Abstract

Application

Heart disease remains one of the leading causes of mortality worldwide. Early detection using data-driven tools can significantly reduce the risk of severe outcomes by assisting clinicians in timely diagnosis and treatment planning. The proposed application is a **heart disease prediction system** that uses patient health parameters such as age, blood pressure, cholesterol, ECG results, and other clinical factors to predict whether an individual is likely to have heart disease. The system provides a simple **web-based interface** allowing users or healthcare practitioners to input patient details and instantly receive predictive results for better medical decision support.

Importance of the Application

Traditional diagnosis often requires expensive or invasive tests, while many developing regions lack access to cardiologists. This system offers an accessible, low-cost, and rapid alternative by leveraging **machine learning** to process medical data and determine disease likelihood automatically. Implementing such a predictive model helps in early risk identification, prioritization of high-risk patients, and optimization of hospital resources. It also provides interpretability for practitioners by highlighting the most influential health parameters affecting prediction.

Methods Used and Advantages

1. Principal Component Analysis (PCA)

PCA was used for **dimensionality reduction**, transforming the original **13 medical features** (from the Cleveland dataset) into a smaller set of uncorrelated principal components while maintaining most of the variation in the data.
Advantages:

- Eliminates redundancy and correlation among attributes.
- Lowers computational complexity.
- Increases model performance by mitigating overfitting when training on small datasets.

2. Support Vector Machine (SVM)

An **SVM** classifier using the **radial basis function (RBF)** kernel was trained on the **PCA**-reduced data to classify patients as either having or not having heart disease.
Advantages:

- Highly effective for **binary medical classification**.
- Performs well on small and nonlinear datasets.
- Provides a clear **margin of separation**, enhancing generalization capability.

Together, **PCA and SVM** form a robust model pipeline capable of accurate detection while maintaining efficiency and interpretability.

Result

The dataset from the **UCI Cleveland Heart Disease repository** was used. After preprocessing missing values and performing train-test splitting, the **PCA + SVM** pipeline achieved strong classification performance, with accuracy around **84–87%**, precision of **0.85**, and recall of **0.82** depending on the hyperparameters. Model training and evaluation confirm the feasibility of combining dimensionality reduction with a kernel-based classifier for disease prediction. The resulting web interface enables interactive prediction, returning messages such as "**Heart disease detected**" or "**No heart disease detected**" along with confidence probabilities.

Overall, the model demonstrates a practical approach to **AI-assisted clinical screening**, helping reduce diagnostic delays and supporting early treatment recommendations.

Product Output: