Heart Disease Prediction Using Machine Learning With Flask

Aarthie.M[1], V.C.Nathiya[2], R.Meenatchi[3]
[1]Dept of CSE
[2, 3]Assistant Professor
[1, 2, 3] Kathir College of Engineering Wisdom Tree Neelambur,Coimbatore,India

Abstract- *Heart disease is a major cause of death worldwide. Early diagnosis and treatment can improve outcomes for patients. Machine learning techniques, particularly the decision tree algorithm, have shown great promise in aiding the diagnosis and prediction of heart disease. This study aims to develop a machine learning model using the decision tree algorithm to predict heart disease. We used a dataset containing patient information such as age, sex, chest pain, cholesterol levels, and other medical conditions. After preprocessing the dataset and selecting relevant features, we trained a decision tree model. The model's performance was evaluated using accuracy metrics and a confusion matrix. A web application using Flask was implemented to provide easy and interactive access to the model for medical professionals.*

Keywords- Heart Disease, Machine Learning, Decision Tree, Flask, Prediction Model

I. INTRODUCTION

Heart disease is a critical health issue, being a leading cause of death globally. The heart, a vital organ, pumps blood to all parts of the body. Any malfunction can lead to severe consequences, including death within minutes. Modern lifestyle changes, work-related stress, and poor dietary habits have contributed to the increased incidence of heart disease.

According to the World Health Organization (WHO), heart-related diseases are responsible for 17.7 million deaths annually, accounting for 31% of all global deaths. In India, heart diseases have become the leading cause of mortality. Data mining and machine learning techniques are valuable tools in extracting significant information from vast medical datasets, aiding in disease diagnosis and prediction.

II. LITERATURE SURVEY

1. **Geumkyung Nah, Eun-A Choi et al.** (2021) - This study explored gene expression patterns of known COPD loci among patients with varying disease severity. The findings highlighted differences in gene expression, enhancing our understanding of COPD risk and pathogenesis.
2. **Hiroki Fuse, Kota Oishi et al.** (2019) - This research utilized brain shape information from MRI images to classify healthy subjects and Alzheimer's patients, achieving a classification accuracy of 87.5%, suggesting the potential of shape information in diagnosis.
3. **Sina Akbarian, Mark P. Nelder et al.** - This study developed an automated tool to identify ticks related to Lyme disease using convolution neural networks, achieving 92% accuracy in tick species classification.
4. **Luyao Wang et al.** (2022) - This research employed edge functional connectivity (eFC) to explore brain network differences between Parkinson's Disease patients and healthy controls, demonstrating the efficiency of eFC in disease diagnosis.
5. **Peng Ni, Jianxin Wang** (2019) - This study introduced ModuleSim, a method to measure disease associations based on disease module theory, showing significant correlation with disease classification and outperforming other methods.
6. **K. Srinivas, B. Kavihta Rani, and A. Govrdhan** (2010) - This paper discussed the use of naive Bayes, neural networks, and decision tree algorithms for heart disease prediction, emphasizing the importance of feature selection and data preprocessing in improving model accuracy.
7. **M. Anbarasi, E. Anupriya, and N. Chandrasekaran** (2010) - This research focused on enhancing the performance of heart disease prediction models by reducing the number of attributes, using techniques like genetic algorithms for feature selection.
8. **Jabbar, M.A., Deekshatulu, B.L., and Chandra, P.** (2013) - The study analyzed the performance of various classification algorithms, such as support vector machines and decision trees, for heart disease prediction, highlighting the trade-offs between accuracy and computational efficiency.
9. **Ghumbre, S.U., Patil, P.D., and Ghatol, A.A.** (2011) - This paper presented a hybrid model combining k-means clustering and support vector machines to improve the accuracy of heart disease prediction.

III. METHODOLOGY

3.1 Data Collection

The dataset used in this study was collected from various online sources, including Kaggle and UCI Machine Learning Repository. It contains records of patients with attributes such as age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise, the slope of the peak exercise ST segment, number of major vessels, and thalassemia.

3.2 Data Preprocessing

Preprocessing steps included:

1. **Handling Missing Values**: Missing values were imputed using the mean or median of the respective feature.
2. **Normalizing Data**: Continuous variables were normalized to bring them to a comparable scale.
3. **Encoding Categorical Variables**: Categorical variables were encoded using one-hot encoding or label encoding as appropriate.

3.3 Feature Selection

Relevant features were selected based on their correlation with the target variable (heart disease presence). Techniques such as Recursive Feature Elimination (RFE) and Principal Component Analysis (PCA) were used to identify the most significant features.

3.4 Model Training

Several machine learning algorithms were explored, including:

1. **Decision Tree**: Chosen for its interpretability and simplicity.
2. **Logistic Regression**: Used as a baseline model.
3. **Random Forest**: An ensemble method that combines multiple decision trees to improve accuracy.
4. **K-Nearest Neighbors (KNN)**: A non-parametric method used for classification.
5. **Support Vector Machine (SVM)**: Used for its effectiveness in high-dimensional spaces.

3.5 Model Evaluation

Models were evaluated using metrics such as accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). A confusion matrix was also used to analyze the classification performance.

3.6 Deployment

The final model was deployed using the Flask web framework, providing an interactive user interface for medical professionals to input patient data and receive heart disease risk predictions.

3.7 Flask Implementation

The Flask application includes the following features:

1. **User Authentication**: Secure login for medical professionals.
2. **Data Input Forms**: Easy-to-use forms for entering patient data.
3. **Prediction Output**: Display of prediction results along with confidence scores.
4. **Data Visualization**: Graphical representation of patient data and model performance.

IV. SYSTEM DESIGN

4.1 Existing System

The existing system uses non-invasive risk prediction models involving factors such as age, gender, blood pressure, heart rate, and electrocardiography. These models apply generalized linear models with regularization paths to estimate heart disease status.

4.2 Existing Block Diagram

- Database
- PC (Machine Learning)
    - Collect Dataset

4.3 Disadvantages

1. Heart disease can severely limit daily activities, reducing quality of life.
2. It increases the risk of other health issues such as stroke, kidney disease, and peripheral artery disease.

4.4 Proposed System

The proposed system uses Flask to allow users to input medical data (age, gender, blood pressure, cholesterol levels, family history, lifestyle habits). A machine learning algorithm analyzes this data to assess heart disease risk and provide recommendations for lifestyle changes or medical interventions. The system stores user data securely and offers reminders for follow-up appointments or tests.

4.5 Proposed Block Diagram

- User Input
- Machine Learning Algorithm
  - Feature Extraction
  - Model Training
  - Prediction Output
- Flask User Interface

4.6 Flow Chart

- Training Data
- Feature Extraction
- Machine Learning Algorithm
- Model
- Predicted Output
- Flask User Interface

4.7 Use Case Diagram

- Data Load
- Specified Algorithm
- Data Preprocessing
- Predicted Data

4.8 Activity Diagram

- Start
- Import Disease Status
- Import Dataset
- Analyze Data

4.9 Class Diagram

- Data Preprocessing
- Feature Extraction
- Output Result Analysis
- Machine Learning Algorithm

4.10 Sequence Diagram

- Heart Disease Prediction
- KNN Algorithm
- Database Record
- Predicting Outcome
- Output

V. MODULES

5.1 Dataset Collection

Collect data on patient demographics and medical history relevant to heart disease.

5.2 Data Preprocessing

Clean and preprocess data to handle missing values, normalize features, and encode categorical variables.

5.3 Feature Extraction

Select relevant features that significantly impact heart disease prediction.

5.4 Model Training

Train the decision tree model using the preprocessed data and extracted features.

5.5 Model Evaluation

Evaluate the model's performance using accuracy metrics, confusion matrix, and other relevant indicators.

5.6 Deployment

Implement the model in a web application using Flask, providing an interactive interface for medical professionals.

VI. RESULTS

The decision tree model achieved an accuracy of 85% in predicting heart disease. The confusion matrix highlighted the model's capability to correctly classify patients with and without heart disease. The Flask application enabled easy access to the model, facilitating its use by healthcare providers.

The dataset used for training consisted of over 300 patient records with attributes such as age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise, the slope of the peak exercise ST segment, number of major vessels, and thalassemia.

VII. DISCUSSION

The integration of machine learning in heart disease prediction can significantly aid early diagnosis and treatment. The decision tree algorithm's interpretability makes it a valuable tool for medical professionals. Future work may involve incorporating more sophisticated models and expanding the dataset for improved accuracy.

The system's robustness can be enhanced by including more diverse data sources, such as real-time monitoring data from wearable devices and integrating genetic data to provide personalized risk assessments. Moreover, leveraging advanced machine learning techniques like ensemble learning, deep learning, and natural language processing for unstructured medical records could further enhance prediction accuracy.

VIII. CONCLUSION

Machine learning, particularly the decision tree algorithm, holds great potential in predicting heart disease. The developed model, coupled with a user-friendly Flask application, can assist healthcare professionals in making informed decisions, ultimately improving patient outcomes.
The future direction of this research includes developing a comprehensive mobile application for continuous patient monitoring and risk assessment. Incorporating real-time data analytics and cloud computing will make the system more scalable and accessible. Additionally, collaboration with healthcare providers to validate the model's effectiveness in clinical settings will be pursued.

REFERENCES

[1] Nah, G., & Choi, E. (2021). Gene expression analysis of known COPD loci revealed its varied levels by disease severity. IEEE.
[2] Fuse, H., & Oishi, K. (2019). Detection of Alzheimer's Disease with Shape Analysis of MRI Images. IEEE.
[3] Akbarian, S., & Nelder, M. P. (2019). A Computer Vision Approach to Identifying Ticks Related to Lyme Disease.
[4] Ni, P., & Wang, J. (2019). Constructing Disease Similarity Networks Based on Disease Module Theory. IEEE.
[5] Wang, L., et al. (2022). Brain network analysis between Parkinson's Disease and Health Control based on edge functional connectivity. IEEE.
[6] Srinivas, K., Rani, B. K., & Govrdhan, A. (2010). Application of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. International Journal on Computer Science and Engineering.
[7] Anbarasi, M., Anupriya, E., & Chandrasekaran, N. (2010). Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm. International Journal of Engineering Science and Technology.
[8] Jabbar, M.A., Deekshatulu, B.L., & Chandra, P. (2013). Heart Disease Prediction using Lazy Associative Classification. IEEE.
[9] Ghumbre, S.U., Patil, P.D., & Ghatol, A.A. (2011). Heart Disease Diagnosis Using Support Vector Machine. International Conference on Computer Science and Information Technology.
[10] World Health Organization: Cardiovascular Diseases
[11] Centers for Disease Control and Prevention: Heart Disease Facts
[12] National Heart, Lung, and Blood Institute: Heart Disease
[13] Kaggle Datasets
[14] Awesome Public Datasets