Heart Disease Classification Using Feature Extraction and Selection

---

Published By:

Kavitha, R., & Kannan, E. (2016, February).

Presented at the International Conference on Emerging Trends in Engineering, Technology, and Science (ICETETS) and published by IEEE.

Introduction:

Heart disease classification remains a significant challenge in healthcare due to the complexity and variability of medical data. This project builds upon the framework proposed by Kavitha and Kannan (2016), which emphasizes the importance of feature extraction and feature selection in developing an efficient and accurate classification system.

By focusing on refining input data through robust feature engineering techniques, this framework aims to enhance the performance of machine learning algorithms, thereby improving the reliability of heart disease prediction models.

Methodology:

1. Feature Extraction:

- Key features were derived from patient datasets, including age, cholesterol levels, resting blood pressure, and heart rate.
- Dimensionality reduction techniques, such as Principal Component Analysis (PCA), were used to identify the most significant features.

2. Feature Selection:

- **Wrapper Methods:** Recursive Feature Elimination (RFE) was applied to select the most predictive features.

- **Filter Methods:** Techniques such as correlation analysis and chi-square tests were used to eliminate redundant or irrelevant features.

3. Classification Algorithms:

The refined dataset was used to train and test machine learning models, including:

- Support Vector Machines (SVM)
- Decision Trees
- Random Forest
- Naïve Bayes

4. Evaluation Metrics:

- Accuracy, precision, recall, and F1 score were used to assess model performance.
- Cross-validation ensured that the results were consistent and generalizable.

Conclusion:

The proposed framework demonstrated significant improvements in classification accuracy by combining feature extraction and selection techniques. Random Forest achieved the highest accuracy, while SVM excelled in handling high-dimensional datasets.

This study underscores the critical role of feature engineering in optimizing machine learning models for heart disease prediction. Future enhancements will focus on integrating real-time patient monitoring data and exploring deep learning approaches for further accuracy improvements.