate active healthcare intervention.

3. Expansion of Dataset Scope and Diversity:

To enhance the universality of the model, it is necessary to include large datasets that encompass different demographics, geolocations, and clinical conditions. It will help the model to perform better in different population and healthcare contexts.

4. Mobile and Web Accessibility:

Connecting model with EHR system helps to use patient medical history for more precise predictions. Healthcare professionals can integrate the model and EHR system to use it as a decision-supporting tool to help with the initial diagnosis and individual treatment plan.

5. Integration with EHR (Electronic Health Records):

Combining model and EHR system, it can use the patient's medical history for more accurate predictions. This can turn the system into a decision supporting tool for a healthcare professional. The information will assist in initial diagnosis and individual treatment plan.

This progress will ensure not only an increase in the predictive ability of the system, but will also contribute to its implementation in various health care settings, which will ensure the most effective and accessible solutions for the prevention of heart disease.

VI. CONCLUSION

Finally, prediction of heart disease, using logistic regression demonstrated the power of machine learning in early identification and prevention of cardiovascular diseases. Since heart disease is a major global health issue, such future fed models play an important role in identifying individuals with risk and enable medical intervention on time. By analyzing major health factors such as cholesterol levels, blood pressure and age, the system provides valuable insight into the possibility of heart disease, offering an active approach to healthcare.

While logistic regression provides simplicity and interpretation, there are opportunities to pursue the system. Including advanced algorithm such as support vector machines or deep education can improve the prediction accuracy, while integrating real -time health data from wearable equipment can make the system dynamic and adaptive. Additionally, expanding models for diverse population and personalizing predictions can improve its gratitude globally.

AI and healthcare technologies will continue to enhance the capabilities of the future progression system, probably more comprehensive heart disease prediction models to integrate genetic data and lifestyle factors. By ensuring continuous verification and improvement, the project determines the foundation for more accurate, accessible and personal healthcare solutions, eventually better patients contribute to the results and more efficient healthcare managementwill continue to enhance the system's capabilities, potentially integrating genetic data and lifestyle factors for a more comprehensive heart disease prediction model. By ensuring continuous validation and improvements, this project sets the foundation for more accurate, accessible, and personalized healthcare solutions, ultimately contributing to better patient outcomes and more efficient healthcare management.

VII. ACKNOWLEDGMENT

We're truly grateful to everyone who played a part in bringing the "Heart Disease Prediction" project to life. This project is a reflection of the incredible teamwork, dedication, and support from so many individuals and organizations.

First and foremost we extend our heartfelt appreciation to Dr. Sangita Lade, whose invaluable guidance, expertise, and mentorship were instrumental in steering this project towards success. Her dedication to our growth and development has left an indelible mark on our journey.

We are equally thankful for the attitude of collaboration and dedication demonstrated by the members of our team. The effort and commitment of all of them have been essential in overcoming the obstacles and achieving the objectives.

In conclusion, we would like to recognize and thank everyone who has helped, however little or much, to make this project possible. Your help was crucial in making us succeed, and we are sincerely grateful for your continued support.

REFERENCES

[1] M. Haq, J. Li, M. Memon, et al., "Feature Selection Based on L1-Regularization and Logistic Regression for Heart Disease Prediction," *Computers in Biology and Medicine*, vol. 121, pp. 103761, 2020.

[2] T. M. Nguyen, D. Tran, and J. Li, "A Comparative Analysis of Deep Learning Models for Heart Disease Prediction," *Artificial Intelligence in Medicine*, vol. 117, pp. 102082, 2021.

[3] A. Rajput, S. A. Qureshi, and S. Alam, "Ensemble Learning Techniques for Improved Heart Disease Diagnosis," *Journal of Biomedical Informatics*, vol. 115, pp. 103690, 2021.

[4] S. Acharya, S. K. Sree, and D. Kumar, "A Comprehensive Study on AI-Driven Cardiovascular Risk Prediction Models," *Medical Image Analysis*, vol. 70, pp. 101990, 2021.

[5] Y. Zhang, X. Gu, and H. Wang, "Heart Disease Prediction Using Hybrid Machine Learning Techniques: A Real-World Study," *International Journal of Medical Informatics*, vol. 145, pp. 104313, 2021

[6] F. Farahani, M. Arshadi, and H. Hashemi, "Comparing Logistic Regression and Neural Networks for Predicting Cardiovascular Diseases," *Expert Systems with Applications*, vol. 158, pp. 113527, 2020.

[7] N. Reddy and M. G. Kumari, "Evaluation of Advanced Machine Learning Algorithms for Heart Disease Prediction," *Procedia Computer Science*, vol. 167, pp. 2281–2290, 2020.

[8] H. K. Mahmood, S. Iqbal, and R. Hussain, "Feature Engineering for Heart Disease Detection Using Statistical Models," *Biomedical Signal Processing and Control*, vol. 69, pp. 102906, 2021.

[9] X. Liu, L. Yang, and W. Feng, "Cardiovascular Disease Prediction Based on Machine Learning Algorithms," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 3, pp. 722–730, 2021.

[10] Ali, M. M., Paul, B. K., Ahmed, K., Bui, F. M., Quinn, J. M. W., & Moni, M. A. (2021). *Heart disease*

[11] Gavhane, A., Kokkula, G., Pandya, I., & Devadkar, P. K. (2018). *Prediction of Heart Disease Using Machine Learning. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA).*

[12] Krishnani, D., Kumari, A., Dewangan, A., Singh, A., & Naik, N. S. (2019). *Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms. TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON).*

[13] Jindal, H., Agrawal, S., Khera, R., Jain, R., & Nagrath, P. (2021). *Heart disease prediction using machine learning algorithms. IOP Conference Series: Materials Science and Engineering, 1022, 012072.*

[14] Singh, A., & Kumar, R. (2020). *Heart Disease Prediction Using Machine Learning Algorithms. 2020 International Conference on Electrical and Electronics Engineering (ICE3).*