

# Liver Disease Diagnosis

Group 1: The Phoenix

DSCI - 6002-01  
Final Project

---

University of New Haven  
West Haven, CT

# Our Team - The Phoenix



**YASASVI THATI**

Storytelling and  
Research



**MERISHNA SINGH**

Data Modeling and  
Deployment



**VIJAYA BHARGAVI PEDINEEDI**

Preliminary  
Research



**MEDHINI BARLA**

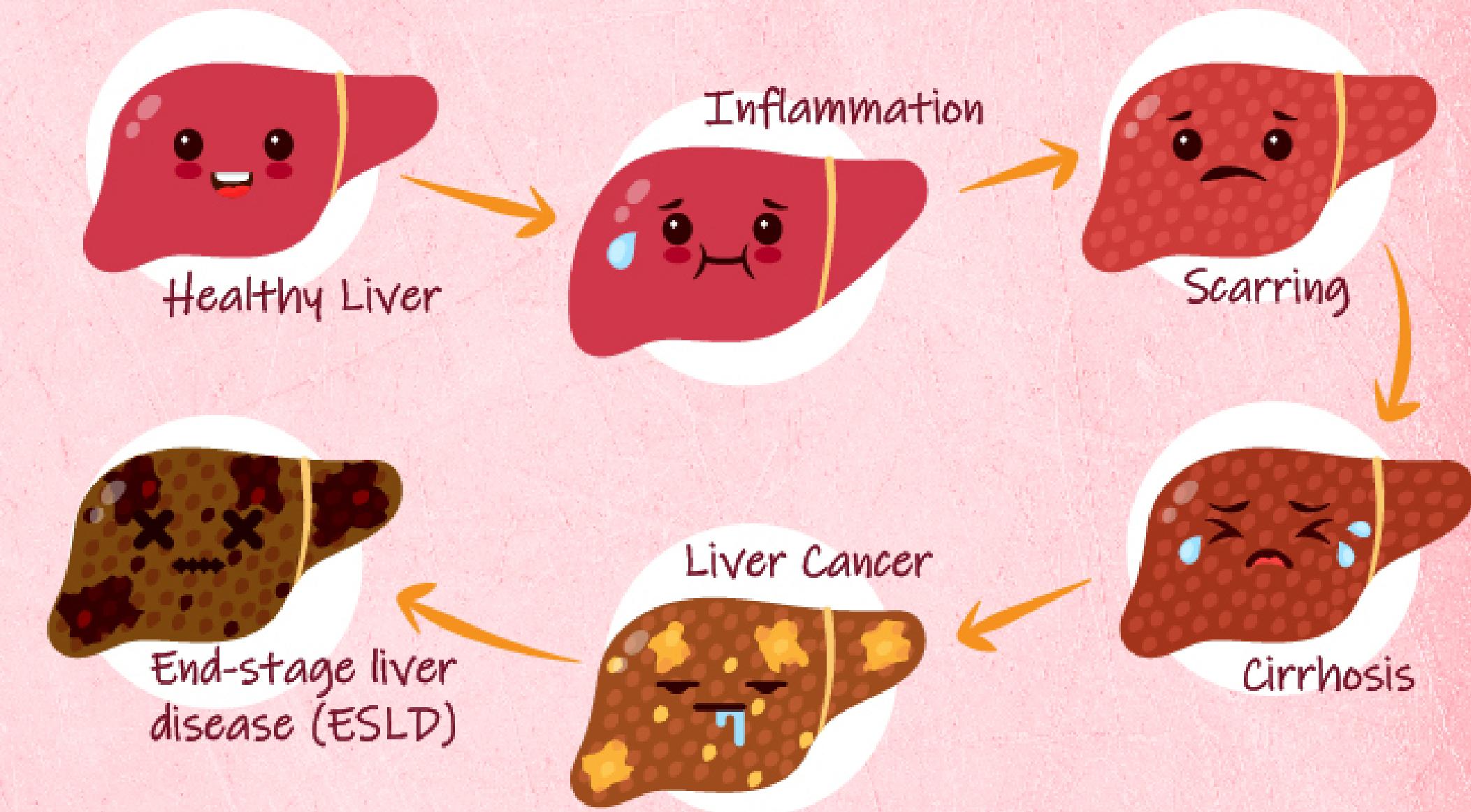
Methodology  
Research



**SRAVYA MEDIKONDA**

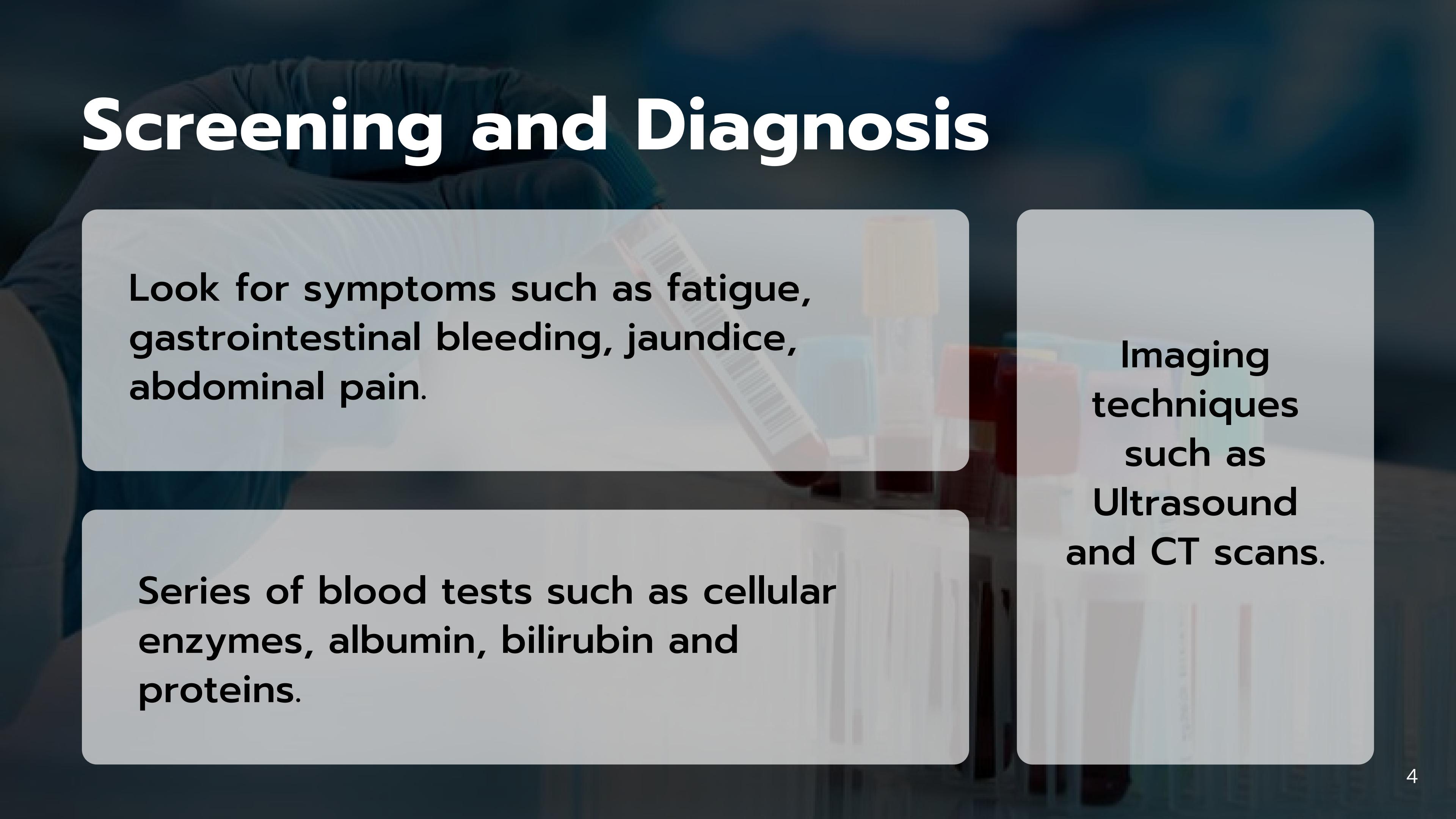
Testing and  
Evaluation

# 5 Stages of Liver Disease



Source: <https://www.relainstitute.com/blog/liver-disease-5-stages-of-progression>

# Screening and Diagnosis



Look for symptoms such as fatigue, gastrointestinal bleeding, jaundice, abdominal pain.

Imaging techniques such as Ultrasound and CT scans.

Series of blood tests such as cellular enzymes, albumin, bilirubin and proteins.

# Problem

- Most patients are asymptomatic.
- Doctors have limited findings suggestive of the presence of the liver disease.

# Objective

- Use blood test results paired with historical patient data and demographics to predict the presence of liver disease.

# Project Plan

Build a liver disease classification model

Data Collection and  
Preprocessing

Data Modeling

Model Evaluation and  
Deployment

# Dataset - Liver Patients

- 583 patients record (416 liver patient, 167 non liver patient)
- Age
- Gender
- Total Bilirubin
- Direct Bilirubin
- Alkaline Phosphotase
- Alamine Aminotransferase
- Aspartate Aminotransferase
- Total Proteins
- Albumin
- A/G Ratio Albumin and Globulin Ratio
- Diagnosis

Source: [UCI- ILPD \(Indian Liver Patient Dataset\)](#)

# Data Preprocessing

- Filled null values with mean of the column
- Label-encoded feature variables
- Feature Selection using Pearson Coefficient
- Feature Scaling
- Oversampling for imbalanced dataset

CSV Dataset

Train

→ 80% Data

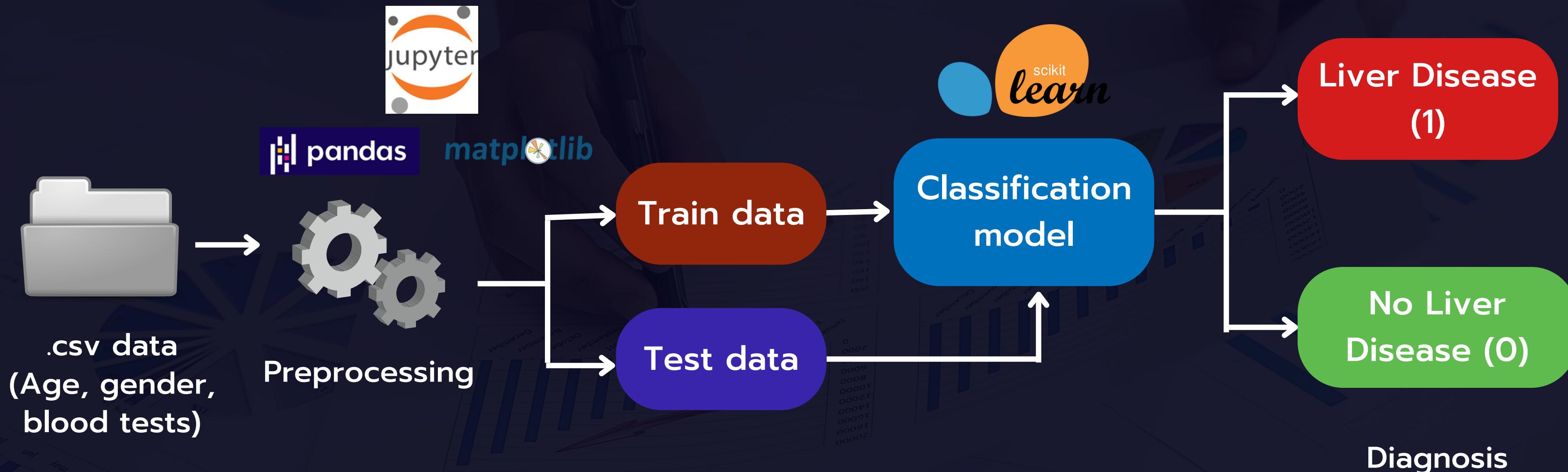
Test

→ 20% Data

# Dataset sample

Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alanine_Aminotransferase	Aspartate_Aminotransferase	Total_Proteins	Albumin	Albumin_and_Globulin_Ratio	Diagnosis
39	1	1.6	0.8	230	88	74	8.0	4.0	1.00	0
46	1	0.8	0.2	160	31	40	7.3	3.8	1.10	1
51	1	0.7	0.1	180	25	27	6.1	3.1	1.00	1
70	1	0.6	0.1	862	76	180	6.3	2.7	0.75	1
66	0	0.7	0.2	162	24	20	6.4	3.2	1.00	0

# Training a classification model



# Liver Disease Classifier

1

Trained and evaluated classification models

2

Deployed the model with highest accuracy

# 1

# Classification Model

## Model

- Logistic Regression
- Support Vector Classifier
- Random Forest Classifier
- K-Nearest Neighbours Classifier

## Accuracy

- 76.57%
- 77.71%
- 76.57%
- 70.28%

2

# Deployment

- Deployed the application using Streamlit
  - Open-source app framework for Machine Learning and Data Science
  - Automatically gets deployed as the GitHub repository is updated.



<https://group-1-the-phoenix-project-group-1-main-anj10e.streamlit.app/>

# Future Improvements

- Collaborate with healthcare professionals for feedback
- Collect more data such as previous medications, alcohol consumption, and family history of liver disease.
- Engineer new features
- Tune hyperparameters for the modeling algorithm

# THANK YOU !

