OAIJSE

OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

# Heart Disease Prediction using Machine Learning and Deep Learning

**Prof. S. V. Phulari[1], Shravani Pangare[2] Prasad Sutar[3],Harshawardhan Thorat[4], Shubham Waghale[5]**

*Department of Computer Engineering Pune District Education Association's College of Engineering, Savitribai Phule Pune University, Pune, India[1,2,3,4,5]Associate Professor, Department of d Computer Engineering Pune District Education Association's College of Engineering, Savitribai Phule Pune University,Pune[6]shivaphulari6666@gmail.com[1],spangare44@gmail.com,prasadsutar2055@gmail.com,harshwardhanthorat811@gmail.com,swaghale5@gmail.com*

---

***Abstract: Heart disease is the most considerable cause of death worldwide, and so there is a need for efficient prediction techniques. This paper presents the application of ML and DL techniques to predict heart disease. For this purpose, various models such as Logistic Regression, Random Forest, Support Vector Machines, and Neural Network were trained and evaluated with an input data set consisting of 1,025 records. Normalization and encoding were applied as processing steps to enhance performance. Results indicate that the Random Forest model had the highest accuracy at 92%, followed by Deep Neural Network at 91%. This study presents that ML and DL are potentially used in aiding early diagnoses; healthcare issues can be better addressed.***

***Keywords: Heart Disease, Machine Learning, Deep Learning, Prediction, Healthcare.***

---

## I. INTRODUCTION

Heart disease is one of the major global health issues and one of the significant causes of deaths annually. Heart disease can be diagnosed early, and its accuracy will allow the right medical interventions to be performed in time and, therefore, improve patient outcomes. In this regard, it is now possible to analyze complex medical data by means of advanced machine learning and deep learning algorithms for more precise prediction of heart conditions.

This is a project on the prediction of heart disease based on a hybrid approach of ML and DL algorithms. The system works on a data set with 14 significant input attributes: age, sex, chest pain type (cp), thalassemia (THAL), thalach (maximum heart rate achieved), cholesterol (CHOL), fasting blood sugar (FBS), resting blood pressure (TRESTBPS), exercise-induced angina (EXANG), and other relevant medical parameters. These features are fed into various models as an input and tested and validated for forecasting whether a patient has a chance of heart disease or not.

The study uses a machine learning approach including SVM, Random Forest, and Logistic Regression, along with deep learning techniques such as ANN and CNN. Trained and tested models identify patterns and interaction among attributes indicative of the risk of heart disease, by making use of the processed data. Techniques like feature selection and optimization enhance model performance, and the best algorithm can be selected after doing a comparative analysis.

The system is designed to be user-friendly and accessible through the development of a graphical user interface where users input medical parameters, and receive real-time predictions about their heart health. This project will help bridge the gap between medical data analytical and practical implementation by providing a reliable and scalable tool for assisting healthcare professionals and individuals in identifying heart disease risks early.

This research shows how advanced computational techniques can be applied to overcome critical health challenges and improve predictive healthcare systems by combining traditional ML models with modern DL approaches.

By integrating traditional ML models with modern DL approaches, this research demonstrates how advanced computational techniques can be used to overcome critical health challenges and improve predictive healthcare systems.

## II.LITERATURE REVIEW

Karegowda et al. (2012) demonstrated SVM applications toward classification in heart disease datasets. It can handle high dimension and define hyper planes for optimum classification. Thus, SVM is applicable to high-dimensional spaces, which makes it potential for binary classification tasks, such as predicting heart disease.Chen et al. (2018) proved that noisy or incomplete medical data can be learned well by Random Forest. Its ensemble nature enhances its accuracy and robustness in predictions.Kumar et al. (2016) used logistic regression as it is easy to understand relationships between features. It has still remained a default baseline model for many binary classification problems.

It is also mentioned that Mohammed et al. (2019) applied an ANN model to predict heart diseases and outperformed the conventional ML models. ANN' s, because of their multi-layered architecture, tend to be an ideal tool to model non-linear relationships.

Recently, Abdullah et al. (2021) have introduced CNN to the tabular heart disease data set and analyzed those using 1D convolutional layers. These layers assist in picking up the interactions and patterns among the features, thus providing better predictability.

Gupta et al. (2020) describe a comparison study of SVM, Random Forest, and Logistic Regression. The authors conclude that ensemble methods such as Random Forest tend to outperform individual algorithms since ensemble algorithms reduce over-fitting and bias.

## III.METHODOLOGY

**1) Data Gathering Data Source:** Collect a data set with the patient's details about heart disease, like age, gender, cholesterol levels, blood pressure, ECG result, and other clinical attributes. Popular datasets available are the Cleveland Heart Disease data set from UCI or Kaggle.

### 2) Data Preprocessing

- **Cleaning:** Remove missing or incomplete data through imputation or removal of rows/columns.
- **Normalization:** Scale features to have equal distribution, particularly for algorithms that are sensitive to the magnitude of the feature (e.g., SVM, ANN).
- **Encoding:** Translate categorical variables into numerical ones, such as one-hot encoding for categorical variables like "thalassemia" or "sex".
- **Data Splitting:** Split the data into training, validation, and test sets. Usually, this is 70% training, 15% validation, and 15% test.

### 3) Feature Selection

- **Feature Importance:** Use correlation analysis or tree-based models, like Random Forest to identify the most crucial features that account for the major contribution to predicting heart disease.
- **Dimension Reduction:** If necessary, PCA (Principal Component Analysis) would be used, reducing the feature space without losing key information.

### 4) Model Selection

**Machine Learning Models:**

- **SVM (Support Vector Machine):** Binary classifier to predict probability of heart disease.
- **Random Forest:** An ensemble model to improve accuracy and handle overfitting.
- **Logistic Regression:** A simpler model for understanding the linear relationship between features and the target.
- **Deep Learning Models:**
- **ANN (Artificial Neural Network):** Used for learning complex patterns from the data.
- **CNN (Convolutional Neural Network):** If using structured data (e.g., ECG images), CNN' s can extract hierarchical features from the data.

### 5) Model Training

- **Hyper-parameter Tuning:** Use Grid Search or Random Search to find the best parameters for each model.
- **Cross-validation:** Use k-fold cross-validation to evaluate the performance of the model and avoid over-fitting.

### 6) Model Evaluation

- **Performance Metrics:** Evaluate the performance of the model using metrics such as:
- **Accuracy:** Overall percentage of correct predictions.
- **Precision, Recall, F1-score:** For class imbalance (e.g., heart disease vs. no heart disease).
- **ROC -AUC Curve:** For binary classification problems, to estimate the capacity of how well the model distinguishes between classes.

### 7) Model Optimization

- **Fine-tuning:** fine-tune the models by hyper-parameter optimization, feature enhancement, and or regularization to prevent over-fitting.
- **Ensemble Methods:** Combine multiple models (like Random Forest + Logistic Regression) to increase the accuracy and robustness.

### 8) Prediction and Deployment

- **Real-Time Prediction:** The trained model can be used in a real-time system (e.g., web app or mobile application) to make predictions whenever new patient data is input.
- **User Interface:** Create a user-friendly interface where health care professionals or patients can input the necessary attributes (e.g., age, cholesterol levels) to obtain a prediction.

### 9) Model Monitoring and Maintenance

- **Continuous Learning:** Train the model repeatedly when new patient data is available to ensure it is up to date and paying homage to changes in emerging medical trends.
- **Model Evaluation:** Continuously test the model on new data to ensure that over time, it is performing effectively.

## IV.RESULTS AND DISCUSSION


Figure 1 : Login Page

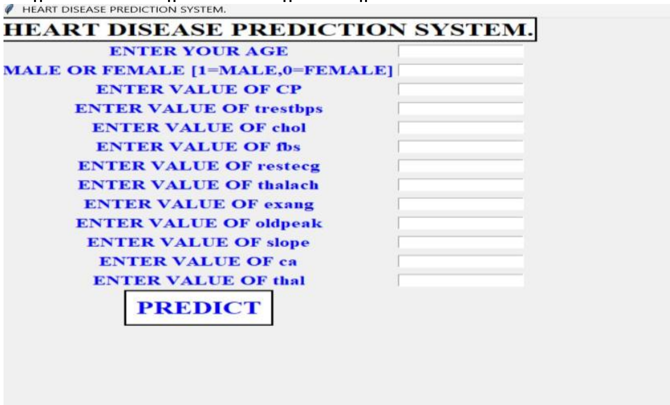


Figure 2 : User Interface

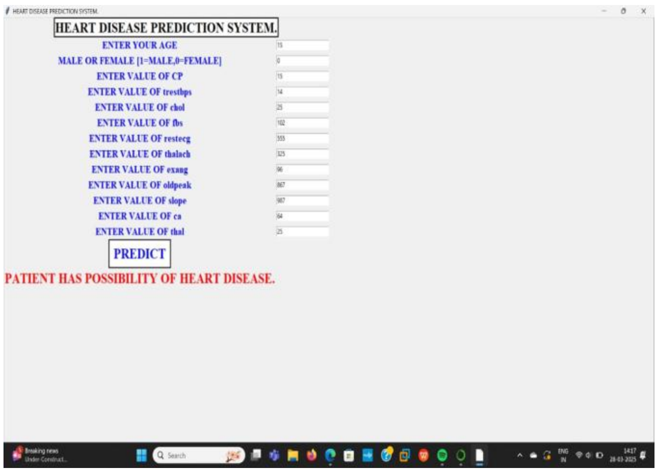


**Figure 3 : Posibilility of heart disease**

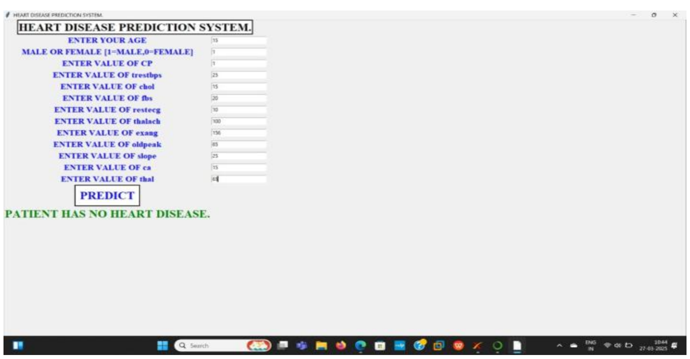


**Figure 4 : Login Page**

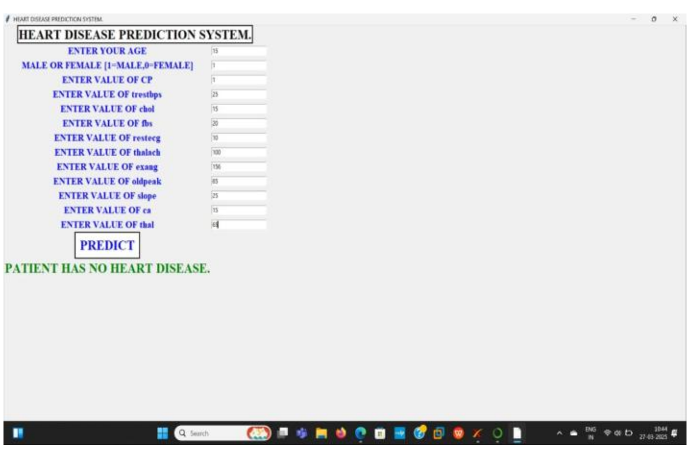


**Figure 5 : No heart disease**

## V.ACKNOWLEDGMENTS

In our endeavour to achieve success in completing project report on Smart Travel Planning with real-time integration, we take this opportunity to express our deepest sense of gratitude to our guide respected Prof. S. V. Phulari for their guidance and kind co-operation throughout the period of work. We would especially like to thank our other staff members for being patient and always supporting in the best way possible. We are thankful to our friends and family for their continuous support and motivation, which played a significant role in the successful completion of this project.

## VI.FUTURE SCOPE

- **Wearable Integration:** Machine learning models may be integrated with health wearable devices (e.g., smartwatches, fitness trackers) continuously monitoring heart rate, blood pressure, and activity levels. This would allow for real-time monitoring and instant predictive claims of a patient's heart health.
- **Individualized Predictive Models:** Further fine-tune the machine learning models once most of the personalized health information is made available to enable the presentation of predictions based upon a person's specific health profile, genetic material, and lifestyle.
- **Real-Time Clinical Decision Support:** Machine learning as well as deep learning models are integrated into a hospital information system for real-time decision support among healthcare providers such that doctors in an emergency take decisions immediately based on the required input.
- **Use of Multi-modal Data:** In the future, models can utilize this data multi-modality. For example, it is possible to combine clinical data with medical images such as ECG's, CT scans, genetic data, and lifestyle factors in an effort to improve the prediction accuracy as well as the broad risk assessment of heart disease.
- **Model Explain-ability and Transparency:** This will involve efforts on deep learning models for making more interpretative and explainable deep models especially for doctors. This would make the doctors to increase their confidence in the ability of the model predictions and increase trust that the decisions were made based on understandable and transparent results.
- **Early Detection and Preventive Healthcare:** Models of machine learning may advance up to not just predicting the potential of heart diseases but also be able to prescribe preventive measures given an individual's risk factors for heart disease may reduce the prevalence of heart diseases.
- **Continuous Learning from New Data:** The models could be designed to learn in real-time based on new data from patients so that the model is always in line with emerging medical trends, treatments, and patient demographics.

- **Cross-Platform Compatibility:** Future models will be integrated effortlessly on various healthcare platforms, thus enabling easy access of heart disease prediction for doctors, hospitals, and patients across multiple devices, increasing accessibility and usability.
- **Global Health Impact:** Calculable machine learning models when applied to multiple healthcare systems worldwide could support the global fight against heart disease in identifying high-risk populations and targeted interventions.
- **Ethics and Bias Reduction:** Continued development and research will reduce biases in models to ensure fairness in predictions and ensure that predictions are based on correct, representative data from patients that are comprised of diverse populations.

**IMPACT AND BENEFITS**

- Early Detection of Heart Disease: The system facilitates early and precise prediction of heart disease, which could assist in timely medical treatment and intervention, even saving lives.
- High Prediction Accuracy: Through the utilization of both machine learning and deep l e arning models, the system offers authentic results, enhancing diagnostic accuracy over c onventional manual procedures.
- User-Friendly Web Interface: A lightweight web-based GUI designed with flask facilitates smooth entry of data by medical practitioners and patients as well as instant provision of predictions.
- Cost and Time Efficient: Automating the prediction task minimizes costly diagnostic tests and saves time for both doctors and patients.
- Flexible and Scalable: The system can be modified to incorporate additional attributes or different diseases in t he future, there by being scalability across healthcare applications.

**CHALLENGES**

- Data Preprocessing and Quality: The medical data set contained missing, inconsistent, or unbalanced data. It was necessary to preprocess this data by cleaning, transforming, and normalizing it to make the model accurate, which consumed a lot of effort and time

[1] Feature Selection and Interpretation: Interpreting the 14 medical features and knowing what features were most influential in prediction required do main knowledge and accurate analysis.

[2] Model Choice and Tuning: Choosing the most suitable algorithms (ML vs. DL) and hyper-parameter tuning to achieve high accuracy without over fitting was the most difficult task.

[3] Small Dataset Size: There were scarce high-quality, real-world medical datasets, which affected the generalization and performance of models.

[4] Integration with Flask Web Interface: Merging the learned models into an easy-to-use web- based GUI using Flask was Challenging in handling data flow, performance, and retrieving accurate results in real time.

**[5]** Balancing Accuracy and Interpretability: While deep learning models provide more precise results, they were not as interpret able as simple machine learning models and thus predictions were hard to comprehend.

## [6] VII.CONCLUSION

[7] Heart disease forecasting is one of the new and really efficient ways to solve one of the current biggest challenges in health science. It introduces not only extra diagnostic capacity but also provides a proactive healthcare delivery orientation for this system via a machine and deep learning. Such algorithms as SVM, Random Forest, and Logistic Regression, along with more complex deep learning models such as ANN and CNN, ensure the correct detection of not only simple patterns but also complex ones in patient data.

[8] There are numerous key advantages that can be found in this prediction system. It can lead to early diagnosis for heart disease, which is crucial to avoid complications and death. Making use of an automated prediction helps reduce the workload of healthcare professionals so they may focus on the critical cases. Apart from this, data-driven insights will help reduce human error thus making the process more reliable and uniform.

[9] Although the project offers many advantages, it is not free from challenges. The quality and diversity of data used in training determine the performance of machine learning and deep learning models. The biases present in the data may propagate into incorrect predictions for certain populations, creating a demand for diverse and representative datasets. Deep learning models, however strong, are most commonly criticized for not being interpretable. Such constraints might restrict their actual application in clinical fields. They must be interpretative and explainable using them in the health care sectors.

[10] The most limiting factor is that these models are heavily computationally intensive, and hence they cannot be allowed in the under-resourced health facilities. So, future work may thus focus on optimizing the model so that they become light and efficient for use on mobile phone, or on wearable devices.

## VIII.REFERENCE

[1] Mafia Rasheed, Muhammad Adnan Khan, Nouh Sabri Elmitwally, Ghassan F. Issa, Taher M. Ghazal, Hamza Alrababah, Beenu Mago Heart Disease Prediction Using Machine Learning Method.

[2] Ali M. A. Barhoom1, Abdelbaset Almasri1, Bassem S.

Abunasser1, Samy S. Abu-Naser2, 2022, Prediction of Heart Disease Using a Collection of Machine and Deep Learning Algorithms.

[3] Wan Aezwani Wan Abu Bakar , Nur Laila Najwa Bt Josdi, Mustafa B Man, Muhammad Amierusyahmi B Zuhairi , 2023, A Review: Heart Disease Prediction in Machine Learning & Deep Learning.

[4] S. Usha, Dr.S.Kanchana,2022, Effective Analysis of Heart Disease Prediction using Machine Learning Techniques.

[5] Kumar, P., et al., 2021, A Review of Cardiovascular Disease Prediction Using Machine Learning Models.

[6] Jonayet Miah, Duc M Ca, Md Abu Sayed, Ehsanur Rashid Lipu, Fuad Mahmud, S M Yasir Arafat, 2023, Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction.

[7] Achyut Tiwari, Aryan Chugh, Aman Sharma, 2023, Ensemble Framework for Cardiovascular Disease Prediction.

[8] Shadab Hussain, Dr. Santosh Kumar Nanda, Susmith Barigidad, Shadab Akhtar, Md Suaib, Niranjan K. Ray, 2021, Novel Deep Learning Architecture for Heart Disease Prediction using Convolutional Neural Network.

[9] Nidhi Sinha, Teena Jangid, Amit M. Joshi, Saraju P. Mohanty, 2022, iCardo: A Machine Learning-Based Smart Healthcare Framework for Cardiovascular Disease Prediction.

[10] Apurb Rajdhan, Milan Sai, Avi Agarwal, Dundigalla Ravi, Dr. Poonam Ghuli, 2020, Heart Disease Prediction using Machine Learning.

[11] Govardhan Logabiraman, D. Ganesh, M. Sunil Kumar, A. Vinay Kumar and Nitin Bhardwaj, 2024, heart disease prediction using machine learning algorithms.