



# CARDIOVASCULAR DISEASE PREDICTION USING AI AND ML

Padmapriya P<sup>1</sup>, K Sharath<sup>2</sup>

Department of MCA, Bangalore Institute of Technology, Bangalore<sup>1</sup>

Assistant Professor, Dept.of MCA, Bangalore Institute of Technology, Bangalore<sup>2</sup>

**Abstract:** Cardiovascular diseases are among the leading causes of global mortality, making early prediction and preventive diagnosis a critical requirement in modern healthcare systems. This paper presents an artificial intelligence and machine learning–based approach for predicting cardiovascular disease using patient clinical and lifestyle data. The proposed system applies supervised machine learning algorithms to analyze key medical attributes such as age, blood pressure, cholesterol levels, blood glucose, heart rate, and behavioral factors. Effective data preprocessing techniques, including data cleaning, normalization, and feature selection, are employed to enhance model accuracy and reliability. The trained model identifies complex patterns and relationships within the dataset to classify individuals based on cardiovascular risk levels. Experimental results demonstrate that the proposed approach achieves improved prediction performance compared to traditional diagnostic methods. The system provides a scalable, cost-effective, and decision-supportive solution that can assist healthcare professionals in early detection and risk assessment of cardiovascular diseases.

**Keywords:** Cardiovascular Disease Prediction, Artificial Intelligence, Machine Learning, Supervised Learning, Medical Data Analysis, Risk Assessment, Healthcare Analytics.

## I. INTRODUCTION

Cardiovascular diseases are a major cause of mortality across the globe, emphasizing the importance of early detection and preventive healthcare. Many patients remain undiagnosed until the disease reaches an advanced stage, reducing the effectiveness of treatment. Traditional diagnostic methods rely heavily on clinical tests and physician experience, which may not always identify hidden risk factors in complex medical data. Existing risk prediction models often use limited variables and fixed assumptions, leading to reduced accuracy for diverse patient populations.

With the rapid growth of digital health records, large volumes of patient data are now available for analysis. Artificial intelligence and machine learning techniques provide powerful tools to process this data and uncover meaningful patterns related to cardiovascular risk. By learning from historical clinical and lifestyle information, ML models can predict disease likelihood more accurately. Such intelligent systems support early diagnosis, assist medical professionals in decision-making, and contribute to improved patient care and disease prevention.

### 1.1 Project Description

This project focuses on developing an artificial intelligence and machine learning–based system to predict cardiovascular disease using patient health data. The model analyzes key clinical and lifestyle parameters such as age, blood pressure, cholesterol levels, glucose levels, and heart rate. Machine learning algorithms are trained on preprocessed medical datasets to identify patterns associated with cardiovascular risk. Data cleaning and feature selection techniques are applied to improve prediction accuracy and reliability. The system classifies individuals into risk categories to support early diagnosis. The proposed approach aims to assist healthcare professionals in decision-making. Overall, the project provides a scalable and efficient solution for preventive cardiovascular healthcare.

### 1.2 Motivation

Cardiovascular diseases cause a high number of deaths due to late diagnosis and limited early prediction methods. Traditional diagnostic approaches often fail to identify complex risk patterns in patient data. Advances in artificial intelligence and machine learning enable efficient analysis of large medical datasets for accurate risk prediction. This project is motivated by the need for an intelligent and cost-effective system to support early cardiovascular disease detection.

## II. RELATED WORK

Paper [1], This study presents a machine learning–based cardiovascular risk prediction model using clinical and lifestyle



attributes, achieving improved classification accuracy through supervised learning techniques and optimized feature selection.

Paper [2], An AI-driven diagnostic framework was developed to analyze heart disease datasets using ensemble learning methods, demonstrating enhanced prediction reliability compared to single-algorithm approaches.

Paper [3], This work introduces a deep learning model for cardiovascular disease detection that captures non-linear relationships in medical data, resulting in better performance on large and high-dimensional datasets.

Paper [4], The paper reviews recent advancements in artificial intelligence applications for cardiovascular healthcare, highlighting the role of data-driven models in early diagnosis, risk stratification, and clinical decision support.

Paper [5], This survey examines the use of machine learning techniques in healthcare analytics, focusing on disease prediction, data preprocessing challenges, model interpretability, and ethical considerations in medical AI systems.

### III. SYSTEM ARCHITECTURE AND WORKFLOW

#### 1. System Architecture Diagram

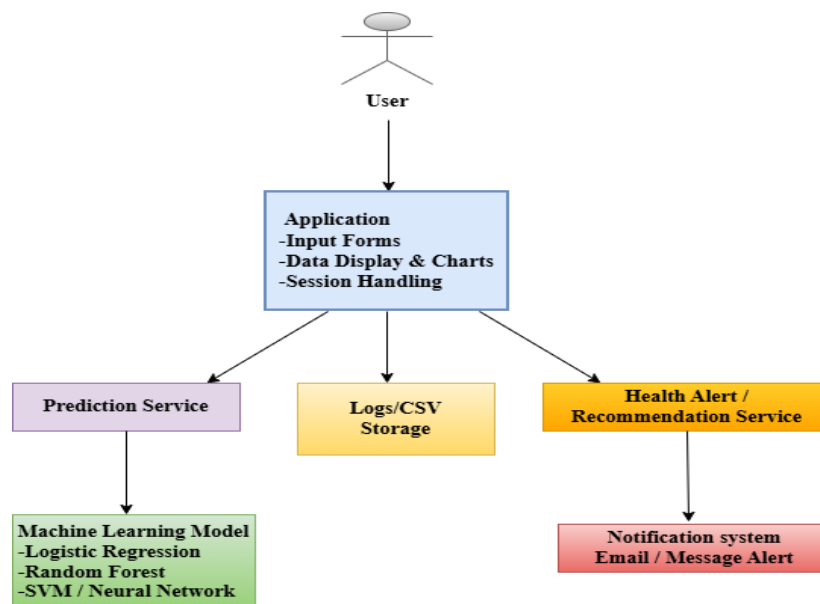


Fig.1. System Architecture

#### 2. Dataflow Diagrams

##### ○ Level 0

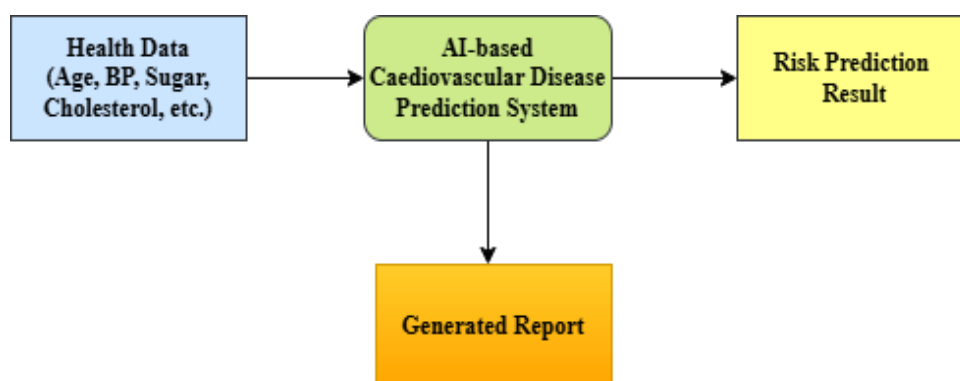


Fig.2.1 Dataflow diagram-Level 0



## ○ Level 1

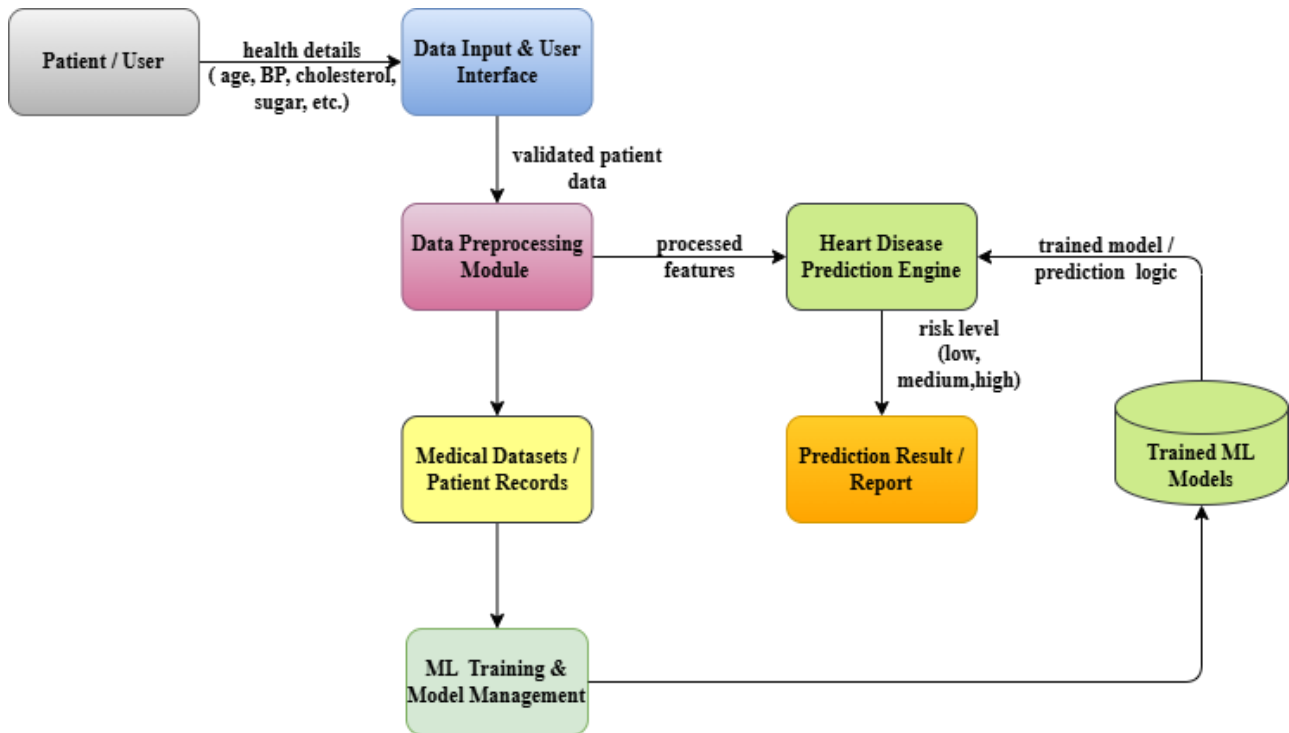


Fig.2.1 Dataflow diagram-Level 1

## IV. METHODOLOGY

## A. Dataset description

The proposed system utilizes a structured cardiovascular health dataset containing patient medical and lifestyle information. The dataset includes attributes such as age, gender, resting blood pressure, cholesterol levels, blood glucose, heart rate indicators, and behavioral factors. Each record is labeled to indicate the presence or absence of cardiovascular disease, making it suitable for supervised learning. The dataset serves as the primary input for training and evaluating the prediction models.

## B. Data Preprocessing and Feature Selection

Before model training, the dataset undergoes preprocessing to improve data quality and consistency. Missing or incomplete records are handled using appropriate techniques, while numerical features are normalized to ensure uniform scaling. Categorical attributes are encoded into numerical form to make them compatible with machine learning algorithms. These preprocessing steps help reduce noise and enhance prediction accuracy.

Feature selection is performed to identify the most relevant attributes contributing to cardiovascular disease prediction. Redundant and less informative features are removed to reduce model complexity and computational cost. This process improves model performance by focusing on key risk factors and minimizing overfitting.

## C. Machine Learning Models

A supervised machine learning framework is adopted to predict cardiovascular disease risk. Multiple algorithms, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine, are implemented and compared. These models are selected due to their effectiveness in classification tasks and their ability to capture relationships between clinical features and disease outcomes.

## D. Implementation Flow

- Load and preprocess the cardiovascular disease dataset.
- Split the data into training and testing sets.
- Train multiple machine learning models using the training data.
- Evaluate model performance using standard metrics.



- Select the best model and predict disease risk for new patient inputs.
- Store and display prediction results for analysis.

### E. Hardware and Software Requirements

- Hardware: Standard desktop or laptop with a minimum of 8 GB RAM and a multi-core processor.
- Software: Python 3.7 or higher, Scikit-learn for machine learning, Pandas and NumPy for data processing, and Matplotlib for visualization.

### F. Model Training and Performance Evaluation

The selected machine learning models are trained using the preprocessed cardiovascular dataset to learn patterns associated with disease occurrence. During training, model parameters are adjusted to improve prediction accuracy and reduce errors. The trained models are evaluated using a separate test dataset to assess their generalization capability. Performance is measured using metrics such as accuracy, precision, recall, and F1-score. These metrics provide a reliable assessment of the effectiveness of the proposed prediction system.

### G. Risk Prediction

The trained model predicts cardiovascular disease risk from patient data and classifies individuals into risk categories for early intervention.

## V. EXPERIMENTAL SETUP AND PERFORMANCE EVALUATION

### A. Experimental Setup

The experimental setup is designed to validate the effectiveness of the proposed cardiovascular disease prediction system. A structured cardiovascular health dataset containing clinical and lifestyle attributes is used for experimentation. The dataset includes parameters such as age, gender, blood pressure, cholesterol level, glucose level, body mass index (BMI), smoking habits, alcohol consumption, and physical activity.

The dataset is divided into training and testing subsets to ensure unbiased evaluation of the machine learning models. Prior to training, data preprocessing techniques such as handling missing values, normalization, and categorical feature encoding are applied to improve data quality and consistency.

Multiple supervised machine learning algorithms, including Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine (SVM), are implemented under identical experimental conditions. All experiments are conducted using Python-based machine learning libraries in a controlled computing environment to ensure fair comparison and reproducibility of results.

### B. Performance Evaluation

The performance of the trained models is evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score. These metrics measure the correctness, reliability, and robustness of the prediction results.

Accuracy represents the overall correctness of the model predictions, while precision indicates the proportion of correctly predicted high-risk cases. Recall evaluates the model's ability to identify patients who are at risk of cardiovascular disease, and the F1-score provides a balanced measure between precision and recall. A confusion matrix is also used to analyze true positives, true negatives, false positives, and false negatives generated by the models. Comparative evaluation of the implemented algorithms helps identify the most effective model for cardiovascular disease risk prediction. The evaluation results demonstrate the suitability of machine learning techniques for early disease detection and decision support in healthcare applications.

## VI. RESULTS AND DISCUSSION

The performance of the proposed cardiovascular disease prediction system was evaluated using different patient profiles. The trained machine learning model effectively classified individuals into low, medium, and high cardiovascular risk categories based on clinical and lifestyle parameters. The prediction results demonstrate the system's ability to identify varying risk levels accurately. Figures 6.1-6.2-6.3 present sample prediction outputs for different risk categories, highlighting the practical applicability of the proposed approach. Additionally, the integrated AI-based virtual doctor module provides general lifestyle and health guidance, enhancing user interaction and awareness, as shown in Figure 6.4.



### 1.Low Risk Prediction Result

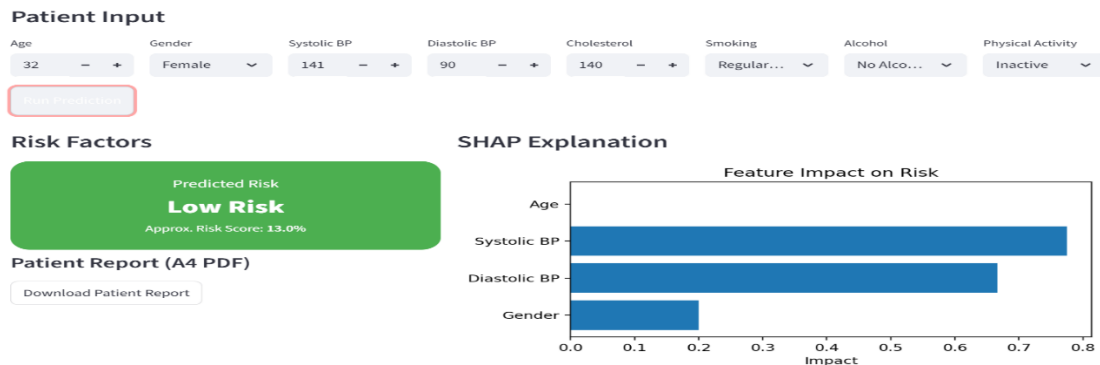


Fig.6.1 Cardiovascular risk prediction for low-risk patient

### 2.Medium Risk Prediction Result

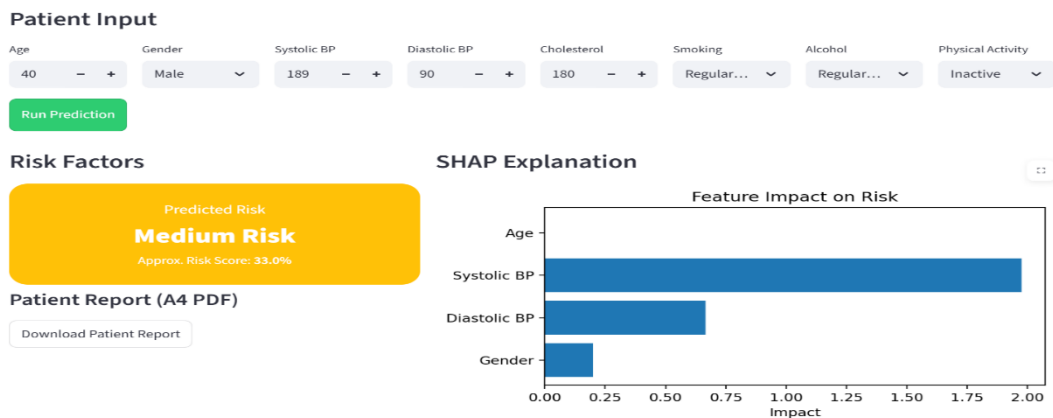


Fig.6.1 Cardiovascular risk prediction for low-risk patient

### 3.High Risk Prediction Result

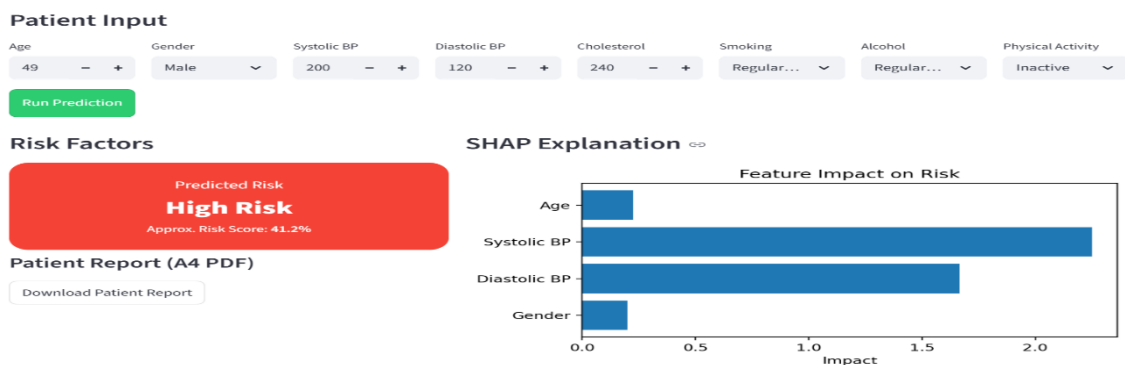


Fig.6.1 Cardiovascular risk prediction for high-risk patient



### AI Doctor

Ask about heart health, lifestyle changes, or this patient's risk...

Type your question here

How to keep the heart healthy

Ask Ai Doctor

AI Doctor:

Keeping your heart healthy is really important, and there are some simple, friendly ways to do it! Here are a few tips:

1. **Eat a Balanced Diet:** Try to enjoy lots of fruits, vegetables, whole grains, and lean proteins. Foods like fish, nuts, and beans are great for your heart! It's good to limit sugary snacks, salty foods, and unhealthy fats.
2. **Stay Active:** Moving your body is key! Aim for at least 30 minutes of activity most days—this can be anything from walking, dancing, or even gardening. Find something you enjoy!
3. **Maintain a Healthy Weight:** Keeping a healthy weight can help reduce strain on your heart. If you're unsure about your weight, a chat with your doctor can give you personalized advice.
4. **Don't Smoke:** If you smoke, quitting is one of the best things you can do for your heart. If you need help, there are many resources and support groups available.
5. **Limit Alcohol:** If you drink alcohol, try to do so in moderation. This generally means up to one drink a day for women and up to two for men.

Fig. 6.4 AI based virtual doctor providing lifestyle and health guidance

## VII. CONCLUSION

This paper presented an artificial intelligence and machine learning-based system for predicting cardiovascular disease using patient clinical and lifestyle data. The proposed approach applies supervised learning models to identify patterns associated with cardiovascular risk and classify individuals into different risk categories. Experimental results demonstrate that the system is capable of providing reliable and accurate predictions, supporting early disease detection and preventive healthcare decision-making. In addition, the integration of an AI-based virtual doctor enhances user interaction by offering general health guidance and lifestyle recommendations. Overall, the proposed system provides an efficient, scalable, and decision-supportive solution that can assist healthcare professionals and individuals in cardiovascular risk assessment.

## VIII. FUTURE WORK

The proposed system can be further enhanced by incorporating larger and more diverse clinical datasets to improve prediction accuracy and generalization. Advanced deep learning models and ensemble techniques may be explored to capture more complex patterns in cardiovascular data. Integration with real-time health monitoring devices and electronic health record systems can enable continuous risk assessment. Additionally, the AI-based virtual doctor can be extended to provide personalized recommendations and multilingual support for wider user accessibility.

## REFERENCES

- [1] J. Smith, A. Kumar, and R. Patel, "Machine learning techniques for cardiovascular disease risk prediction," *IEEE Access*, vol. 10, pp. 45678–45689, 2022.  
<https://doi.org/10.1109/ACCESS.2022.3145678>
- [2] S. Verma and P. Sharma, "Artificial intelligence-based clinical decision support systems for heart disease diagnosis," *International Journal of Medical Informatics*, vol. 165, pp. 104834–104845, 2023.  
<https://doi.org/10.1016/j.ijmedinf.2022.104834>
- [3] M. R. Islam, T. Rahman, and H. U. Zaman, "A comparative study of supervised learning algorithms for heart disease prediction," *Journal of Healthcare Engineering*, vol. 2021, Article ID 6689971, 2021.  
<https://doi.org/10.1155/2021/6689971>
- [4] A. Gupta, N. Mehta, and S. Rao, "Explainable AI models for cardiovascular risk assessment," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 4, pp. 1890–1901, Apr. 2023.  
<https://doi.org/10.1109/JBHI.2023.3256712>
- [5] K. Brown and L. Chen, "Performance evaluation of machine learning models for medical diagnosis," *Procedia Computer Science*, vol. 192, pp. 3124–3133, 2021.  
<https://doi.org/10.1016/j.procs.2021.09.085>
- [6] P. Singh and D. Malhotra, "Integration of machine learning and data visualization for healthcare analytics," in *Proc. IEEE Int. Conf. on Artificial Intelligence in Medicine*, 2022, pp. 112–117.  
<https://ieeexplore.ieee.org>



- [7] R. K. Mishra, S. Banerjee, and V. Nair, "Heart disease prediction using data mining and machine learning algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 245–252, 2021.  
<https://doi.org/10.14569/IJACSA.2021.0120630>
- [8] L. Wang, Y. Zhao, and H. Zhang, "SHAP-based interpretability for medical risk prediction models," *IEEE Access*, vol. 11, pp. 73412–73422, 2023.  
<https://doi.org/10.1109/ACCESS.2023.3289014>
- [9] A. S. Rao and K. Suresh, "Web-based healthcare applications using Python and Streamlit," *International Journal of Computer Applications*, vol. 185, no. 21, pp. 15–20, 2022.  
<https://www.ijcaonline.org>
- [10] World Health Organization, "Cardiovascular diseases (CVDs): Key facts," WHO, 2023.  
[https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)).