

**PROJECT REPORT**

**ON**

**“ Cardiac Risk Prediction”**

Under the guidance of

* Mr. AKASH V

**TABLE OF CONTEXT:**

SL.NOCONTENT

1. Introduction
2. Proposed System AdvantagesSystem
3. Advantages and Disadvantages
4. HardwareDisadvantages4Hardware and Software Requirements
5. Conclusion

**Introduction**

Heart disease is one of the major causes of death in the world today. Predicting heart problems at an early stage can help save many lives. In this project, we use **Machine Learning (ML)**, a type of technology that allows computers to learn from data and make predictions.

We use patient data like **age, blood pressure, cholesterol, chest pain type**, and other health details to check if a person is at risk of heart disease. We apply different ML techniques such as **Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, Random Forest**, and **Support Vector Machine (SVM)** to build prediction models.

The main goal of this project is to compare how well these models work. We check their performance using accuracy and other simple methods to see which one gives the best results. This project shows how ML can help doctors and hospitals in making faster and more accurate decisions about heart health.

**Proposed System**

**Project Idea**

This project focuses on predicting whether a person is at risk of heart disease using machine learning techniques. By analyzing patient health data (like age, blood pressure, cholesterol, etc.), we apply classification algorithms such as Logistic Regression, K-Nearest Neighbors (KNN), and Decision Trees to build predictive models.

The dataset is preprocessed, the models are trained and evaluated using accuracy and confusion matrix, and the best-performing model is identified. This system can help in early diagnosis, enabling better medical decisions and timely treatment.

We used Python along with libraries like Pandas, scikit-learn, and Matplotlib. The project demonstrates how machine learning can assist in solving real-world healthcare problems.

**Components of the System**

To address the growing need for early and accurate detection of heart disease, the proposed system leverages supervised machine learning algorithms to predict whether a person is at risk or not, based on their medical data. The system is designed to be efficient, interpretable, and helpful for healthcare professionals.

### **1. Input Dataset**

The system uses a structured dataset containing medical records of patients.

Common attributes include:'male', 'age', 'BPMeds', 'prevalentHyp', 'diabetes','totChol', 'sysBP', 'diaBP', 'BMI', 'glucose'

### **2. Data Preprocessing**

**Handling missing values** if present.

**Converting categorical data** into numerical form using encoding.

**Normalizing/Scaling** features to bring all values to the same range.

**Splitting the dataset** into training and testing sets (e.g., 80:20 or 70:30).

### **3. Model Building**

Multiple supervised ML models are trained and tested:

**Logistic Regression**

**K-Nearest Neighbors (KNN)**

**Support Vector Machine (SVM)**

**Decision Tree**

**Random Forest:**Each model is trained using the training data and tuned using hyperparameters to improve accuracy.

### **4. Model Evaluation**

The models are evaluated using the following performance metrics:

**Accuracy** – Correct predictions overall.

**Precision** – How many predicted "heart disease" cases were correct.

**Recall (Sensitivity)** – How many actual "heart disease" cases were detected.

**F1-Score** – Balance between precision and recall.

**Confusion Matrix** – Summary of prediction results.

**ROC-AUC Curve** – To visualize model performance.

### **5. Best Model Selection**

Based on the above evaluation, the best-performing model is selected.

The model with the **highest accuracy and balanced performance** is recommended for real-world deployment.

**Advantages**

**Early Detection & Prevention**  
Helps in identifying patients at high risk early, enabling timely medical intervention.

**Automation of Risk Assessment**  
Reduces manual work for doctors by automatically analyzing multiple health parameters.

**Cost-Effective**  
Reduces the need for expensive tests by using existing patient data to assess risk.

**Improved Decision-Making**  
Supports healthcare professionals with data-driven insights, improving diagnosis accuracy.

**Model Reusability**  
Once trained, models can be reused and updated with new data for continuous improvement.

**Real-Time Predictions**  
Fast predictions allow integration into mobile or web applications for real-time use.

**Disadvantages**

**Data Quality Dependency**  
The model's accuracy depends heavily on the quality and completeness of the dataset.

**Lack of Explainability**  
Some models (like Random Forest or SVM) may act like "black boxes" and lack clear explanations for their decisions.

**Bias & Overfitting Risks**  
If the dataset is biased or small, the model may overfit or provide misleading results.

**Not a Replacement for Doctors**  
It's a supportive tool, not a substitute for medical professionals or diagnostic tests.

**Privacy Concerns**  
Handling patient health data needs strict privacy protection and ethical considerations.

**Software & Hardware requirements**

**Hardware requirements:**

* Computer: A standard laptop or desktop computer with sufficient processing power and memory capacity to handle data analysis tasks.
* Processor: A multi-core processor (e.g., Intel Core i5 or higher) for faster computation of machine learning algorithms.
* Memory (RAM): At least 8 GB of RAM is recommended to handle large datasets and complex machine learning models efficiently.

**Software requirements:**

• Python: The programming language used for data analysis, machinelearning, and model development.

* Jupyter Notebook: An interactive development environment that allows for easy prototyping, visualization, and documentation of data analysis workflows.
* NumPy: A fundamental library for numerical computing in Python, essential for handling arrays and mathematical operations
* Pandas: A powerful library for data manipulation and analysis, particularly useful for loading, cleaning, and preprocessing datasets.
* Scikit-learn: A machine learning library in Python that provides a wide range of algorithms for classification, regression, clustering, and model evaluation.

**Conclusion**

This project demonstrated how machine learning can be effectively applied to predict the likelihood of heart disease using patient medical data. By preprocessing the dataset, selecting meaningful features, and applying various classification algorithms like Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and Support Vector Machine (SVM), we were able to build predictive models capable of assisting healthcare professionals in early diagnosis.

The evaluation results showed that some models performed better in terms of accuracy and reliability, helping us identify the most suitable algorithm for this type of classification problem. Through this project, we also explored the importance of feature selection, data normalization, and model evaluation metrics such as accuracy, precision, recall, and F1-score.

Overall, this study highlights the practical impact of supervised machine learning in the healthcare domain, particularly in reducing the risk of delayed diagnosis and supporting timely treatment decisions. It also opens up possibilities for further enhancements, such as integrating more advanced algorithms, larger datasets, or real-time health monitoring systems.