CSET211 – SML

PROJECT REPORT

PROJECT MILESTONE 03
B.TECH (CSE)
3RD SEMESTER

Topic~ CardioVision

UNDER THE GUIDANCE OF

PROF. PRASHANT KAPIL

(Assistant Professor)

SUBMITTED BY

-RAJNISH SINGH E23CSEU0094

CONTENTS

INTRODUCTION

Heart disease remains one of the leading causes of death worldwide, making early detection and prevention crucial. The Cardio Vision project aims to develop a machine learning model to predict the risk of heart disease based on various health parameters. By leveraging the UCI Heart Disease dataset, this project seeks to provide accurate predictions that can aid in early diagnosis and preventive healthcare. The primary objective is to create a reliable predictive system that can be used by healthcare professionals to identify high-risk individuals and take proactive measures.

➢ Background and Motivation

Cardiovascular diseases account for a significant portion of global mortality, with millions of deaths each year. Traditional diagnostic methods, while effective, often require invasive procedures and can be time-consuming. The advent of machine learning offers a non-invasive, efficient alternative that can analyse vast amounts of data to identify patterns indicative of heart disease. This project is motivated by the potential to harness these advanced techniques to improve patient outcomes and reduce the burden on healthcare systems.

- ✓ *Here are some statistics about them:*
  - *Number of deaths: In 2021, 20.5 million people died from CVDs globally, a 60% increase from 1990.*

- *Location of deaths: More than three-quarters of CVD deaths occur in low- and middle-income countries.*
- *Age of deaths: One-third of CVD deaths occur in people under 70 years of age.*
- *Causes of death: 85% of CVD deaths are caused by heart attacks and strokes.*
- *Risk factors: Risk factors for CVDs include age, sex, race, family history, and lifestyle factors like unhealthy diet, tobacco use, and physical inactivity.*

Note - One person dies every 33 seconds from cardiovascular disease.

➢ Objectives

The main objectives of the Cardio Vision project are:

- *Develop a Predictive Model: Create a machine learning model that accurately predicts the risk of heart disease.*
- *Enhance Early Detection: Enable early diagnosis through predictive analytics, allowing for timely medical intervention.*
- *Improve Healthcare Efficiency: Provide a tool that healthcare professionals can use to quickly assess patient risk, thereby improving the efficiency of medical evaluations.*
- *Contribute to Research: Add to the body of knowledge in the field of medical research by demonstrating the practical application of machine learning in healthcare.*

RESEARCH SYNOPSIS

A comprehensive literature survey was conducted to understand the current state of research in heart disease prediction using machine learning. The following papers were reviewed:

1. A Comprehensive Review on Heart Disease Risk Prediction using Machine Learning and Deep Learning Algorithms
   - *Domain: Cardiovascular disease prediction*
   - *Dataset: Various datasets including UCI Heart Disease*
   - *Methodology: Comparison of machine learning and deep learning techniques*
   - *Performance Metrics: Accuracy, precision, recall*
   - *Summary: This paper reviews the effectiveness of different machine learning and deep learning algorithms in predicting heart disease, highlighting the need for further research to improve prediction accuracy Family Budgeting:*
   - *Source:* SpringerLink

2. Survey on Heart Disease Prediction Using Machine Learning Techniques
   - *Domain: Heart disease prediction*
   - *Dataset: Multiple datasets from healthcare databases*
   - *Methodology: Analysis of 25 papers using various ML models*

- *Performance Metrics: Performance achievements, research gaps*
- *Summary: This survey analyzes various machine learning models used for heart disease prediction, identifying the most effective techniques and highlighting areas for future research.*
- *Source:* SpringerLink

3. A Survey on Prediction Techniques of Heart Disease using Machine Learning
    - *Domain: Heart disease prediction*
    - *Dataset: UCI Heart Disease dataset*
    - *Methodology: Overview of ML techniques*
    - *Performance Metrics: Accuracy, F1-score*
    - *Summary: This paper provides an overview of current machine learning techniques for heart disease prediction, emphasizing the importance of data preprocessing and feature selection.*
    - ***Source:*** IJERT

4. A Survey for Societal Care and Information using Machine Learning
    - *Domain: Heart disease prediction*
    - *Dataset: Various medical datasets*
    - *Methodology: Survey of ML techniques*
    - *Performance Metrics: Accuracy, precision, recall*

- *Summary: This survey evaluates different machine learning techniques for heart disease prediction, focusing on their performance and applicability in clinical settings.*
- *Source:* SpringerLink

5. Cardiovascular Disease Prognosis and Analysis Using Machine Learning Techniques
   - *Domain: Cardiovascular disease prognosis*
   - *Dataset: UCI Heart Disease dataset*
   - *Methodology: Comparative analysis of ML models*
   - *Performance Metrics: Accuracy, ROC-AUC*
   - *Summary: This paper compares various machine learning models for predicting cardiovascular disease, highlighting the strengths and weaknesses of each approach.*
   - *Source:* Journal of Big Data

6. Predicting the Heart Disease Using Machine Learning Techniques
   - *Domain: Heart disease prediction*
   - *Dataset: UCI Heart Disease dataset*
   - *Methodology: Implementation of ML models*
   - *Performance Metrics: Accuracy, precision, recall*
   - *Summary: This study implements several machine learning models to predict heart disease, demonstrating the effectiveness of logistic regression and decision trees.*
   - *Source:* SpringerLink

DATASETS

The dataset used in this project is the UCI Heart Disease dataset, which includes 14 attributes such as age, sex, chest pain type, resting blood pressure, cholesterol levels, and more. This dataset is publicly available and widely used for heart disease prediction research. It contains 303 rows and 14 columns, providing a comprehensive set of features for analysis.

➢ Data Preprocessing

Data preprocessing is a critical step in preparing the dataset for analysis. The following steps were taken to preprocess the data:

- ***Handling Missing Values**: The dataset was checked for missing values, and any missing entries were imputed using the mean or median of the respective columns.*
- ***Encoding Categorical Variables**: Categorical variables such as chest pain type and sex were converted into numerical values using one-hot encoding to make them compatible with machine learning algorithms.*
- ***Normalization**: Numerical features such as age, resting blood pressure, and cholesterol levels were normalized to ensure they are on a similar scale, which helps improve the performance of the machine learning model.*
- ***Splitting Data**: The dataset was divided into training and testing sets to evaluate the model's performance.*

METHODOLOGY

The methodology section outlines the approach taken to develop the heart disease prediction model. The project follows a structured workflow, including data collection, preprocessing, model training, and evaluation.

Hardware and Software Requirements

The project was implemented using the following hardware and software:

- *Hardware:*
  - *A standard personal computer with at least 8GB of RAM and a modern processor (e.g., Intel i5 or equivalent).*
- *Software:*
  - *Programming Language: Python*
  - *Libraries: NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn*
  - *Development Environment: Jupyter Notebook or Google Colaboratory*

Performance Metrics

The performance of the machine learning model was evaluated using the following metrics:

- *Accuracy: The proportion of correctly predicted instances out of the total instances.*
- *Precision: The proportion of true positive predictions out of the total positive predictions.*
- *Recall: The proportion of true positive predictions out of the total actual positives.*
- *F1-Score: The harmonic mean of precision and recall, providing a single metric that balances both.*

RESULTS AND ANALYSIS

The results section presents the findings from the model training and testing. The logistic regression model was trained on the preprocessed dataset, and its performance was evaluated using the testing set.

- ***Model Accuracy**: The logistic regression model achieved an accuracy of 85% on the training data and 82% on the testing data.*
- ***Precision and Recall**: The model demonstrated high precision and recall, indicating its effectiveness in predicting heart disease.*

- ***F1-Score***: *The F1-score was calculated to provide a balanced measure of the model's performance.*

Visualizations such as confusion matrices and ROC curves were used to analyze the model's performance further. The results indicate that the logistic regression model is a reliable tool for predicting heart disease, with a good balance between precision and recall.

CONCLUSIONS AND FUTURE WORKS

The Cardio Vision project successfully developed a machine learning model to predict heart disease risk based on various health parameters. The logistic regression model demonstrated high accuracy and reliability, making it a valuable tool for early detection and prevention of heart disease.

- ***Model Improvement***: *Explore other machine learning models such as decision trees, random forests, and deep learning techniques to improve prediction accuracy.*

- ***Feature Engineering***: *Investigate additional features that could enhance the model's performance.*
- ***Real-world Application***: *Develop a user-friendly application that healthcare professionals can use to input patient data and receive heart disease risk predictions.*
- ***Data Expansion***: *Incorporate larger and more diverse datasets to improve the model's generalizability and robustness.*

Thank You
