Detecting Cardio Vascular problem using ML

- Shwetha Anand
RA2111027010045

Overview

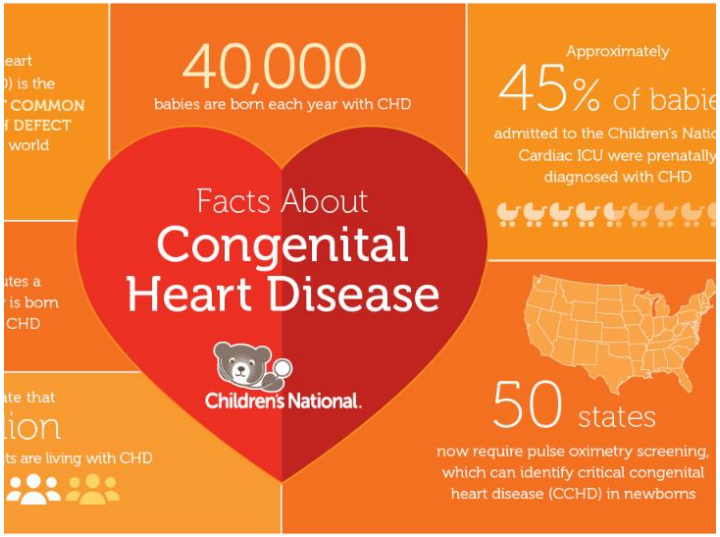
1. **Introduction**
   - Problem Identification
   - Motivation
2. **Problem Formulation**
3. **Algorithm**
4. **Methodology**
5. **Literature Review**
6. **Data Collection**
7. **Applications**
8. **References**

Introduction

- Detecting cardiovascular problems using Machine Learning is a significant application in healthcare.
- Cardiovascular diseases are a leading cause of mortality worldwide, and early detection is crucial for timely intervention.
- Here's a simplified outline for creating a cardiovascular disease detection model using Python and scikit-learn.

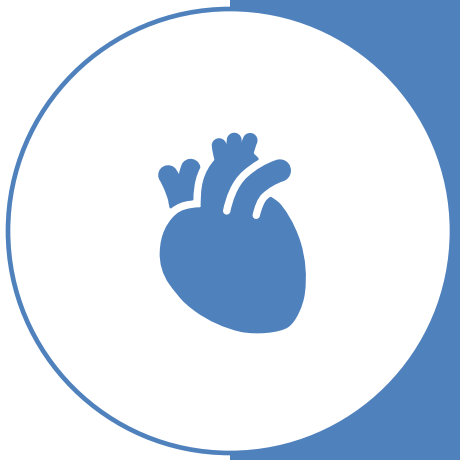
Motivation

- **Cardiovascular diseases are the leading cause of death globally. They account for a substantial proportion of deaths in both developed and developing countries.**
- **According to the World Health Organization (WHO), an estimated 17.9 million people died from CVDs in 2019, representing 32% of all global deaths.**
- **Cardiovascular diseases include conditions such as - heart failure, stroke, hypertension, arrhythmia**



Problem Formulation

- Heart disease is the leading cause of death among all other diseases, even cancer.
- Due to lack of resources in the medical field, the prediction of heart disease may occasionally be a problem.
- Utilization of suitable technology support in this area can prove to be highly beneficial to the medical fraternity and patients.
- The project intends to automatically detect cardiovascular disease using two datasets through a deep learning network and a variety of machine learning classification models.



Algorithm

Step 1: Data Collection

- Gather a dataset with a significant number of patient records, including features such as age, sex, blood pressure, cholesterol levels, heart rate, medical history, and other relevant information. You may also include the target variable, which indicates whether a patient has cardiovascular disease or not.

Step 2: Data Preprocessing

- Clean and preprocess the data to handle missing values, outliers, and format the data for analysis. This may involve standardization, normalization, or encoding categorical variables.

Step 3: Feature Selection

- Determine which features are most relevant for predicting cardiovascular disease. Feature selection techniques, such as correlation analysis or recursive feature elimination, can help you identify the most important variables.

Step 4: Data Splitting

- Divide the dataset into training and testing subsets. Typically, you might use 70-80% of the data for training and the remaining 20-30% for testing the model's performance.

Step 5: Model Selection

- Choose the machine learning model(s) you want to use for your algorithm. Common models for binary classification tasks like this include logistic regression, decision trees, random forests, support vector machines, and deep learning models (e.g., neural networks).

Step 6: Model Training

- Train the selected model(s) on the training dataset. The model will learn the patterns and relationships within the data that are associated with the presence or absence of cardiovascular disease.

Step 7: Model Evaluation

- Assess the model's performance using evaluation metrics such as accuracy, precision, recall, F1-score, and the receiver operating characteristic (ROC) curve. You can also calculate the area under the ROC curve (AUC) to measure the model's ability to discriminate between positive and negative cases.

Step 8: Model Validation

- Validate the model's performance on the testing dataset to ensure it generalizes well to new, unseen data.

Methodology

Input Heart disease data

↓

Replace Missing values with column means

↓

Model Building

↓

Calculating accuracy for constructed models.

Literature Review

S.NO	Techniques	Title of work	Results/Limitation
1)	K-nearest neighbor	Analysis of heart disease prediction using various machine learning techniques year-2018	To achieve better accuracy and to make the system more efficient so that it can predict the chances of heart attack.
2)	Naïve Bayes Decision tree Random Forest	A Hybrid Machine Learning Approach for Prediction of Heart Diseases Year-2018	In this paper, to develop a prediction system that be capable to envisage heart diseases based on measurements are extracted from THE ERIC lab.
3)	Decision tree 91% Naïve Bayes 87%	Prediction of Heart Disease using Machine Learning Algorithms year-2018	In this system, a heart disease data set is used. The main aim of this system is to predict the possibilities of occurring heart disease of the patients in terms of percentage.

Data Collection

- **Data Preprocessing:** Clean and preprocess the dataset by handling missing values, normalizing numeric features, and encoding categorical variables.
- **Feature Selection and Engineering:** Select relevant features that contribute to the prediction of cardiovascular diseases. You can also engineer new features, such as BMI or ratios of certain measurements.
- **Model Selection:** Choose appropriate machine learning algorithms for classification. Common algorithms include Logistic Regression, Random Forest, Support Vector Machines, and Neural Networks.

- **Model Training:** Split the dataset into training and validation sets. Train the selected model on the training data and tune hyperparameters for optimal performance. Consider techniques like cross-validation.
- **Model Evaluation:** Evaluate the trained model on the validation set using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Choose the evaluation metrics based on the specific goals of the project.
- **Hyperparameter Tuning:** Fine-tune the model's hyperparameters to improve its performance. Techniques like Grid Search or Random Search can help find optimal parameter values.

- **Deployment:** Deploy the trained model in a user-friendly interface, such as a web application or a mobile app, where users can input their health data and receive predictions about their cardiovascular health.
- **Ethical Considerations:** Be mindful of ethical considerations, patient privacy, and data security. Ensure compliance with relevant regulations when working with medical data.
- **Validation and Collaboration:** Collaborate with medical professionals to validate the model's predictions and clinical relevance. Their expertise is essential for refining the model and ensuring its reliability.

Applications

- **Medical Institutions** :-

To teach medical students how the heart attack been measured or how to identify that the person is suffering from heart disease.

- **Hospitals** :-

To detect that is the person having disease or not.

- **Stress Testing** :-

Stress tests, such as exercise stress tests or pharmacological stress tests, can assess how the heart responds to increased workload.

References

- **Datasets:**
  - Cardiovascular Disease Dataset
- **Tutorials and Frameworks:**
  - Scikit-learn Documentation
  - TensorFlow and Keras for Neural Networks
- **Interpretability:**
  - Interpretable Machine Learning with SHAP
  - Interpretable Machine Learning: A Guide for Making Black Box Models Explainable