

Project Report Format

1. INTRODUCTION

1.1 Project Overview

Liver cirrhosis is a critical health concern globally, leading to chronic liver failure and high mortality rates. This project aims to build a machine learning model to predict liver cirrhosis based on various patient features. By analysing these features, the model will classify patients into risk categories, aiding in early diagnosis and treatment.

1.2 Purpose

The purpose of this project is to develop a machine learning-based predictive system that assists in the early detection of liver cirrhosis using a comprehensive set of clinical and lifestyle-related health indicators. Liver cirrhosis is often diagnosed at a late stage due to the lack of noticeable symptoms in the early phases. Delayed detection leads to limited treatment options, higher healthcare costs, and increased emotional distress for patients and their families.

This project aims to:

- Analyze patient health data to identify risk patterns indicative of early stage liver damage.
- Design a user-friendly prediction tool that clinicians and patients can use for proactive screening.
- Empower healthcare providers with reliable, non-invasive decision support for early intervention.
- Raise awareness about the importance of preventive liver care and encourage at-risk individuals to undergo timely checkups.

By combining medical insights with modern machine learning techniques, the project seeks to bridge the gap between clinical diagnosis and preventive care, ultimately improving patient outcomes and reducing the burden of liver disease.

2. IDEATION PHASE

2.1 Problem Statement

2.2 Empathy Map Canvas

2.3 Brainstorming

3. REQUIREMENT ANALYSIS

3.1 Customer Journey map

3.2 Solution Requirement

3.3 Data Flow Diagram

3.4 Technology Stack

4. PROJECT DESIGN

4.1 Problem Solution Fit

4.2 Proposed Solution

- 4.3 Solution Architecture
5. **PROJECT PLANNING & SCHEDULING**
 - 5.1 Project Planning
6. **FUNCTIONAL AND PERFORMANCE TESTING**
 - 6.1 Performance Testing
7. **RESULTS**
 - 7.1 Output Screenshots

The figure consists of two screenshots of a web form titled "Liver Disease Prediction".

The top screenshot shows the form with the following input fields: Age, Gender (Male / Female), Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase, Aspartate Aminotransferase, Aspartate Aminotransferase, Total Protein, Albumin, and Albumin and Globulin Ratio. The bottom screenshot shows the same form with the following values entered: Age: 20, Gender: Male, Total Bilirubin: 0, Direct Bilirubin: 0.2, Alkaline Phosphatase: 210, Aspartate Aminotransferase: 38, Aspartate Aminotransferase: 80, Total Protein: 6.5, Albumin: 3.1, and Albumin and Globulin Ratio: 1.2.

Figure 7.3: Form filled with patient data

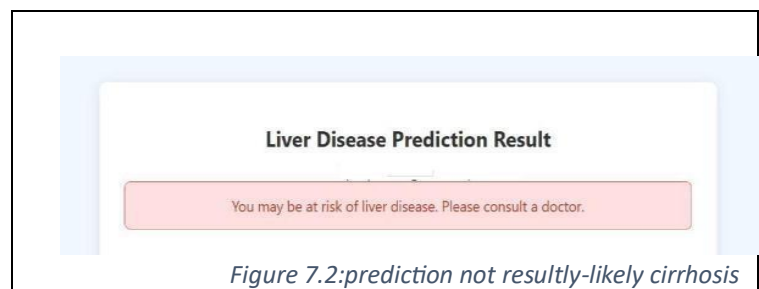


Figure 7.2:prediction not resultly-likely cirrhosis

Figure 7.2 :Prediction Result-Like

```
[ ] dataset.head()
```

	Age	Gender	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphatase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin	Albumin_and_Globulin_Ratio	Data
0	65	Female	0.7	0.1	187		16	18	6.8	3.3	0.90
1	62	Male	10.9	5.5	699	64		100	7.5	3.2	0.74
2	62	Male	7.3	4.1	490	60		68	7.0	3.3	0.89
3	58	Male	1.0	0.4	182	14		20	6.8	3.4	1.00
4	72	Male	3.9	2.0	195	27		59	7.3	2.4	0.40

Figure 7.4

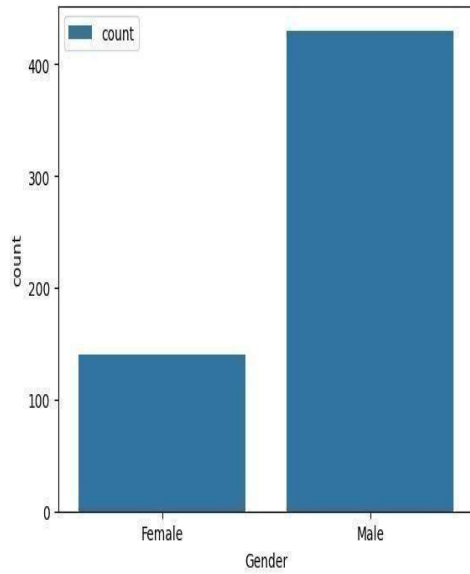


Figure 7.5

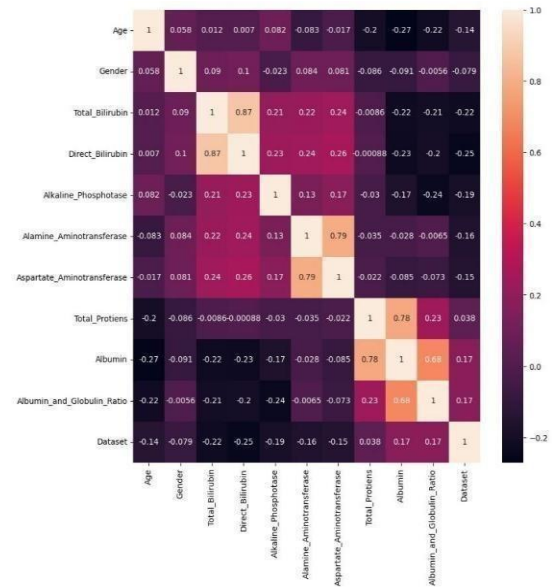


Figure 7.6

8. ADVANTAGES & DISADVANTAGES

8.1 Advantages

- **Early Detection of Liver Cirrhosis**

The system enables identification of liver cirrhosis in its early stages, which is crucial for timely intervention and improved treatment outcomes.

- **Non-Invasive Assessment**

The prediction model uses existing clinical and health-related parameters, eliminating the need for painful or invasive diagnostic procedures like liver biopsies.

- **Personalized Risk Evaluation**

The machine learning model provides risk predictions tailored to individual patient profiles, making the outcomes more relevant and actionable.

- **Time and Cost Efficiency**

The tool can generate results rapidly and reduce the financial burden associated with repeated diagnostics and hospital visits.

8.2 Disadvantages

- **Dependence on Data Quality**

The accuracy of the prediction model is highly dependent on the quality and consistency of the input data. Poor data may lead to inaccurate predictions.

- **Potential for Misclassification**

False positives may cause unnecessary concern, while false negatives could result in missed opportunities for early treatment.

- **Lack of Model Interpretability**

Some machine learning algorithms may produce results that are difficult for users to interpret, which could reduce trust in the system.

- **Data Privacy and Ethical Concerns**

Handling sensitive patient data requires strict adherence to privacy policies and ethical guidelines, including compliance with healthcare regulations.

9. CONCLUSION

The project “*Revolutionizing Liver Care: Predicting Liver Cirrhosis using Advanced Machine Learning Techniques*” demonstrates the transformative potential of artificial intelligence in healthcare, specifically in the early prediction and prevention of liver cirrhosis. By utilizing a rich set of clinical and lifestyle-related features, the machine learning model developed in this project aims to detect cirrhosis risk before symptoms become critical—offering a timely and non-invasive approach to liver health management.

In conclusion, this work bridges the gap between medical data and actionable outcomes by combining machine learning with a user-centric interface. It contributes to the vision of smarter, preventive healthcare systems and reinforces the role of technology in enabling better, faster, and more accessible liver disease diagnosis.

10. FUTURE SCOPE

While the current project successfully demonstrates the feasibility and effectiveness of using machine learning for early prediction of liver cirrhosis, there are several areas where the system can be further expanded and improved to maximize clinical impact and real-world usability:

1. Integration with Electronic Health Records (EHRs)

The model can be integrated with hospital EHR systems to automate predictions during routine check-ups, allowing real-time risk assessments without manual data input.

2. Incorporation of Real-Time Monitoring

Future versions can include live health tracking data (e.g., wearable devices, smart sensors) for more dynamic and personalized predictions.

3. Explainable AI (XAI)

To increase clinical trust and transparency, integrating explainable AI techniques would help healthcare professionals understand how the model makes predictions.

4. Mobile Application Deployment

Building a mobile app version of the tool will improve accessibility for patients, allowing them to self-screen and track their liver health regularly.

5. Support for Multi-Language and Regional Adaptation

Adapting the interface and instructions to multiple languages and cultural contexts can improve usability in rural and diverse population settings.

11. APPENDIX

Source Code

This appendix contains key parts of the source code used in the project "Revolutionizing Liver Care: Predicting Liver Cirrhosis using Advanced Machine Learning Techniques". It includes the backend logic, model training script, and frontend form design.

A. app.py – Flask Web Application

This script handles the core backend logic including:

- Route definitions and handling form input
- Loading the trained model, scaler, and label encoders
- Handling form input
- Preprocessing and prediction

```

from google.colab import files
uploaded = files.upload()

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

dataset = pd.read_csv('indian_liver_patient.csv')
print(dataset.head())

print("Initial shape:", dataset.shape)
print("Duplicates:", dataset.duplicated().sum())
dataset = dataset.drop_duplicates()
print("Shape after dropping duplicates:", dataset.shape)

# Check missing values
print(dataset.isna().sum())

# Fill missing Albumin_and_Globulin_Ratio
sns.boxplot(data=dataset, x='Albumin_and_Globulin_Ratio')
plt.show()

median_val = dataset['Albumin_and_Globulin_Ratio'].median()
dataset['Albumin_and_Globulin_Ratio'] = dataset['Albumin_and_Globulin_Ratio'].fillna(median_val)
print(dataset.isna().sum())

[ ] # Gender count and encoding
sns.countplot(data=dataset, x='Gender', label='count')
plt.show()

Male, Female = dataset['Gender'].value_counts()
print('Number of patients that are male: ', Male)
print('Number of patients that are females: ', Female)

# Convert Gender to 0 and 1
dataset['Gender'] = dataset['Gender'].map(lambda x: 1 if x == 'Male' else 0)

# Convert Dataset (label) so that 1 -> 0 and 2 -> 1
dataset['Dataset'] = dataset['Dataset'].map(lambda x: 0 if x == 1 else 1)

# Correlation heatmap
plt.figure(figsize=(10,10))
sns.heatmap(dataset.corr(), annot=True)
plt.show()

# Features and label
x = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values

# Split into training and testing
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25, random_state=42)

# Scaling
from sklearn.preprocessing import StandardScaler

```

B. train_model.py – Model Training Scrip: This script handles dataset cleaning, label encoding, feature normalization, model training (using Random Forest), and saving the model and tools.

```
[ ] y = dataset.iloc[:, -1].values

# Split into training and testing
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25, random_state=42)

# Scaling
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
x_train = sc.fit_transform(x_train)
x_test = sc.transform(x_test)

# Logistic Regression
from sklearn.linear_model import LogisticRegression
log_classifier = LogisticRegression(random_state=0)
log_classifier.fit(x_train, y_train)

# Prediction and evaluation
log_y_pred = log_classifier.predict(x_test)

from sklearn.metrics import confusion_matrix
log_cm = confusion_matrix(y_test, log_y_pred)

sns.heatmap(log_cm, annot=True)
plt.title("Confusion Matrix - Logistic Regression")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()
```

Dataset Link: <https://www.kaggle.com/uciml/indian-liver-patient-records>

GitHub Link: https://github.com/HarshithaGownipalli/liver_cirrhosis_ml

ProjectDemoLink:

<https://drive.google.com/file/d/1n4Th3QSOjEk0tfPgXQxKKk1An20x38ai/view?usp=sharing>