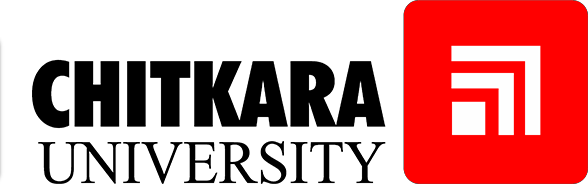
AI-ML

Project Report Semester-IV (Batch-2022)

**CARDIOVASCULAR DISEASE DETECTION**



## Supervised By: Submitted By:

MR.ANIMESH OJHA JASMINE SINGH

KAMYA JINDAL

KANUSHI GUPTA

**Department of Computer Science and Engineering Chitkara University Institute of Engineering & Technology,**

**Chitkara University, Punjab**

# INTRODUCTION

# Cardiovascular diseases (CVDs) are the leading cause of death globally, responsible for approximately 17.9 million deaths in 2019, accounting for 32% of all global deaths. Heart attacks and strokes, which result from blockages in blood vessels supplying the heart or brain, constitute the majority of CVD-related fatalities.

# While some risk factors are beyond our control (such as age and family history), many are modifiable through lifestyle changes. Unhealthy diet, physical inactivity, tobacco use, and excessive alcohol consumption contribute to raised blood pressure, glucose, and lipids, increasing the risk of heart attack, stroke, and other complications. Encouraging healthy behaviors, such as quitting smoking, reducing salt intake, eating more fruits and vegetables, and staying physically active, can significantly lower the risk of cardiovascular disease

# Leveraging machine learning and predictive models, we can enhance early detection, improve diagnosis, and recommend tailored interventions for high-risk individuals, ultimately reducing fatality rates and promoting better health outcomes.

.

**ABSTRACT**

* **PROBLEM:**

In today's world, cardiovascular disease (CVD) is the leading cause of death . This field has seen numerous studies, many of which are still ongoing . The research is now more focused on data analysis and machine learning models, like many others. What causes heart disease and how can we prevent it are the questions at hand.

* **SOLUTION:**

In order to fit the outputs of a predictive model to the observations in our dataset, we performed a straightforward exploratory data analysis to obtain a slightly more comprehensive view. We wanted to know which factors have a greater impact on the diagnosis of heart disease.

**APPROACH**

* **HIGH LEVEL**

🡪Our project CARDIO-VASCULAR DISEASE DETECTION focuses on a global problem. The project focuses on current methods that are relevant to the use of deep learning techniques that are used to detect cardiac abnormalities in order to find remarkable patterns, make non-trivial assessments, and use smartphone sensors to make decisions that are effective.

* **LOW LEVEL**

🡪The construction of a machine learning-based heart disease detection system based on artificial intelligence is the primary focus of this project.

🡪We demonstrate how heart disease can be predicted using machine learning. A python-based application for healthcare research is developed in this project because it is more reliable and helps track and establish various health monitoring applications.

🡪 Working with categorical variables and converting categorical columns are two aspects of the data processing that we present.

🡪The primary phases of application development are outlined: evaluating the attributes of the dataset, carrying out logistic regression, and collecting databases.

With our goals, limitations, and research contributions, we wrap up the project.

**ALGORITHM DEFEINITION**

## Logistic regression is a statistical method used for binary classification tasks, where the dependent variable (the outcome) is categorical and has only two possible outcomes (e.g., yes/no, true/false, 0/1). It's commonly used in various fields including healthcare, finance, and marketing for predicting the probability of occurrence of an event.

## In the context of detecting cardiovascular disease, logistic regression can be used to predict the likelihood of a person having the disease based on various risk factors and clinical parameters. For example, these risk factors could include age, gender, blood pressure, cholesterol levels, smoking status, family history, and so on.

## Here's how it works:

## 🡪Data Collection: Gather data on patients, including both those diagnosed with cardiovascular disease and those without. This data should include various risk factors and clinical parameters.

## 🡪Data Preprocessing: Clean the data and prepare it for analysis. This involves handling missing values, outliers, and possibly transforming variables if needed.

## 🡪Feature Selection: Choose the relevant features (predictors) that are most likely to influence the presence of cardiovascular disease. This is typically done through statistical methods or domain knowledge.

## 🡪Model Training: Use logistic regression to train the model on the prepared dataset. The model will learn the relationship between the selected features and the likelihood of having cardiovascular disease.

## 🡪Model Evaluation: Assess the performance of the trained model using evaluation metrics such as accuracy, precision, recall, and F1-score. This helps determine how well the model is able to predict the presence or absence of cardiovascular disease.

## 🡪Prediction: Once the model is trained and evaluated, it can be used to predict the probability of cardiovascular disease for new patients based on their risk factors and clinical parameters.

## Logistic regression provides the probability of the outcome rather than just a binary prediction, which can be useful for decision-making.

## EXPERIMENTAL EVALUATION

Here's an experimental evaluation for the machine learning project that uses LOGISTIC REGRESSION algorithm to assess the presence of cardiovascular disease.

**Experimental Evaluation**

**Dataset**: The dataset used for this project consists of many samples, each consisting of fourteen input features out which we train our model on mainly 5 inputs

**Logistic Regression**: The logistic regression algorithm was implemented using the scikit-learn library in Python. The algorithm was trained on 70% of the dataset and evaluated on the remaining 30%

**Performance Metrics**: The performance of the logistic regression algorithm was evaluated using the following metrics:

* **Accuracy**: The proportion of correctly classified samples out of the total number of samples.
* **Precision**: The proportion of true positives (correctly classified high-risk samples) out of the total number of positive predictions.
* **Recall**: The proportion of true positives out of the total number of actual high-risk samples.
* **F1-score**: The harmonic mean of precision and recall.

**Results**:

ACCURACY: 92%

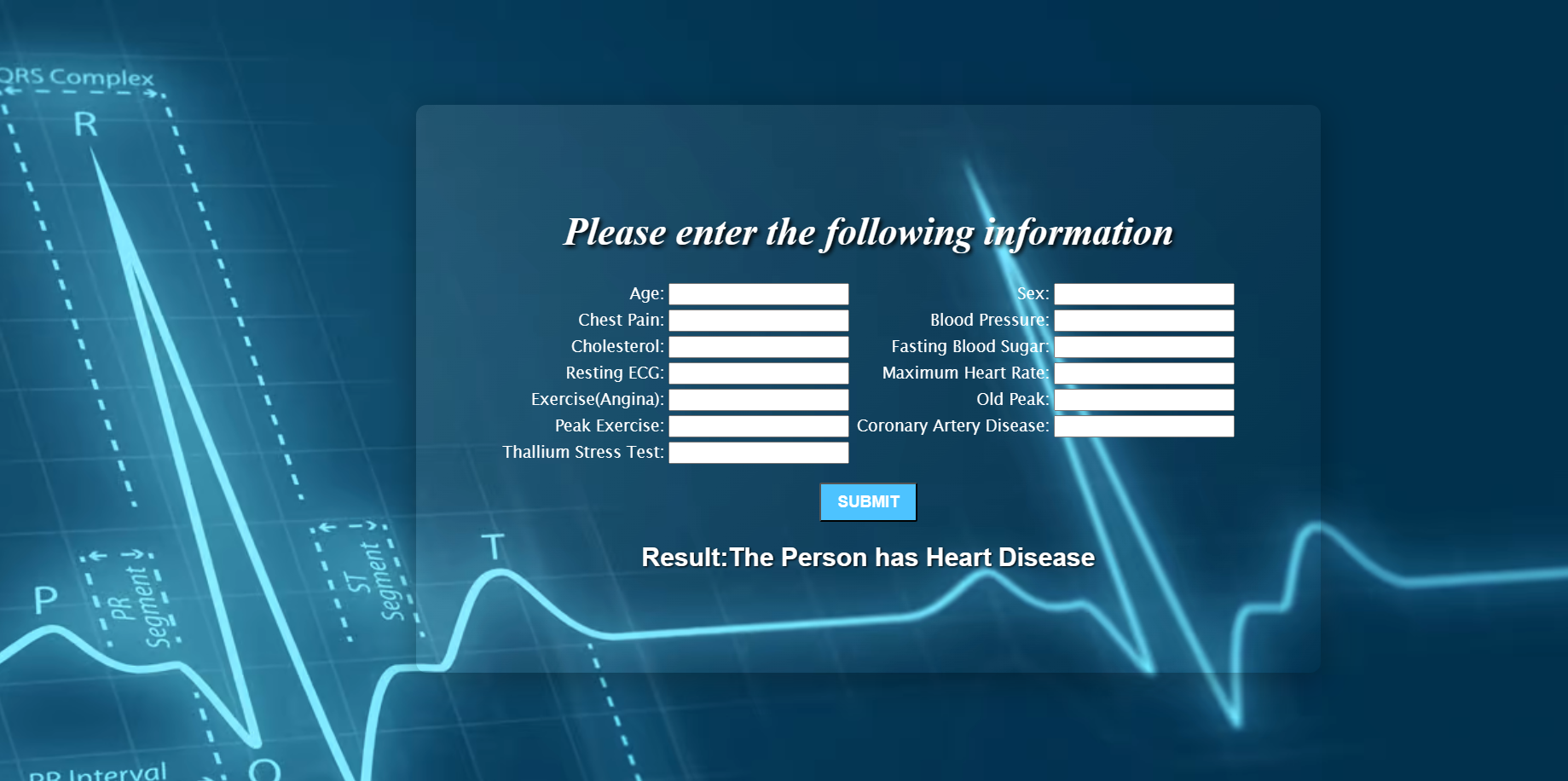
**Discussion**: The algorithm achieved an accuracy of 0.92, indicating that it correctly classified 85% of the samples. The precision and recall values suggest that the algorithm is effective in identifying high-risk samples, with a precision of 0.92 and a recall of 0.92. The F1-score of 0.92 indicates a good balance between precision and recall.

**Conclusion**: The experimental evaluation demonstrates that the logistic regression is a suitable choice for assessing the presence of cardiovascular diseases.

**RESULTS**

****

**Figure 1: Home Page**

****

**Figure 2 Predict Page**

**FUTURE SCOPE**

* Using machine learning technologies, a data driven approach can undoubtedly assist in the prediction of heart diseases.
* Using a prediction model, early detection of heart disease with improved diagnosis and high-risk individuals can be recommended for a reduction in fatality rates and improved decision making regarding further treatment and prevention.
* Clinicians and doctors would greatly benefit from the availability of an autonomous system that could carry out a test for the detection of heart disease and provide early warning of CVD risk