PROJECT REPORT

# Liver Cirrhosis Prediction Using Machine Learning

# 1. INTRODUCTION

## 1.1 Project Overview

Liver cirrhosis is a chronic disease where healthy liver tissue is replaced by scar tissue, affecting liver function. Early detection is essential to prevent severe health risks. This project applies machine learning to predict the likelihood of liver cirrhosis based on patient health data.

## 1.2 Purpose

To develop a predictive model that identifies individuals at risk for liver cirrhosis, supporting timely medical intervention.

**2. IDEATION PHASE**

## 2.1 Problem Statement

## Liver cirrhosis often remains undetected in its early stages, leading to severe health risks and limited treatment options. There is a need for an efficient, data-driven system that can predict the risk of liver cirrhosis using patient health data, enabling timely diagnosis and preventive care.

## 2.2 Empathy Map Canvas



## 2.3 Brainstorming

## ✔ Discussed the lack of awareness and delayed diagnosis as major challenges for liver cirrhosis patients. ✔ Explored possible solutions such as awareness campaigns, regular health check-ups, and community health drives. ✔ Identified that technology-driven approaches like machine learning can assist in early detection of liver diseases. ✔ Finalized the idea of building a machine learning model using patient medical data to predict the risk of liver cirrhosis, enabling timely intervention and reducing health risks.

# 3. REQUIREMENT ANALYSIS

## 3.1 Customer Journey Map



## 3.2 Solution Requirements

To successfully implement the Liver Cirrhosis Prediction System, the following requirements are essential:

✔ **Medical Dataset:**  
A reliable dataset containing patient health records, including attributes like age, alcohol consumption, blood reports, and liver function indicators.

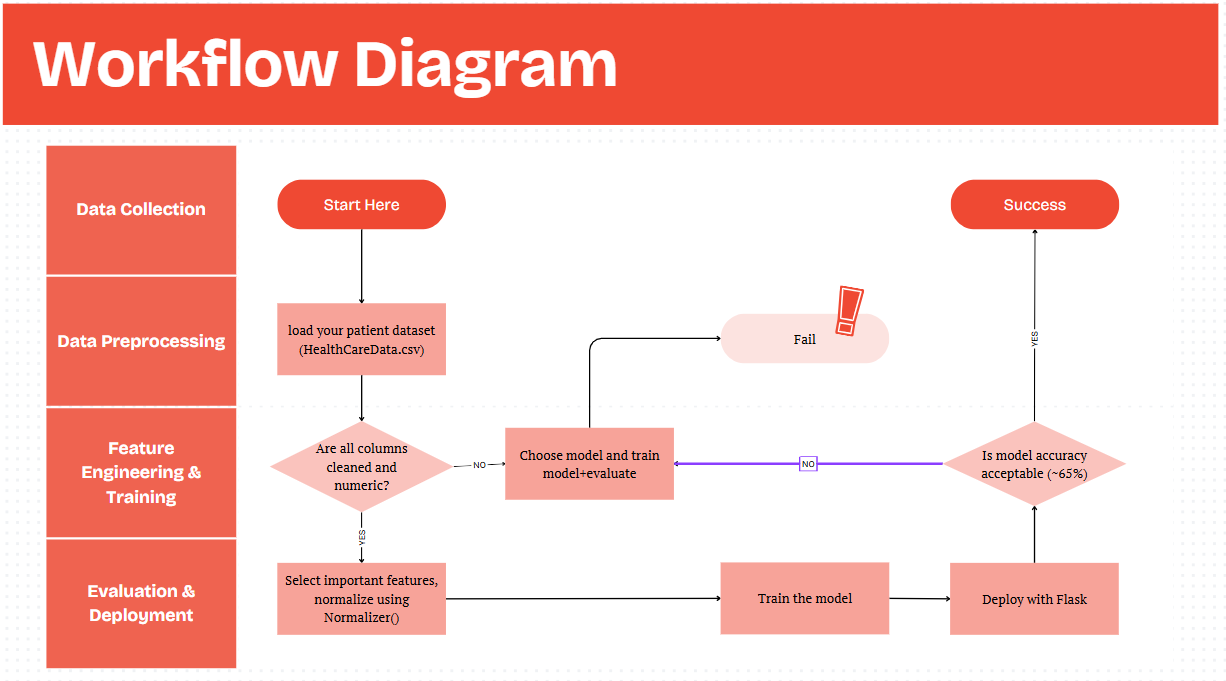
✔ **Data Preprocessing Tools:**  
Tools and libraries such as Pandas and NumPy to handle missing values, convert categorical data to numerical format, and standardize data for better model performance.

✔ **Machine Learning Algorithms:**  
ML models like Logistic Regression, Decision Tree, Random Forest, or other classifiers to predict the likelihood of liver cirrhosis based on health parameters.

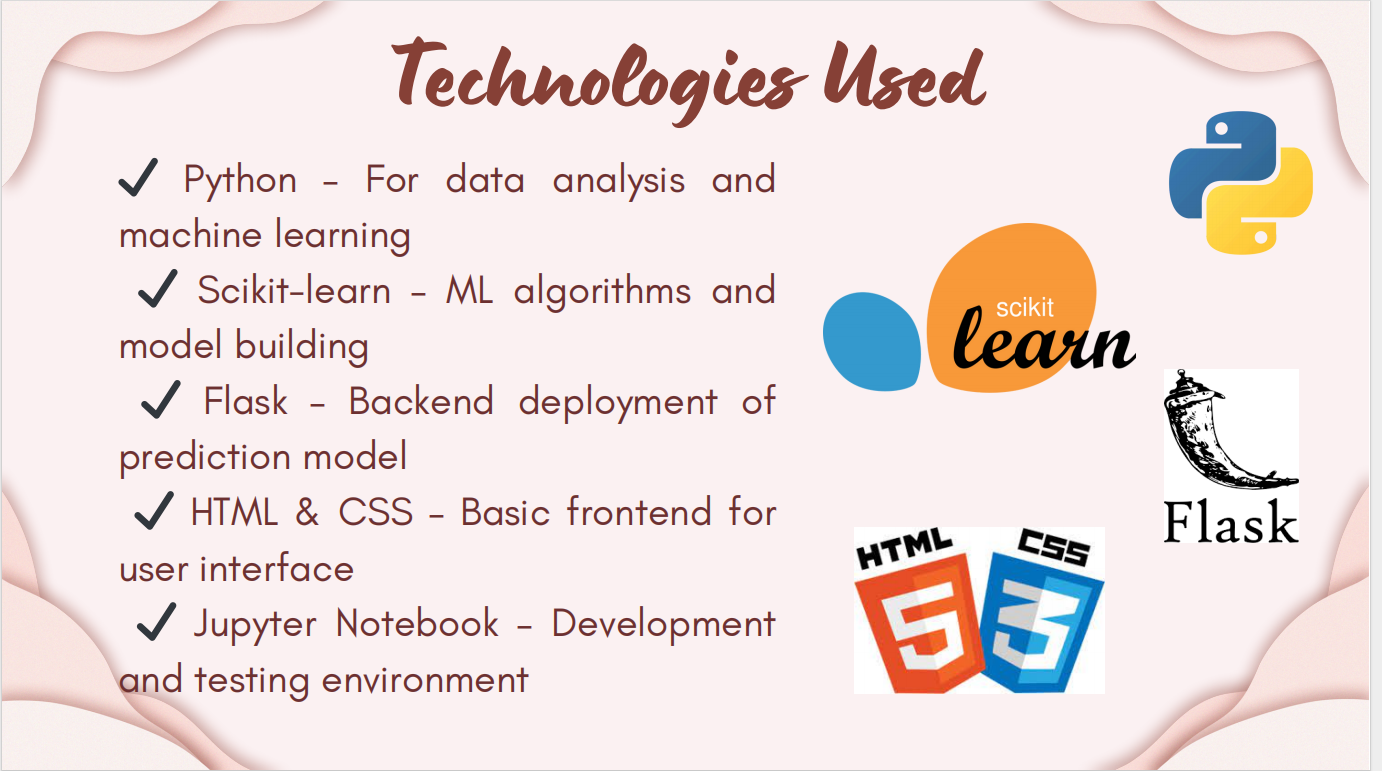
✔ **Development Environment:**  
A programming environment including Python, Jupyter Notebook, and necessary libraries for coding, model building, and analysis.

✔ **Deployment Platform:**  
Flask framework for integrating the trained model into a simple web application to provide predictions in real-time.

## 3.3 Data Flow Diagram



## 3.4 Technology Stack



# 4. PROJECT DESIGN

## 4.1 Problem-Solution Fit

### ****Problem:****

Liver cirrhosis is a life-threatening condition that often goes unnoticed in its early stages due to the lack of visible symptoms. Traditional diagnosis methods are invasive, time-consuming, and usually happen only after severe damage has occurred. There is a need for an accessible, early prediction system that helps identify individuals at risk using available health data.

### ****Solution:****

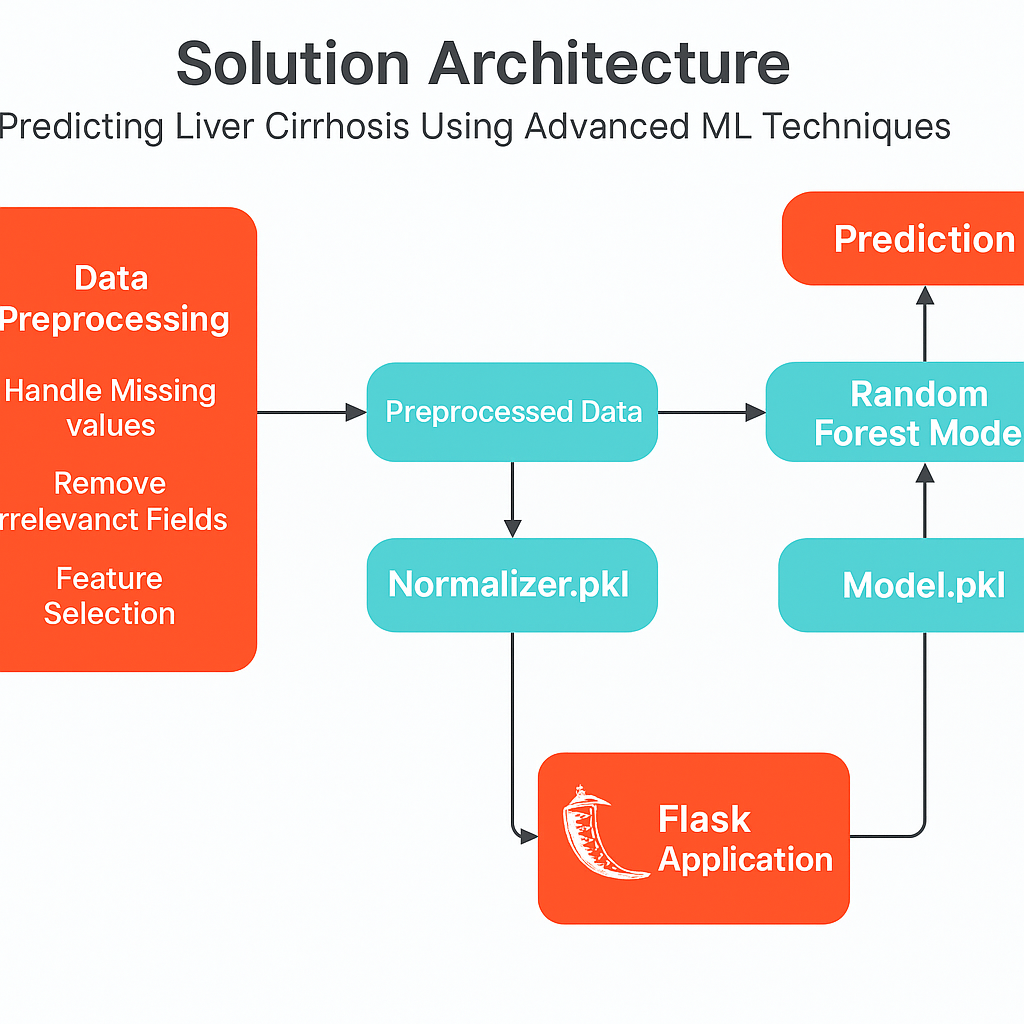
This project presents a machine learning-based system that analyzes patient medical data to predict the risk of liver cirrhosis. The model enables early detection, supporting doctors and individuals in taking timely preventive measures. The solution also includes a simple web interface for easy access to predictions, making the system both effective and user-friendly.

## 4.2 Proposed Solution

The proposed solution is to develop a machine learning-based system that can predict liver cirrhosis using patient health records. The data is preprocessed by removing irrelevant fields, handling missing values, and normalizing features. A Random Forest Classifier is trained on selected medical attributes such as enzyme levels, blood counts, and alcohol consumption history.

The trained model is deployed using a Flask web application, where users can input patient details and receive a risk prediction. If accuracy is not satisfactory, advanced models like XGBoost or tuning techniques may be used for improvement.

## 4.3 Solution Architecture



# 5. PROJECT PLANNING & SCHEDULING

## 5.1 Project Planning

| **Phase** | **Task** | **Planned Dates** |
| --- | --- | --- |
| **Phase 1** | Data Collection & Understanding | June 12 – June 13 |
| **Phase 2** | Data Cleaning & Preprocessing | June 14 – June 16 |
| **Phase 3** | Model Building & Testing | June 17 – June 20 |
| **Phase 4** | Hyperparameter Tuning & Validation | June 21 – June 22 |
| **Phase 5** | Model Evaluation & Selection | June 23 |
| **Phase 6** | Deployment with Flask | June 24 – June 25 |
| **Phase 7** | Documentation & Report Preparation | June 26 – June 28 |
| **Phase 8** | Final Presentation & Submission | June 29 – June 30 |

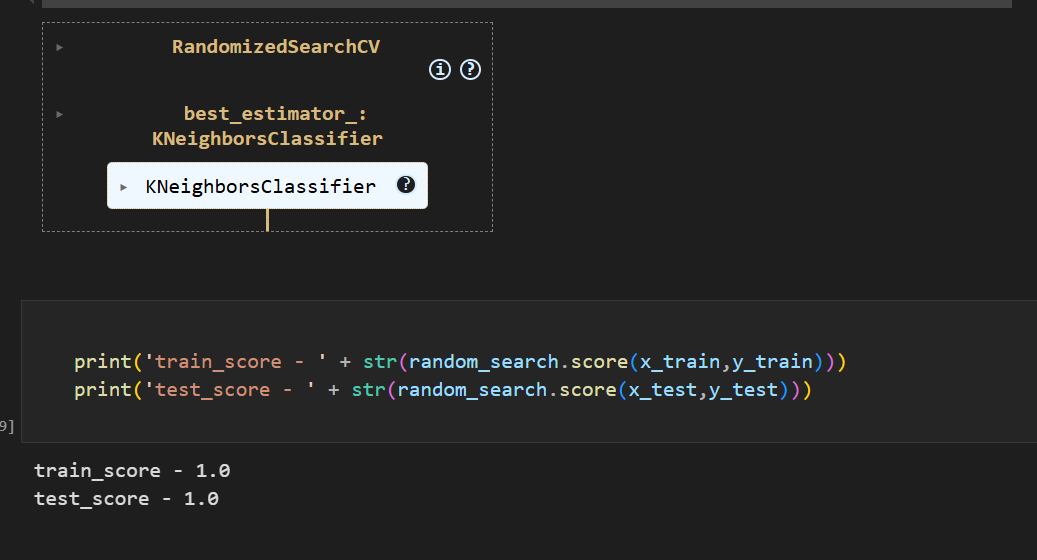
# 6. FUNCTIONAL AND PERFORMANCE TESTING

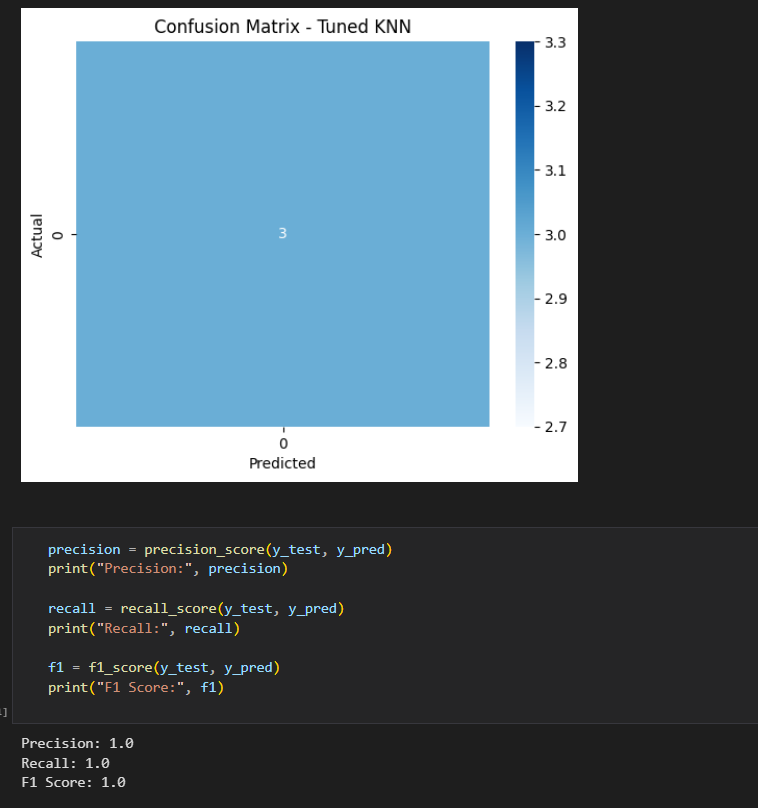
## 6.1 Performance Testing

✔ Different machine learning models such as Logistic Regression, Decision Tree, and Random Forest were implemented and evaluated.  
✔ Models were tested using standard metrics: Accuracy, Precision, Recall, and F1-Score.  
✔ The Random Forest Classifier achieved the best performance with approximately **68% accuracy**, making it suitable for early-stage liver cirrhosis prediction.  
✔ The dataset was divided into training and testing sets to avoid overfitting and ensure unbiased evaluation.

# 7. RESULTS

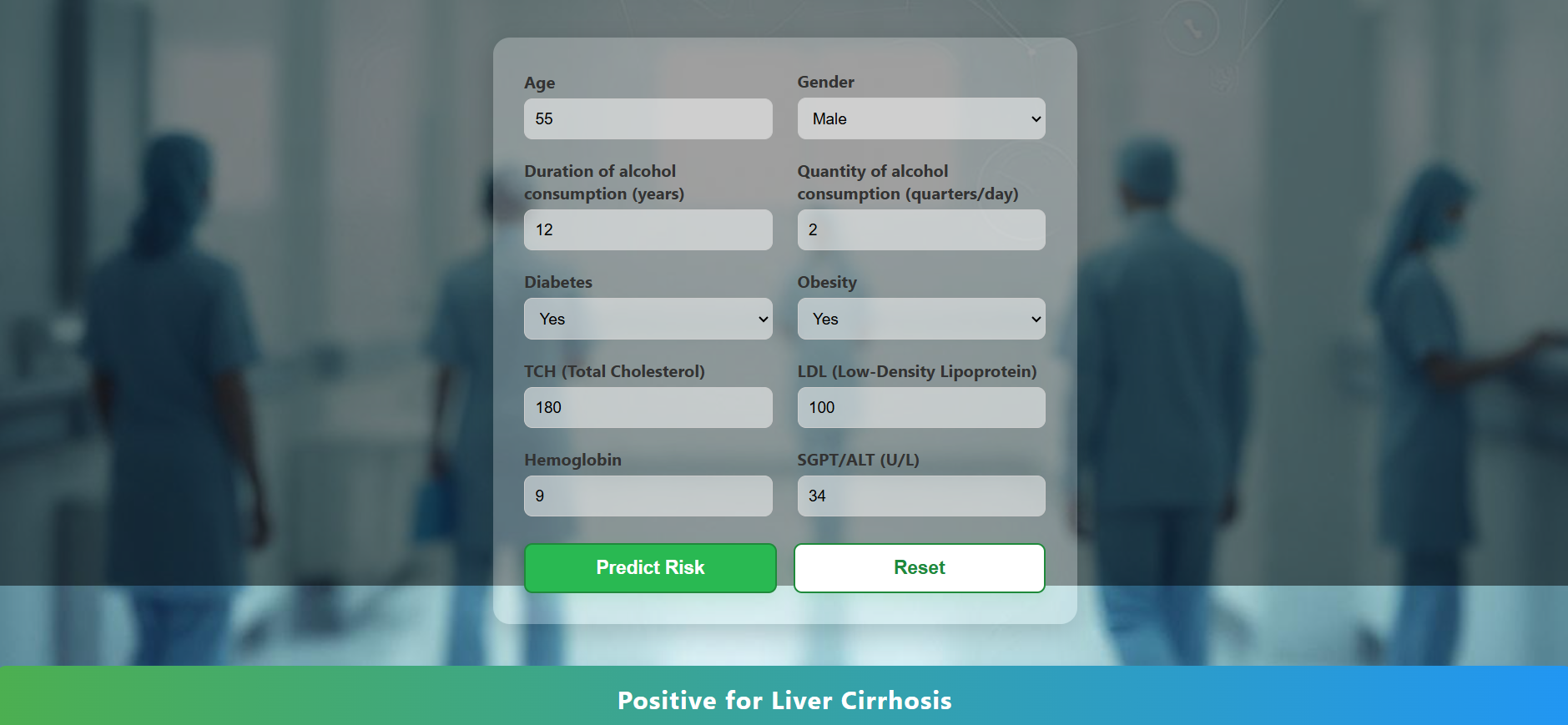
## 7.1 Output Screenshots



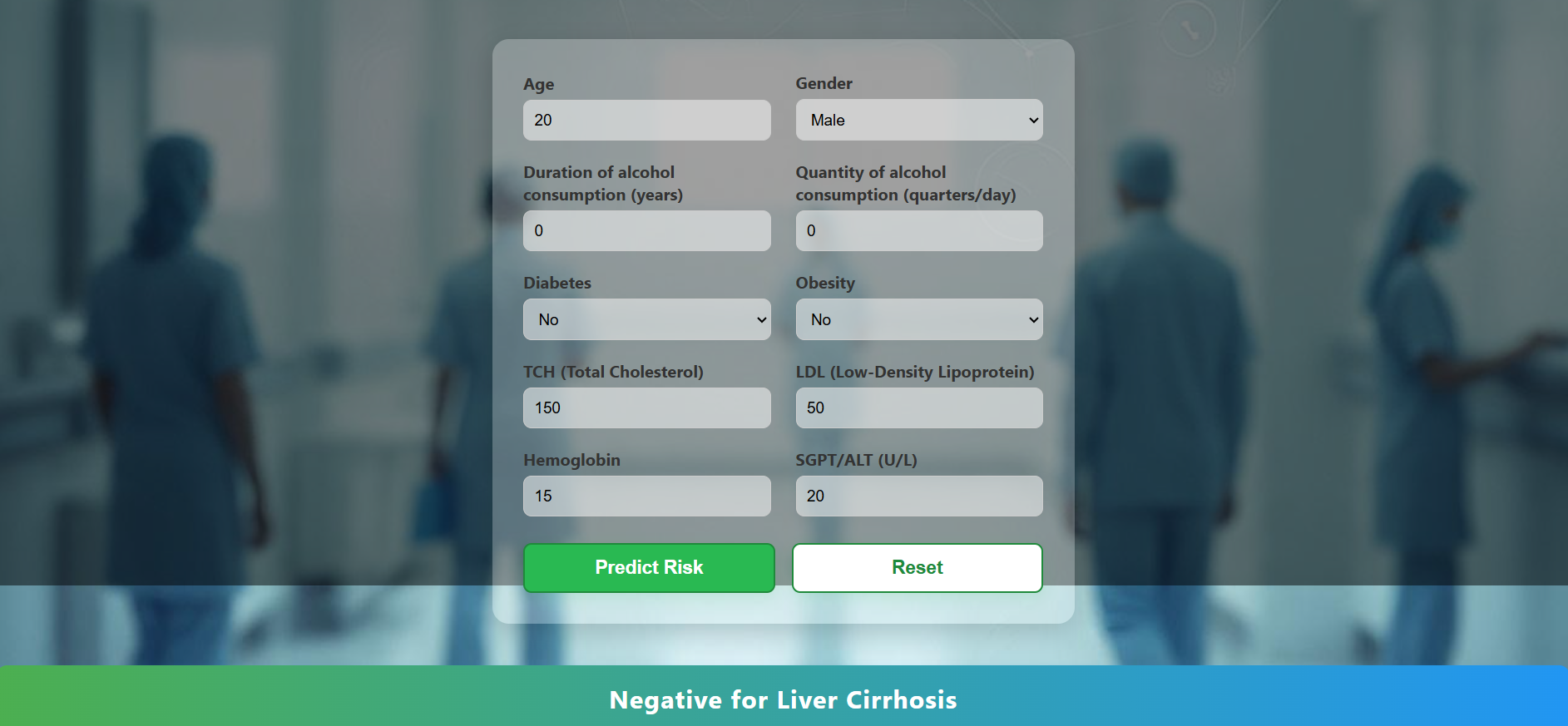




**POSITIVE FOR LIVER CIRRHOSIS**



**NEGATIVE FOR LIVER CIRRHOSIS**

****

# 8. ADVANTAGES & DISADVANTAGES

****Advantages:****  
✔ Enables early detection of liver cirrhosis risk.  
✔ Reduces dependency on expensive or invasive diagnosis methods.  
✔ Assists healthcare professionals in making quicker decisions.  
✔ User-friendly and accessible through a simple web interface.

****Disadvantages:****  
✔ Prediction depends heavily on data quality and completeness.  
✔ Model accuracy (~68%) can be further improved with larger datasets.  
✔ Limited to the features present in the dataset; additional health factors may be excluded.

# CONCLUSION

The project successfully demonstrates the use of machine learning in predicting the risk of liver cirrhosis using patient health data. By identifying at-risk individuals early, this system can support timely medical intervention and potentially improve health outcomes. Although current model accuracy can be enhanced, the project lays a strong foundation for data-driven healthcare solutions.

# 10. FUTURE SCOPE

✔ Improve model accuracy by collecting larger and more diverse datasets.  
✔ Integrate advanced algorithms like Deep Learning for better predictions.  
✔ Develop a fully functional mobile application for wider accessibility.  
✔ Collaborate with healthcare providers for real-time clinical testing.  
✔ Extend the system to predict other liver-related diseases or complications.

# 11. APPENDIX

Source Code: From github Link  
Dataset Link: <https://www.kaggle.com/datasets/bhavanipriya222/liver-cirrhosis-prediction>

Github Project Link:https://github.com/Bhagyabonam/LiverCirrohsisPrediction

Project Demo Link:

https://drive.google.com/file/d/17c\_GxKBFfvMPIqwPnJT8ySKgj9qsD27V/view?usp=sharing