AI-POWERED HEART DISEASE RISK ASSESSMENT SYSTEM

Team Members:

R. Sahithi Reddy - 5F8
T. Mahalakshmi - 5K3
V. Nitya Vaishnavi - 5K6

PROBLEM STATEMENT

- Heart disease is a major global health issue and early detection is vital to prevent serious outcomes.
- Traditional diagnosis methods are often time-consuming and expensive.
- There is a need for an efficient, data-driven system that can predict the likelihood of heart disease based on patient health data.

PROPOSED SOLUTION

The Heart Disease Prediction System uses machine learning to predict heart disease risk from medical parameters like age, cholesterol, and blood pressure.

LITERATURE SURVEY

1. UCI Heart Disease Dataset

- The UCI Repository provided an early heart disease dataset with 303 records and 13 standard clinical features.
- It was widely used in earlier research for testing traditional ML models.

2. Kaggle Heart Disease Dataset (Used in Our Project)

- We used the Kaggle dataset, which contains 1,025 records, offering a larger and more diverse sample for training.

DATASET

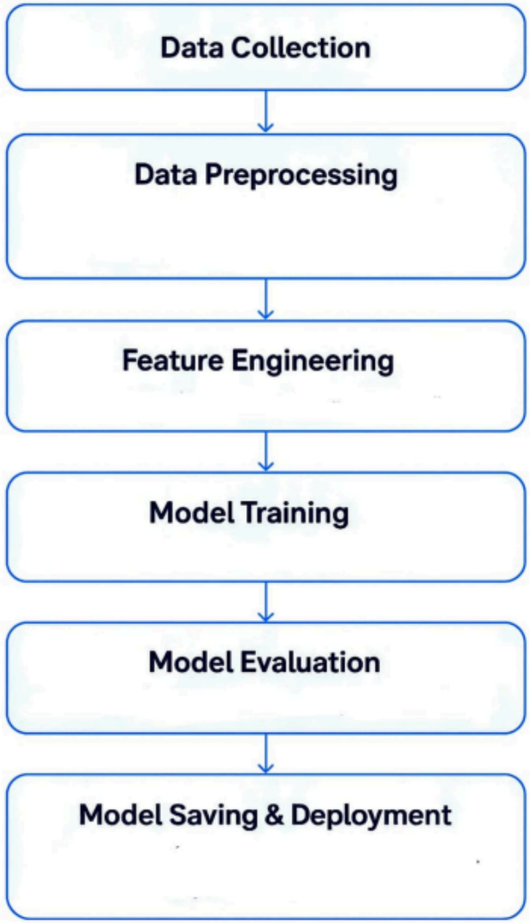
KAGGLE DATASET BASED ON CLEVELAND, HUNGARY, SWITZERLAND, LONG BEACH V

- age: Age of the patient.
- sex: Gender of the patient.
- cp: Chest pain type.
- trestbps: Resting blood pressure (in mm Hg).
- chol: Serum cholesterol (in mg/dl).
- fbs: Fasting blood sugar.
- restecg: Resting electrocardiographic results.
- thalach: Maximum heart rate achieved.
- exang: Exercise induced angina.
- oldpeak: ST depression induced by exercise relative to rest.
- slope: The slope of the peak exercise ST segment.
- ca: Number of major vessels colored by fluoroscopy.
- thal: Thalassemia (a blood disorder).

MODEL COMPARISION

Criteria	Support Vector Machine (SVM)	Random Forest Classifier
Algorithm Type	Kernel-based classifier	Ensemble of multiple decision trees
Accuracy	0.875 (87.5%)	0.9415 (94.15%)
AUC Score	0.942	0.9946
Sensitivity (Recall)	0.902	0.9697
Specificity	0.885	0.9151
Cross-Validation (Mean Score)	0.821	0.8488 ± 0.0839
Training Speed	Slower	Faster
Final Outcome	Performed well	Best Performing Model

METHODOLOGY



MODEL AND IMPLEMENTATION

Random Forest Classifier

Robust with mixed data types, provides feature importance, and is less prone to overfitting on this type of dataset.

Tools Used - Python, Pandas, Scikit-learn, Streamlit, Plotly

Implementation-

- Data processing
- Feature Engineering.
- Model Training
- Probability Calibration

Feature Engineering:

- Age Categories: Young (<45), Middle (45-65), Senior (>65)
- BP Categories: Normal, Elevated, Stage 1, Stage 2 Hypertension
- Cholesterol Levels: Desirable, Borderline, High
- Medical Risk Score: Weighted combination of risk factors
- Interaction Features: Age×BP, Cholesterol×ST Depression

Data Preprocessing:

- Automatic outlier detection and removal
- Target encoding verification
- Missing value handling

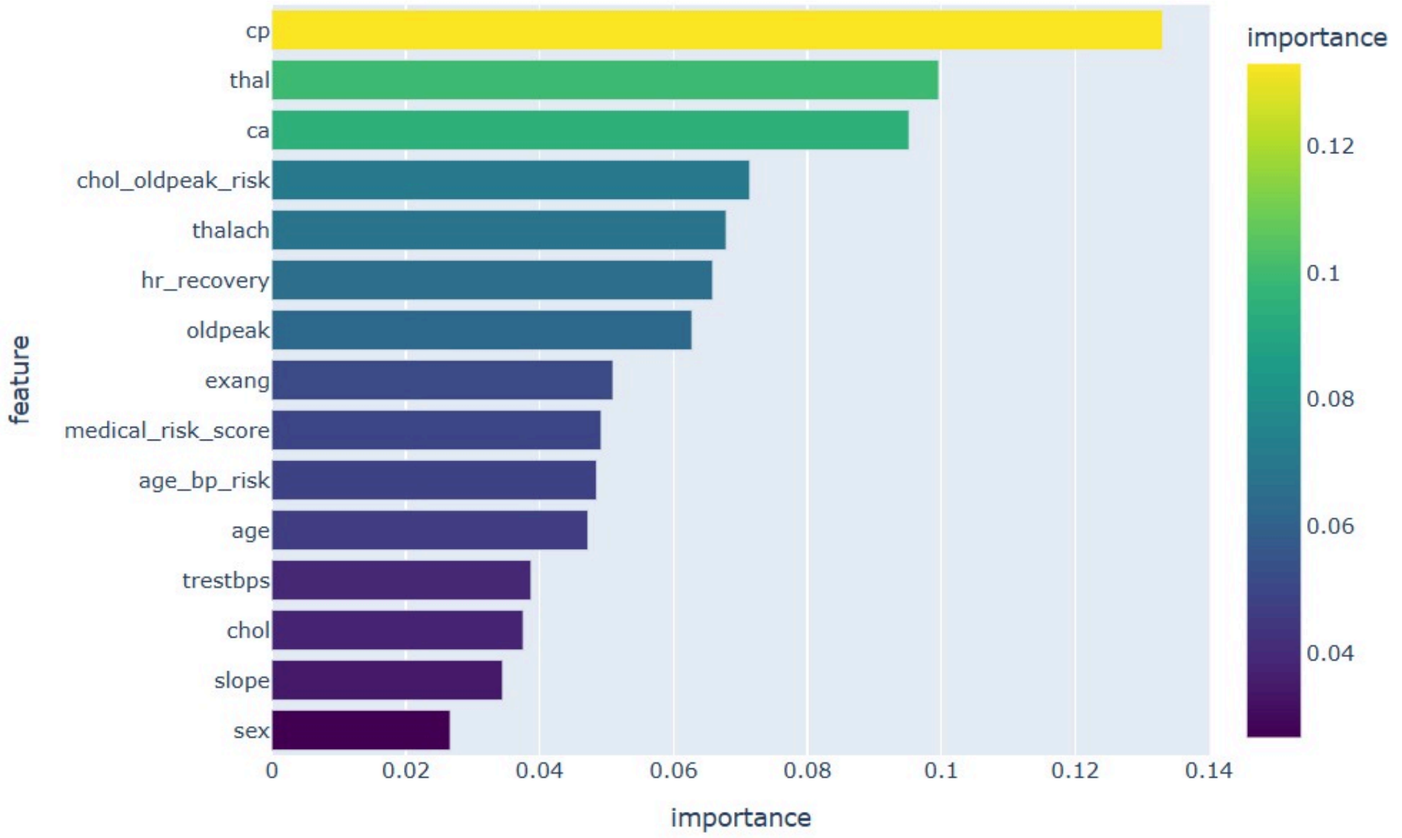
Risk Classification:

🟢 LOW RISK (0-30%): Continue healthy habits
🟡 MODERATE RISK (30-70%): Consult doctor, lifestyle changes
🔴 HIGH RISK (70-100%): Immediate medical consultation

Test Cases Used:

- Low Risk - Young Healthy (24% probability)
- Moderate Risk - Middle Age (55% probability)
- High Risk - Multiple Factors (71% probability)
- Borderline Case (27% probability)
- Healthy Elderly (20% probability)

RESULTS


cp
thal
ca
chol_oldpeak_risk
thalach
hr_recovery
oldpeak
exang
medical_risk_score
age_bp_risk
age
trestbps
chol
slope
sex
feature
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
importance
importance
0.12
0.1
0.08
0.06
0.04

MODEL PERFORMANCE

📊 Model Evaluation:
✅ Accuracy: 0.9415
📊 AUC Score: 0.9946

📋 Classification Report:

	precision	recall	f1-score	support
0	0.97	0.92	0.94	106
1	0.91	0.97	0.94	99
accuracy			0.94	205

CONCLUSION

This project successfully demonstrates how machine learning can be combined with medical expertise to create an accurate, accessible, and explainable tool for early heart disease screening, with the potential to save lives

Future Work

- Deploy app to cloud (AWS/Azure).
- Retrain model with diverse data.

REFERENCES

- Janosi, A., Steinbrunn, W., Pfisterer, M., & Detrano, R. (1989). Heart Disease Dataset. UCI Machine Learning Repository.
- Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.

THANK YOU