

**Heart Disease Prediction**

An AiMl Mini Project report submitted

In partial fulfilment of the requirements for the

Degree of Bachelor of

Engineering/Technology

In

Computer Engineering (CP)

Semester – VI

**Submitted By**

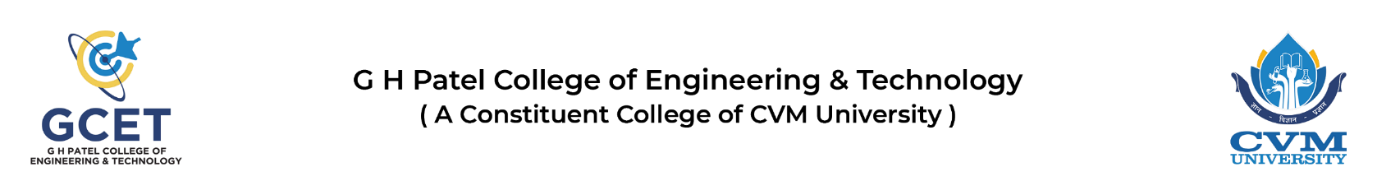
**Meet Dadhaniya(12202040501038)**

**Param Dholakia(12202040501049)**

**Artificial Intelligence &**

**Machine Learning (202046702)**

**A.Y. 2024-25 ODD TERM**



**G H Patel College of Engineering & Technology**

**CERTIFICATE**

This is to certify that **Meet Dadhaniya** (**12202040501038**) & **Param Dholakia** (**12202040501049)** has submitted the Aiml Mini Project report on **" Heart Disease Prediction "** for partial fulfilment of the degree of Bachelor of Engineering in **Computer Engineering**, **G H Patel College of Engineering and Technology**, at The Charutar Vidya Mandal (CVM) University, Vallabh Vidyanagar, during the academic year 2024 – 25.

Dr. Priyang Bhatt

(Internal Faculty Guide)

# INTRODUCTION

This project aims to predict the likelihood of heart disease in individuals based on various health metrics using machine learning models. The primary goal is to leverage data-driven insights to identify patterns and risk factors associated with heart disease, enabling early detection and intervention. The dataset includes a range of health-related features, and three machine learning models—

* Random Forest,
* K-Nearest Neighbors (KNN), and
* Gradient Boosting

—are employed to classify individuals as having heart disease or not.

# DATASET DESCRIPTION

The dataset contains the following features:

|  |  |
| --- | --- |
| ID | Unique identifier for each individual |
| Age | Age of the individual (in years) |
| Gender | Gender of the individual (Male/Female) |
| Height\_c | Height in centimeters |
| Weight\_kg | Weight in kilograms |
| BMI | Body Mass Index |
| Daily\_Steps | Number of steps taken daily |
| Calories\_Intake | Daily calorie intake (in calories) |
| Hours\_of\_Sleep | Hours of sleep per day |
| Heart\_Rate | Resting heart rate (in beats per minute) |
| Blood\_Pressure | Blood pressure reading (in mmHg) |
| Exercise\_Hours\_per\_Week | Hours of exercise per week |
| Smoker | Smoking status (Yes/No) |
| Alcohol\_Consumption\_per\_Week | Alcohol consumption per week (in units) |
| Diabetic | Diabetic status (Yes/No) |
| Heart\_Disease | Presence of heart disease (Yes/No, target variable) |

The dataset provides a comprehensive view of each individual’s health profile, with Heart\_Disease as the target variable.

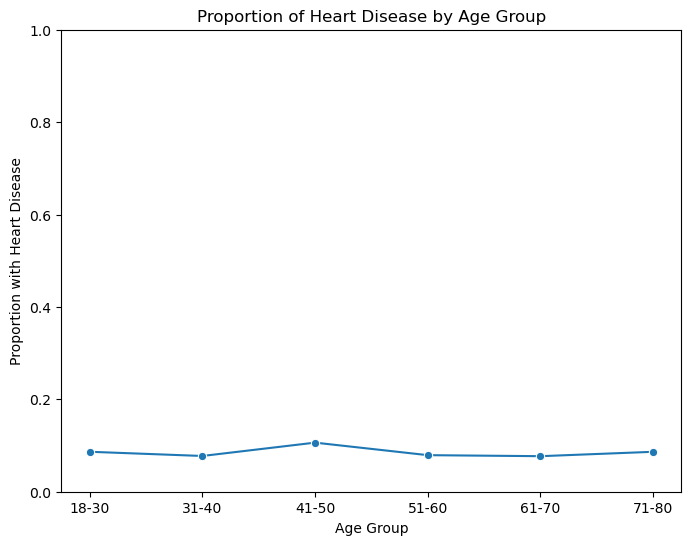
# DATA PREPROCESSING

To prepare the dataset for modeling, the following preprocessing steps were applied:

* Categorical Variable Encoding: Categorical features such as Gender (Male=0, Female=1), Smoker (No=0, Yes=1), Diabetic (No=0, Yes=1), and Heart\_Disease (No=0, Yes=1) were mapped to numerical values.
* Blood Pressure Transformation: The Blood\_Pressure feature was split into two numerical features: Max\_BP (systolic) and Min\_BP (diastolic).

These steps ensured the dataset was numerical and ready for modeling.

# EXPLORATORY DATA ANALYSIS

****A line plot was created to show the proportion of individuals with heart disease across age groups, highlighting age as a potential risk factor.

# MODEL TRAINING AND EVALUATION

Three models were trained:

* Random Forest: Ensemble of decision trees.
* KNN: Distance-based classifier.
* Gradient Boosting: Sequential tree ensemble.

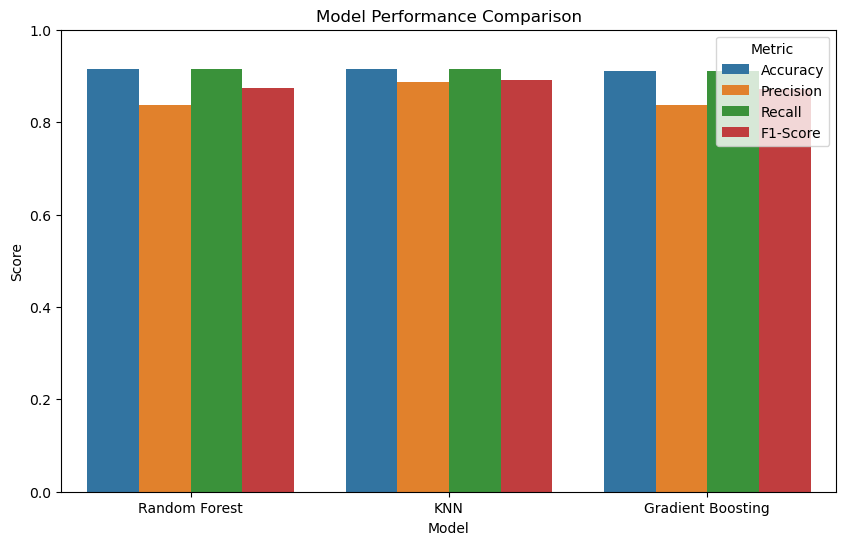
Performance was evaluated using accuracy, precision, recall, and F1-score.

MODEL PERFORMANCE METRICS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-Score |
| Random Forest | 0.915 | 0.837 | 0.915 | 0.874 |
| KNN | 0.915 | 0.887 | 0.915 | 0.890 |
| Gradient Boosting | 0.910 | 0.837 | 0.910 | 0.872 |

# MODEL PERFORMANCE VISUALIZATION

A bar plot compares the models’ performance across metrics.



# CONCLUSION

* The project demonstrates the use of machine learning to predict heart disease, with GBC, KNN and RF showing strong performance.
* Age was identified as a key risk factor.
* Additional features could further improve predictions.
* This work provides a foundation for early heart disease detection tools.