Classification of Heart Disease Using K-Nearest Neighbour and Genetic Algorithm

Authors:

B. L. Deekshatulu and **P. Chandra**

Abstract

Heart disease continues to be a major cause of mortality worldwide, necessitating efficient and accurate diagnostic tools. This study introduces a hybrid approach for heart disease classification, combining the K-Nearest Neighbor (KNN) algorithm with a Genetic Algorithm (GA). KNN is employed for classification due to its simplicity and effectiveness, while GA is used for feature selection, improving model performance by identifying the most relevant attributes. The system was evaluated using clinical datasets and metrics such as accuracy, precision, and recall. Results demonstrate that the hybrid approach significantly enhances classification accuracy compared to using KNN alone, providing a robust solution for heart disease diagnosis.

Introduction

Heart disease poses a significant threat to global health, accounting for a substantial percentage of deaths annually. Accurate diagnosis at

an early stage is crucial for effective treatment and prevention. Traditional diagnostic methods often struggle with the complexity and variability of medical data. Advances in artificial intelligence and machine learning have enabled the development of more robust and efficient diagnostic tools. This paper proposes a hybrid approach combining the K-Nearest Neighbor (KNN) algorithm for classification and a Genetic Algorithm (GA) for feature selection. The integration of GA optimizes the performance of KNN by selecting the most informative features, improving diagnostic accuracy and reducing computational complexity.

Methodology

1. **Dataset Collection and Preprocessing**:
    - Clinical datasets containing patient attributes such as age, blood pressure, cholesterol levels, and other diagnostic indicators were used.
    - Data preprocessing included handling missing values, normalizing numerical data, and encoding categorical variables.
2. **Feature Selection Using Genetic Algorithm**:
    - GA was employed to identify the most relevant features from the dataset.
    - Steps in GA:
        - **Initialization**: A population of potential feature subsets was generated.
        - **Fitness Evaluation**: Each subset was evaluated based on its classification accuracy using KNN.
        - **Selection**: The best-performing subsets were selected for reproduction.

        - **Crossover and Mutation**: New feature subsets were generated to explore the search space.

3. **Classification with K-Nearest Neighbor**:
    - KNN was applied to classify patients based on the selected features.
    - The value of K (number of neighbors) was optimized using cross-validation.
4. **Performance Evaluation**:
    - The system was evaluated using metrics such as accuracy, precision, recall, and F1-score.
    - Comparative analysis was performed between KNN with GA-based feature selection and KNN without feature selection.

Conclusion

The proposed hybrid approach integrating K-Nearest Neighbor (KNN) with Genetic Algorithm (GA) demonstrates significant improvements in the classification of heart disease. By optimizing feature selection, the GA enhances the efficiency and accuracy of the KNN model. The experimental results highlight the system's potential as a reliable tool for heart disease diagnosis, offering a balance between simplicity and effectiveness. Future research could explore the inclusion of real-time data, larger datasets, and other advanced classification techniques to further refine the model's performance. This study underscores the importance of hybrid machine learning techniques in advancing healthcare diagnostics.