FINAL PROJECT REPORT

Project Title: Revolutionizing Liver Care:
Predicting liver cirrhosis using Advance Machine Learning

Team Information

Team ID	**LTVIP2025TMID40744**	Email ID
Team Size	**4**	
Team Leader	**Gs Farhana**	gsfarhana20@gmail.com
Team Member	**C Supriya**	Supriyareddy228@gmail.com
Team Member	**C Sujitha**	Csujitha04@gmail.com
Team Member	**Amarnath**	Amarsony969@gmail.com

1. INTRODUCTION

Building an Advanced ML-Powered System for Predicting Liver Cirrhosis

1.1 Project Overview – Transforming Liver Care through Predictive Technology

In an era of increasing healthcare demands and the need for **early disease detection**, liver cirrhosis continues to be a **critical health issue** often diagnosed too late for effective intervention. Many healthcare providers still depend on **manual, subjective assessments** and **disconnected patient records**, which lead to delayed diagnosis, inconsistent monitoring, and reactive rather than preventive care.

These challenges include:

Fragmented patient data management

Inconsistent risk assessments

Lack of centralized, predictive reporting systems

To address these gaps, this project introduces a **custom-built, machine learning-powered liver cirrhosis prediction system.** Designed specifically for healthcare environments, this solution:

- **Digitizes clinical workflows**
- **Automates risk prediction**
- **Centralizes patient data and clinical test results**

The system leverages:

- **Advanced machine learning models** for accurate risk scoring
- **Real-time clinical validation and automated data processing**
- **Interactive dashboards and predictive reports** to support timely clinical decisions

By combining **data-driven intelligence with practical clinical usability**, this solution aims to create a **scalable, maintainable, and impactful system** that revolutionizes liver disease management and improves patient outcomes.

Key Features of the Solution:

- Custom Clinical Data Model supporting patient profiles, clinical tests, risk factors, and prediction scores

- Automated Risk Prediction using advanced machine learning models for real-time risk assessment

- Formula-based Fields for dynamic calculations such as patient age, BMI, and derived risk scores

- Interactive Dashboards and Reports for clinicians to monitor patient trends, high-risk cases, and disease progression

1.2 Purpose – Solving Clinical Gaps in Liver Disease Management

The purpose of this project is to address the real-world clinical challenges in liver disease management caused by **manual processes, fragmented data, and delayed diagnosis.** By developing a **centralized, machine learning-driven liver cirrhosis prediction system**, this project aims to:

- **Streamline clinical workflows** through real-time data processing and automation
- **Improve data accuracy** using built-in validation rules and structured clinical relationships
- **Enable proactive, informed clinical decision-making** through dynamic, risk-based reporting tools
- **Create a scalable, user-friendly system** that can evolve with future healthcare advancements and integrations

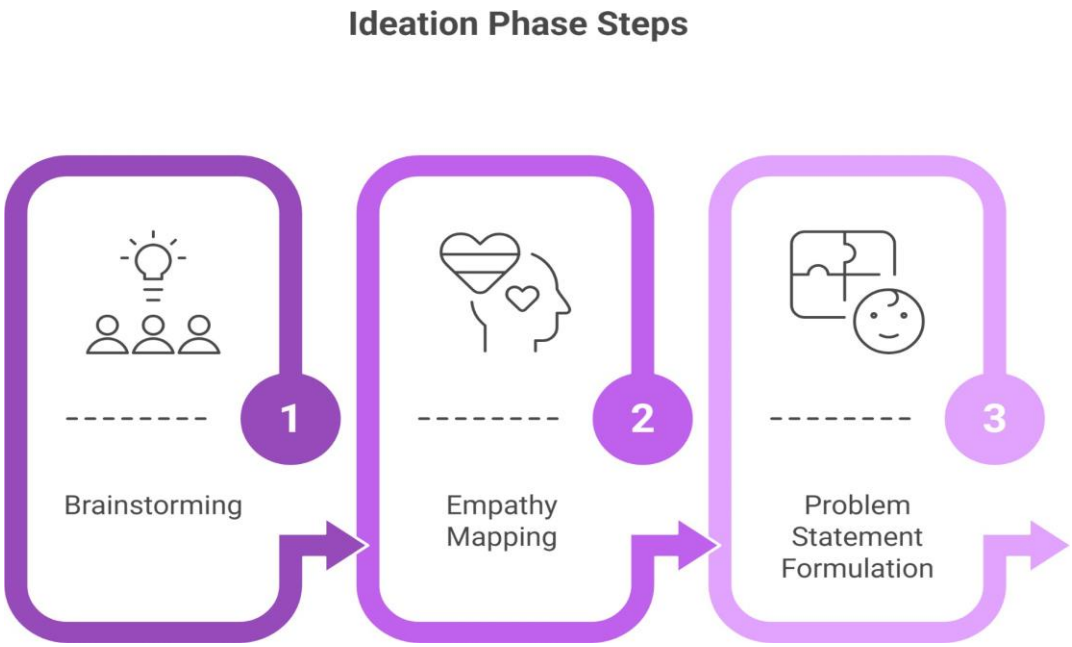
2. Ideation Phase

The Ideation Phase serves as the foundation of any successful project. It blends creativity, user empathy, and structured thinking to identify the core problem, generate meaningful ideas, and prioritize solutions that bring value to users. Where creativity and structured thinking combine to find meaningful and impactful documentation

In our project titled: **“**Revolutionizing Liver Care: Predicting liver cirrhosis**”**. This project aims to develop a predictive model for the early detection and prognosis of liver cirrhosis using machine learning techniques.

The ideation phase included three main steps:

1. Brainstorming
2. Empathy Mapping
3. Problem Statement Formulation



1. 🔹 Brainstorming & Idea Prioritization Template

Step 1: Team Gathering, Collaboration, and Selecting the Problem Statement

Our team convened with the goal of identifying inefficiencies in current liver disease diagnostics and proposing a tech-driven predictive healthcare solution.
Through collaborative meetings, medical research reviews, online whiteboards, and patient journey mapping, we collectively explored pain points faced by hepatologisits, lab technicians, and healthcare administrators. We reviewed real-world hospital practices and identified that most liver care systems still rely heavily on manual workflows for managing:

- Liver function test data interpretation
- Patient diagnosis and staging of cirrhosis
- Manual risk score calculations
- Periodic performance tracking of liver care programs

After several discussions, we clearly defined the core issue:

Problem Statement:

"Liver cirrhosis is often diagnosed late due to manual and inconsistent analysis of liver test results.
Healthcare systems lack predictive tools and centralized data, leading to delayed treatment and poor outcomes.
An automated, intelligent system is needed to assess liver health and predict cirrhosis risk in real time..

Step 2: Brainstorm, Idea Listing, and Grouping

We conducted a collaborative brainstorming session using a digital board, where each team member contributed ideas focused on enhancing liver care diagnostics. These ideas were grouped under key healthcare technology themes:

- **Data Management**: centralized storage of patient profiles, liver function test results, and diagnostic history
- **Automation**: real-time risk score calculation, alert generation for critical cases
- **Reporting**: visual dashboards showing liver health trends, patient risk levels, and diagnosis distribution
- **Validation & Access Control**: ensuring data accuracy with rule-based checks and role-based access for doctors and technicians

From around 25–30 ideas, we grouped and shortlisted the ones that aligned directly with operational efficiency.

Step 3: Idea Prioritization

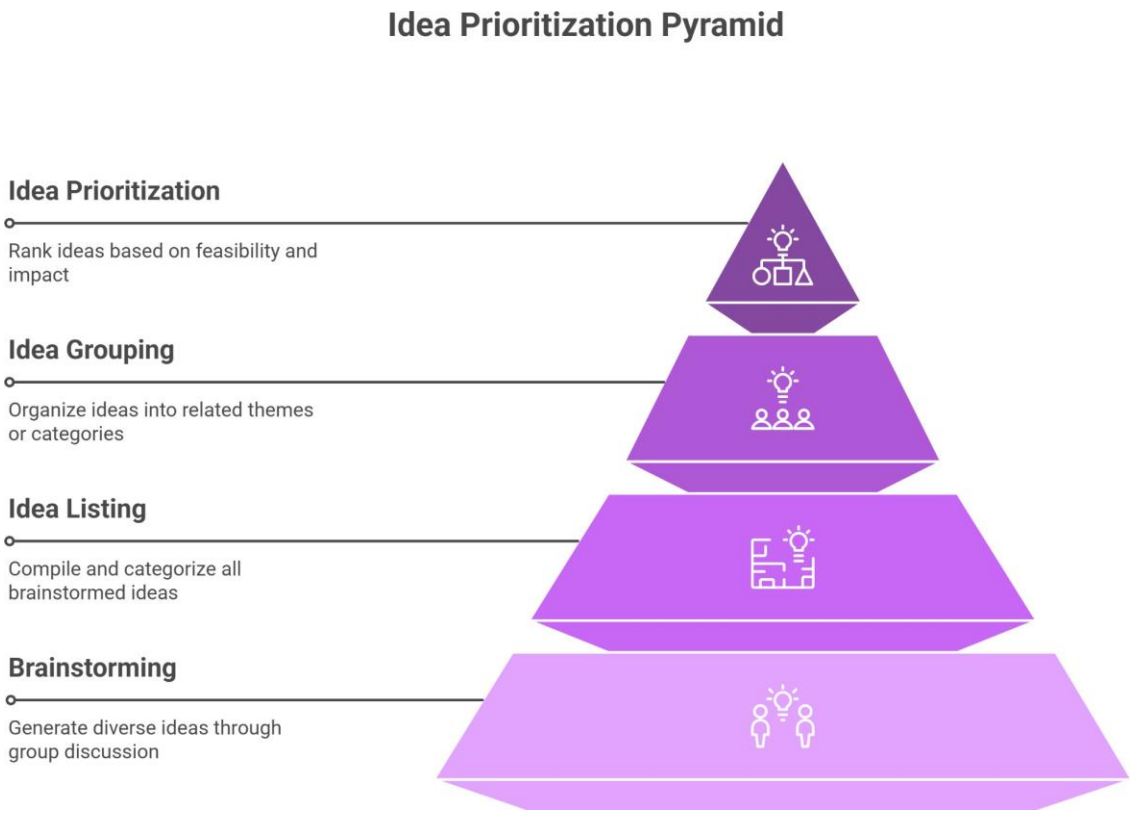
Each grouped idea was evaluated on:

- **Feasibility:** How easily the solution could be implemented using healthcare technology tools
- **Impact:** The value the feature would bring to early detection and patient outcomes
- **Urgency:** Whether the solution addressed a critical gap in current liver care workflows

We developed a decision matrix to help identify a Minimum Viable Product (MVP) that could provide immediate clinical benefit and support scalable development.

- **Top Priority Features**:
    - Automated liver test data updates
    - Role-based medical assignments
    - Trigger-based alerts for critical cases
    - Real-time summary dashboards (Visual summaries of liver health trends, patient risk levels)
    - Controlled field dependencies

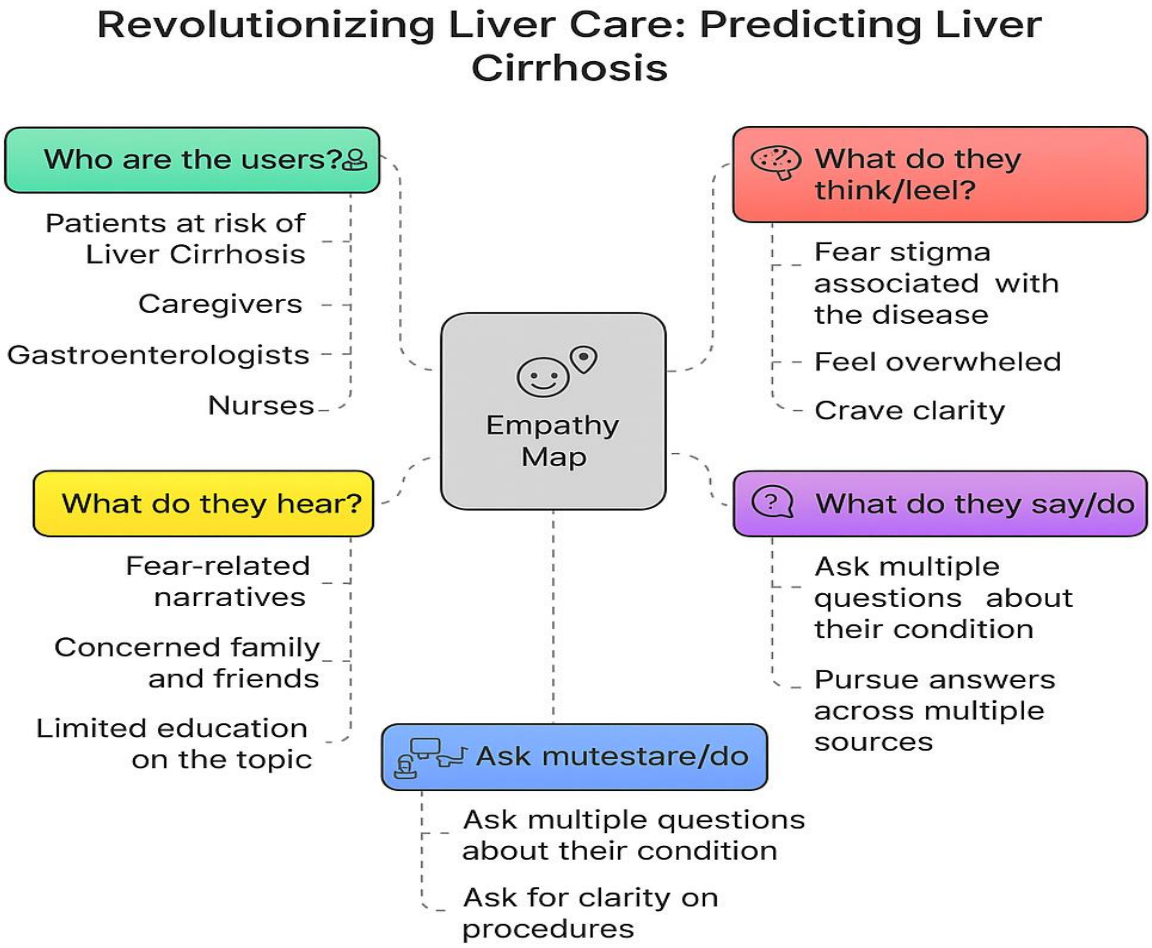
These features formed the scope of our system design in the subsequent development phases.



🔹 Empathy Mapping- Empathize & Discover

Empathy Map Canvas

An empathy map is a visual tool that helps teams deeply understand their users' experiences, pains, and expectations. We used it to map the daily challenges of liver care stakeholders, including hepatologists, lab technicians, healthcare admins and patients.



By stepping into the user’s shoes, we ensured that our Salesforce CRM features (formulas, flows, triggers, dashboards) directly addressed their key frustrations.

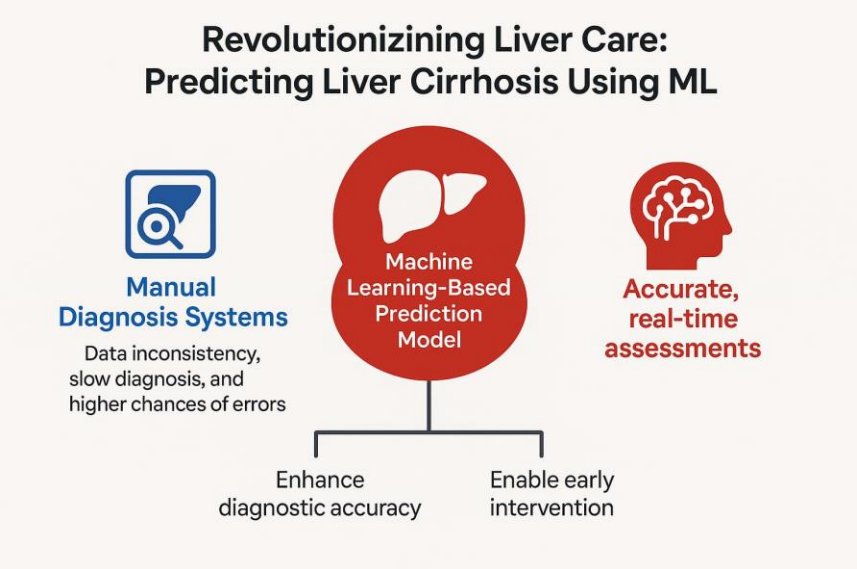
🔹 Define the Problem Statements

Customer Problem Statement Template

To build a successful solution, it's essential to clearly identify the core challenges user face. This ensures the system is focused on real-world medical needs rather than just technical implementation.

Final Customer Problem Statement:

Healthcare providers manage liver disease diagnosis through fragmented records and manual interpretation of liver function tests. This results in delayed diagnoses, increased risk to patients, and inconsistent clinical decisions. A centralized, intelligent system can digitize patient workflows, ensure accurate risk assessment, and deliver timely insights through predictive scoring, dashboards, and automation. This aligns with clinical expectations and informs the system's design, validations, and reporting capabilities.



Requirement Analysis

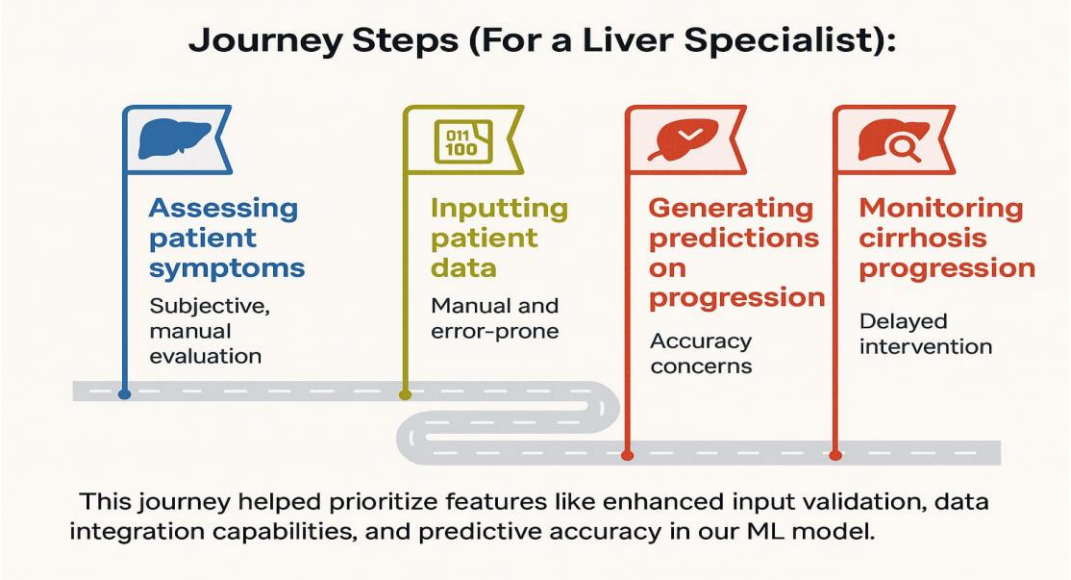
The Requirement Analysis Phase focuses on identifying, organizing, and validating the core needs that the liver care prediction system must address. It ensures that the solution is not only medically accurate but also aligned with real-world clinical workflows and stakeholder expectations—such as doctors, lab technicians, and health administrators.

In our project, “Revolutionizing Liver Care: Predicting Liver Cirrhosis,” this phase bridged the gap between recognizing medical challenges and designing a technical solution using patient journeys, data flows, and system requirements. Where user-centered planning meets clinical clarity to transform a healthcare need into a buildable, intelligent solution..

Customer Journey Map-Understanding User Experience Flow

Purpose: The Patient Journey Map illustrates how liver care stakeholders (e.g., hepatologists, lab technicians, administrators) interact with the system throughout the diagnosis and treatment process. It highlights key actions, pain points, and improvement areas, guiding system design from a real-world medical perspective.

Journey Steps (For a Hepatology Clinician):

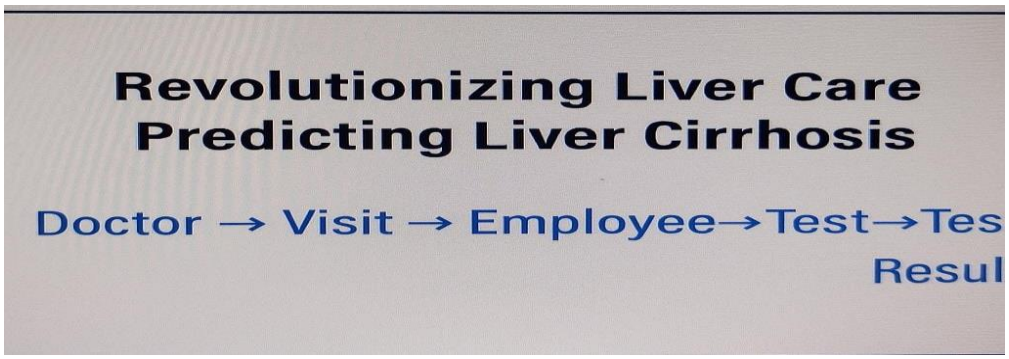


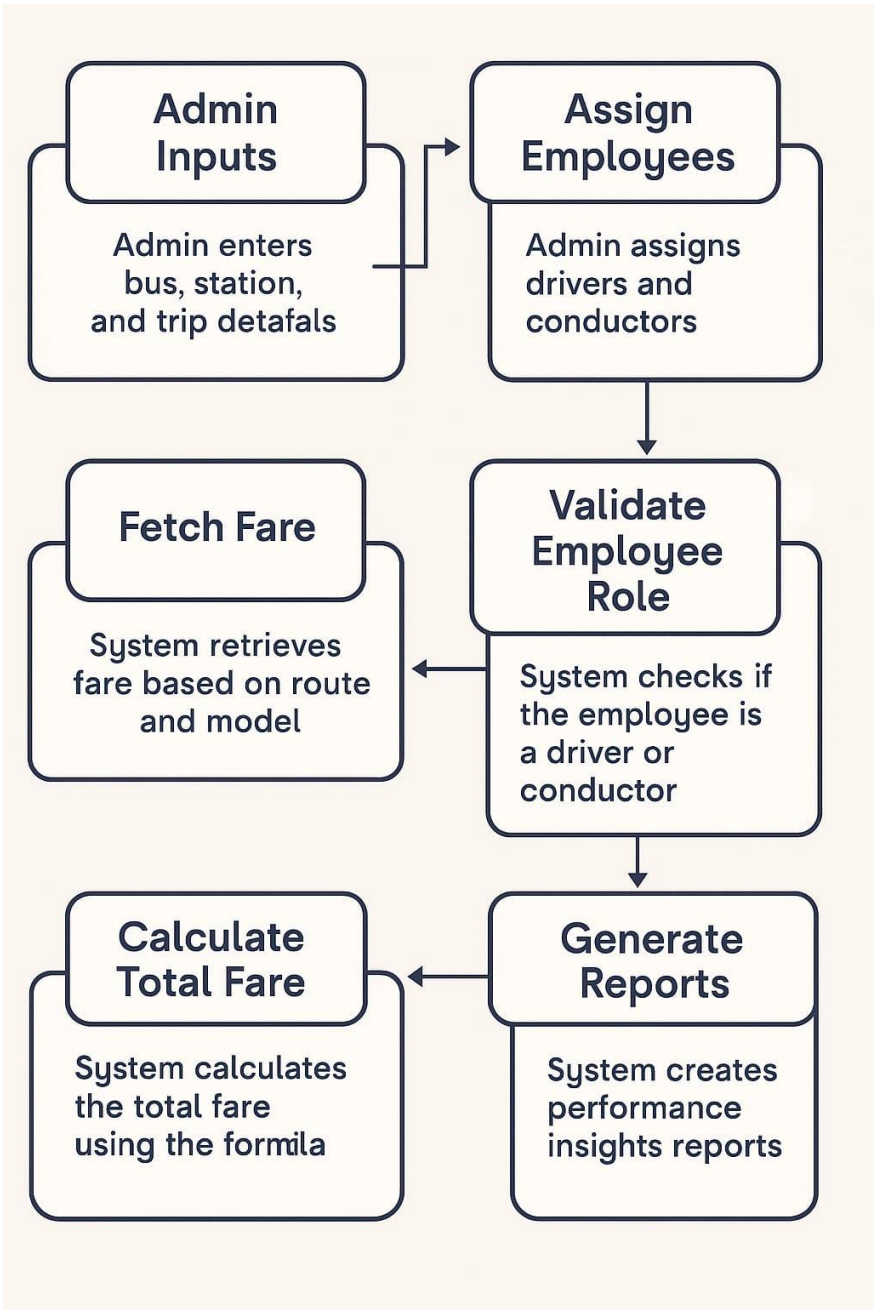
🔹 Data Flow Diagram

Purpose: Mapping Information Flow in Liver Cirrhosis Prediction System
The Data Flow Diagram (DFD) models how patient data and clinical information move through the liver cirrhosis prediction system. It structures the relationship between key components:
● Patient Records, Clinical Tests, Risk Factors, Prediction Engine, and Medical

Level 1 DFD Overview:

1. **Healthcare Provide Inputs**:
    - Adds patient demographics, medical history, lifestyle factors.
    - Inputs clinical test results (e.g., liver enzymes, bilirubin levels, imaging results).
2. **System Logic**:
    - Validates data completeness and test ranges.
    - Analyzes risk factors using a trained machine learning model.
    - Predicts liver cirrhosis risk based on patient profile and clinical indicators.
3. **Outputs**:
    - Generates risk source and predictive reports
    - Displays error messages for missing or invalid input data.




Admin Inputs
Admin enters bus, station, and trip detafals
Assign Employees
Admin assigns drivers and conductors
Validate Employee Role
System checks if the employee is a driver or conductor
Fetch Fare
System retrieves fare based on route and model
Calculate Total Fare
System calculates the total fare using the formila
Generate Reports
System creates performance insights reports

🔹 Solution Requirements

What the Liver Cirrhosis Prediction System System Must Do

Functional Requirements:

- Create custom data entities for:
  - Patient Profile, Clinical Test Results, Risk Factors, Prediction Scores
- Establish data relationships using lookup fields and clinical formula calculations.
- Validate clinical data entries and test ranges using automated logic.
- Automate risk prediction using machine learning workflows.
- Support real-time dashboards and detailed risk assessment reports for clinicians.

Non-Functional Requirements:

- Provide a **user-friendly healthcare interface** for easy data entry and monitoring.
- Enable **real-time clinical data validation** to ensure accuracy and completeness.
- Maintain a **centralized, secure patient database** with well-structured relationships.

🔹 Technology Stack

Tools & Platforms Used for Liver Cirrhosis Prediction System

Category	Technology Used	Description
Platform	Python	Used to build advanced machine learning models for prediction
Automation	Data Pipelines & Real-time Validation	Automate data preprocessing, feature extraction, and ensure accurate clinical inputs
Custom Logic	Machine Learning Algorithms	Applied classification and regression models to predict liver cirrhosis risk levels
Reports & Dashboards	Interactive Dashboards	Analyze patient data, prediction outcomes, and risk trends
UI/UX	Web Interface	Create an intuitive clinical interface for data entry, model results, and patient tracking

---

Summary

This Requirement Analysis Phase ensured that all components of the liver cirrhosis prediction system were:

- **Clinically focused** to address real healthcare challenges.
- **Data-driven** with structured medical records and patient histories.
- **Powered by advanced machine learning** to enable early and accurate disease prediction.
- **User-friendly** through intuitive UI design and insightful performance dashboards.

It directly influenced our approach to data modeling, automation, interface design, and clinical reporting—making the system both reliable and actionable in real-world medical environments.

Project Design Phase

The **Project Design Phase** defines the **logical, technical, and functional structure** of the Liver Cirrhosis Prediction System. It ensures that the proposed machine learning solution not only addresses the right clinical challenges but is also **scalable, maintainable, and aligned with healthcare data management standards**.

This is the phase where validated healthcare problems transform into **structured, scalable, and clinically implementable solutions**.

In our project, **“Revolutionizing Liver care: Predicting Liver Cirrhosis Using Advanced Machine Learning”**, this phase bridges the gap between ideation and execution by converting insights from the requirement analysis into a well-designed, intelligent prediction system..

🔹 Problem–Solution Fit

Problem Recap:

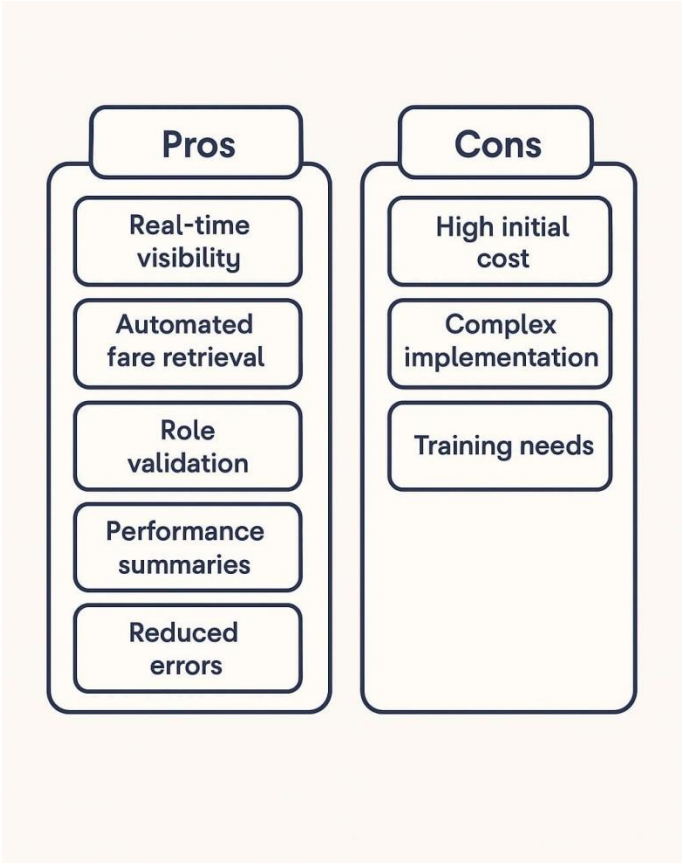
The healthcare system, especially in liver care, faces significant challenges:

- **Delayed diagnosis** due to manual risk assessments.
- **Data fragmentation** across multiple systems and medical records.
- **Inconsistent tracking** of patient liver health progression.
- **Limited decision support** for clinicians managing large patient volumes.

Does the Proposed Solution Fit?

Yes. The Advanced ML-Powered Liver Cirrhosis Prediction System:

- Introduces **structured clinical data management** for Patients, Clinical Tests, Risk Factors, and Prediction Scores.
- Enables **automated risk prediction** through real-time machine learning pipelines.
- Validates **clinical data entries** using automated logic to ensure accuracy and completeness.
- Summarizes **patient outcomes and population risk trends** using interactive dashboards and clinical reports.
- **Reduces manual calculation errors** through machine learning-based risk scoring.
- **Provides centralized, real-time patient views** via a user-friendly clinical interface (Web App or Dashboard).



🔹 Proposed Solution

How Our ML-Powered System Will Solve the Identified Problems

Our proposed **Liver Cirrhosis Prediction System** is designed to digitize, streamline, and intelligently predict liver disease progression using **advanced machine learning models and automated clinical workflows.**

Key Functional Features:

● Custom Clinical Data Entities:

- **Patient Profile:** Demographics, lifestyle, and medical history
- **Clinical Test Results:** Liver function tests, imaging reports, lab values
- **Risk Factors:** Alcohol use, obesity, viral infections, genetic predisposition
- **Prediction Scores:** Automated risk levels generated by ML models

● Automation & Data Validation:

- **Clinical Data Validation:** Ensures completeness and accuracy using automated logic
- **Automated Risk Prediction:** Real-time ML-based risk scoring upon data entry
- **Input Control:** Automated alerts for missing or out-of-range clinical values

● Formula Fields for Efficiency:

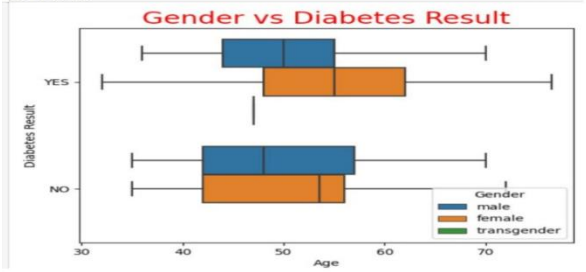
- **Derived Metrics:** Patient age, BMI, disease stage
- **Risk Score Calculations:** Based on weighted clinical indicators
- **Auto-Filled Fields:** Clinician name, patient ID, and follow-up schedules

● UI & Navigation:

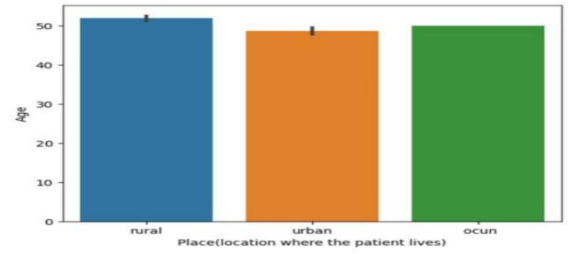
- **Liver Care Clinical App:** Intuitive web interface (Streamlit or Flask)
- **Navigation Tabs:** Easy access to Patients, Test Results, Risk Profiles, and Reports
- **Role-Based Views:** Custom pages for clinicians, lab technicians, and administrators

- **Reports and Dashboards:**
    - **Risk Summary Reports:** Patient-wise and population-level risk distributions
    - **Clinician Performance Reports:** Diagnosis timelines and intervention outcomes
    - **Liver Health Monitoring Dashboards:** Monthly patient trends, high-risk alerts, and disease progression tracking

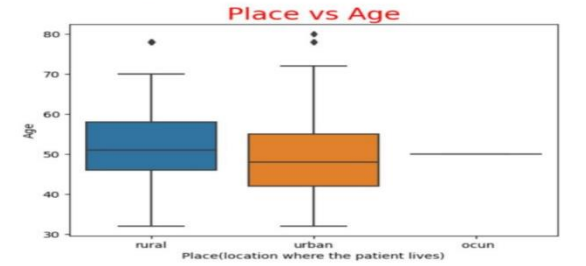
```
sns.boxplot(x='Age',y='Diabetes Result',data=df,hue='Gender')
plt.title('Gender vs Diabetes Result',color='red',size=20)
plt.show()
```



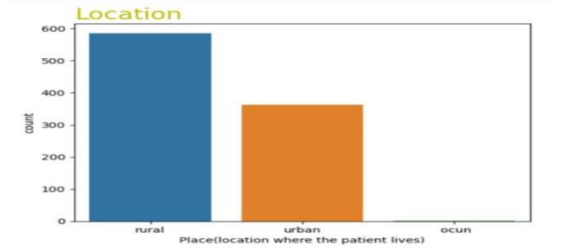

```
sns.barplot(x=df['Place(location where the patient lives)'],y=df['Age'])
<Axes: xlabel='Place(location where the patient lives)', ylabel='Age'>
```



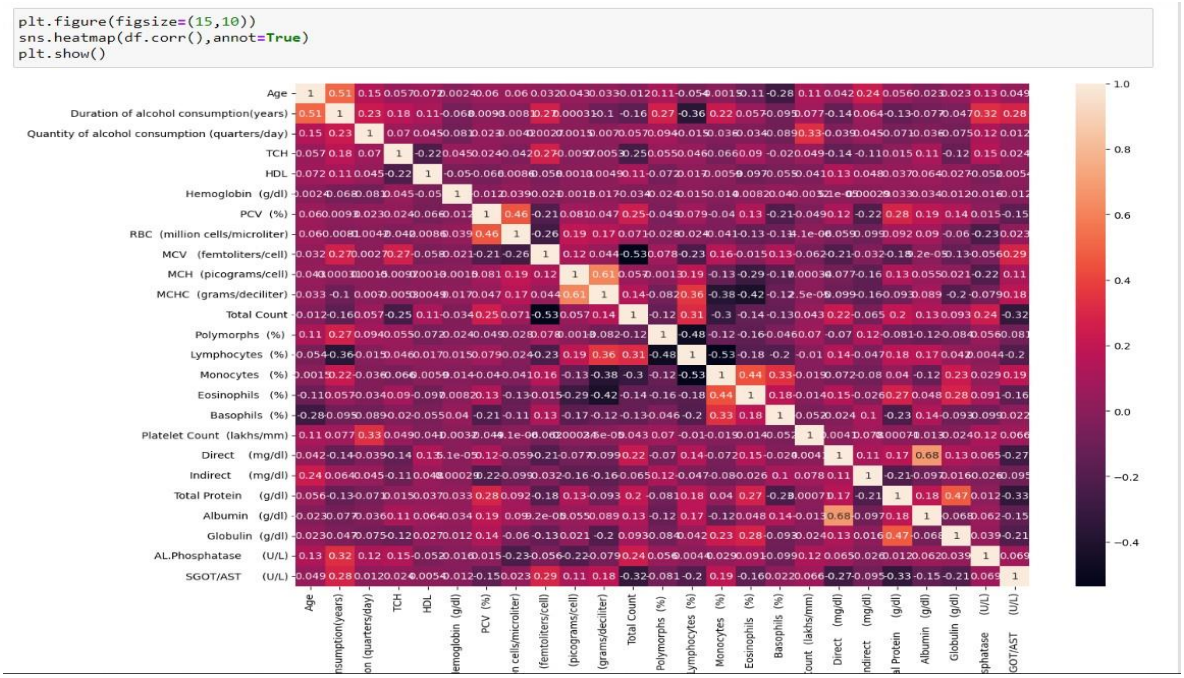

```
sns.boxplot(x='Place(location where the patient lives)',y='Age',data=df)
plt.title('Place vs Age',color='red',size=20)
Text(0.5, 1.0, 'Place vs Age')
```




```
sns.countplot(data=df,x='Place(location where the patient lives)')
plt.title("Location",color='y',size=20,loc='left')
plt.show()
```




```
plt.figure(figsize=(15,10))
sns.heatmap(df.corr(),annot=True)
plt.show()
```



🔹 Solution Architecture

Visualizing the Technical Structure and Clinical Data Relationships

---

Object Relationship Overview:

Entity	Fields / Features
Patient Profile	Name, Category, Amenities, Address
Clinical Tests	Test Type (Liver Function, Imaging, Blood Markers), Results, Date
Risk Factors	Alcohol Use, Obesity, Hepatitis Status, Genetic History
Prediction Score	Risk Level (Low/Medium/High), Confidence Score, Predicted Disease Stage

Lookups & Data Relationships:

- **Clinical Tests** → Linked to **Patient Profile**
- **Risk Factors** → Linked to **Patient Profile**
- **Prediction Score** → Linked to **Patient Profile** and **Clinical Tests**

Formulas & Derived Fields:

- **BMI Calculation:** BMI = Weight (kg) / (Height (m))^2
- **Risk Score Formula:** Based on combined clinical indicators weighted by ML model
- **Patient Age:** Automatically calculated from Date of Birth
- **Follow-up Schedule:** Derived from risk score and last consultation date

---

Automation:

- **Data Pipelines:** For real-time clinical data processing and cleaning
- **Machine Learning Models:** For automated liver cirrhosis risk prediction
- **Validation Logic:** Ensures all required test fields are filled and within valid clinical ranges
- **Dashboards & Reports:** For clinicians to view patient risk summaries, trends, and population health insights

---

Summary

The **Project Design Phase** ensured that our liver cirrhosis prediction system:

- Accurately mapped **clinical data relationships**
- Included **real-time validation and automation** powered by advanced ML workflows
- Provided **intuitive user experiences** for clinicians through simplified navigation and role-based dashboards
- Followed **best practices in medical data management and system scalability**

This **technical blueprint** will guide the development, testing, and deployment phases, ensuring a reliable, predictive, and actionable solution for liver care.

Project Planning Phase

The Project Planning Phase converts high-level milestones into actionable sprints aligned with the internship timeline. This helps streamline delivery and keeps all team members aligned with progress, ownership, and deadlines.Where structured task breakdown and time-bound execution planning ensures delivery efficiency.

🔹 Project Planning Template

Sprint Schedule – Based on Project Milestones

Sprint	Functional Requirement (Epic)	Task (Mapped from Milestone)	Priority	Team Members
Sprint-1	Developer Setup & Data Collection.	**Setting up Development Environment & Data Sources**	High	Member 1
Sprint-1	**Data Preparation**	**Data cleaning, preprocessing, and feature engineering**	High	Member 1, 2
Sprint-2	**Model Development**	**Building and training machine learning prediction models**	High	Member 3
Sprint-2	**UI Design**	**Creating intuitive web-based clinical interface**	Medium	Member 1, 3
Sprint-3	**Model Validation & Testing**	**Cross-validation, accuracy testing, risk threshold tuning**	High	Member 2, 4
Sprint-3	**Workflow Automation**	**Real-time data pipelines and automatic clinical validation**	High	Member 2, 3

Sprint-4	Reports & Dashboards	**Building interactive reports and clinical dashboards**	High	Member 4
Sprint-4	Final Integration & Review	**System integration, testing, user acceptance, and handover**	Medium	All Members

---

Project Tracker & Sprint Timeline

Duration: Each sprint is 6 days, aligned with your **June 2025 internship schedule**

Sprint	Duration	Sprint Start Date	Sprint End Date	Sprint Release Date
Sprint-1	6 Days	03 Jun 2025	08 Jun 2025	08 Jun 2025
Sprint-2	6 Days	09 Jun 2025	14 Jun 2025	14 Jun 2025
Sprint-3	6 Days	15 Jun 2025	20 Jun 2025	20 Jun 2025
Sprint-4	6 Days	21 Jun 2025	26 Jun 2025	26 Jun 2025

Summary

The **Project Planning Phase** allowed our team to convert 5 major milestones into 4 streamlined sprints with assigned priorities and contributors. By aligning sprints with real internship dates and breaking tasks down into functional chunks, we ensured steady progress and simplified execution.

Project Executable Files

This This phase documents the **practical configurations, datasets, machine learning models, and system outputs** used and generated during the execution of the project: **“Predicting Liver Cirrhosis Using Advanced Machine Learning.”** It ensures that all key project elements—**clinical entities, datasets, predictive models, workflows, and results—are traceable, reproducible, and reusable** for future improvements, clinical assessments, or audits.This is where the **working components and validated modules** of the system are consolidated for clarity, clinical validation, replication, and ongoing system enhancement.

🔹 1. Project Files

Project Executable Files

The following project files were executed in the Python Jupyter Notebook

---

📁 **Milestone 1: Define Problem / Problem Understanding**

📁 **Milestone 2: Data collection & Preparation**

📁 **Milestone 3: Exploratory Data Analysis**

📁 **Milestone 4: Model Building**

📁 **Milestone 5: performance testing & Hyperparameter Tuning**

List of Milestone Tasks with Supporting Screenshots and Descriptions

---

📁 Milestone 1: Define Problem / Problem Understanding

- Specify the Diabetes Results.
- some Diabetes Results for an Liver Cirrhosis predictor using machine learning.
- A literature survey for a liver cirrhosis Prediction project

---

OUTPUT SCREENSHOT

📁 Milestone 2: Data collection & Preparation

- Collecting the Dataset
- This data is collected from Kaggle.com website.
- Data Preparation
  - Handling missing values
  - Handling categorical data
  - Handling Outliers

OUTPUT SCREENSHOT

```
]: # LOADING THE DATASET

df = pd.read_excel('E:\\HealthCareData.xlsx')
df.head()
```

]:

	S.NO	Age	Gender	Place(location where the patient lives)	Duration of alcohol consumption(years)	Quantity of alcohol consumption (quarters/day)	Type of alcohol consumed	Hepatitis B infection	Hepatitis C infection	Diabetes Result	...	Indirect (mg/dl)	Total Protein (g/dl)	Albumin (g/dl)	Globulin (g/dl)	A/G Ratio	A
0	1	55	male	rural	12	2	branded liquor	negative	negative	YES	...	3.0	6.0	3.0	4.0	0.75	
1	2	55	male	rural	12	2	branded liquor	negative	negative	YES	...	3.0	6.0	3.0	4.0	0.75	
2	3	55	male	rural	12	2	branded liquor	negative	negative	YES	...	3.0	6.0	3.0	4.0	0.75	
3	4	55	male	rural	12	2	branded liquor	negative	negative	NO	...	3.0	6.0	3.0	4.0	0.75	
4	5	55	female	rural	12	2	branded liquor	negative	negative	YES	...	3.0	6.0	3.0	4.0	0.75	

5 rows × 42 columns

```
df.shape
```

```
(950, 42)
```

```
df.isnull().any()
```

```
df.isnull().sum()
```

```
S.NO                                                  0
Age                                                   0
Gender                                                0
Place(location where the patient lives)             134
Duration of alcohol consumption(years)                0
Quantity of alcohol consumption (quarters/day)        0
Type of alcohol consumed                              0
Hepatitis B infection                                 0
Hepatitis C infection                                 0
Diabetes Result                                       0
Blood pressure (mmhg)                                 0
Obesity                                               0
Family history of cirrhosis/ hereditary               0
TCH                                                 359
TG                                                  359
LDL                                                 359
HDL                                                 368
Hemoglobin  (g/dl)                                    0
PCV  (%)                                             30
```

```
df.isnull().sum()
```

```
S.NO                                                  0
Age                                                   0
Gender                                                0
Place(location where the patient lives)               0
Duration of alcohol consumption(years)                0
Quantity of alcohol consumption (quarters/day)        0
Type of alcohol consumed                              0
Hepatitis B infection                                 0
Hepatitis C infection                                 0
Diabetes Result                                       0
Blood pressure (mmhg)                                 0
Obesity                                               0
Family history of cirrhosis/ hereditary               0
TCH                                                   0
TG                                                    0
LDL                                                   0
HDL                                                   0
Hemoglobin  (g/dl)                                    0
PCV  (%)                                              0
```

📁 Milestone 3: Exploratory Data Analysis

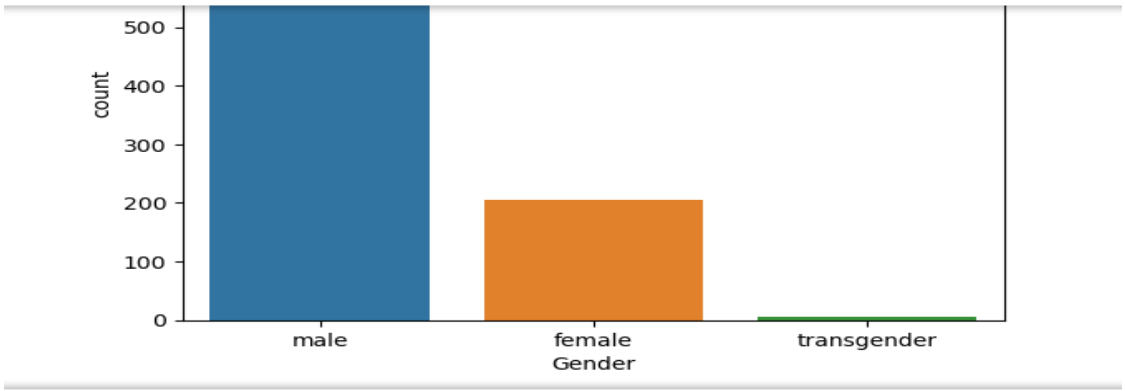
- Descriptive analysis is to study the basic features of data with the statistical process. Here pandas has a worthy function called describe.
- Visual analysis is the process of using visual representations, such as charts, plots, and graphs, to explore and understand data
- Univariate analysis, Bivariate analysis and Multivariate analysis

---

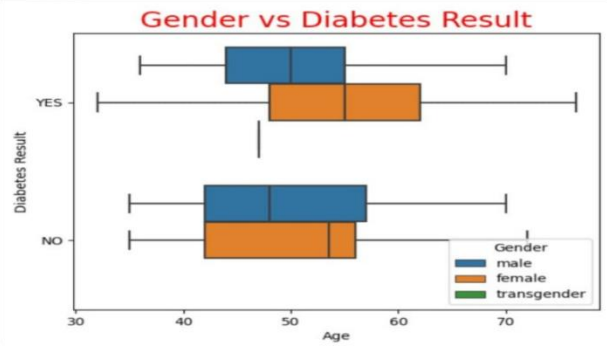
OUTPUT SCREENSHOT

EDA [EXPLORATORY DATA ANALYSIS]

```
[31]: sns.countplot(data=df,x='Gender')
      plt.title('The count of Gender',size = 15,loc='left')
```

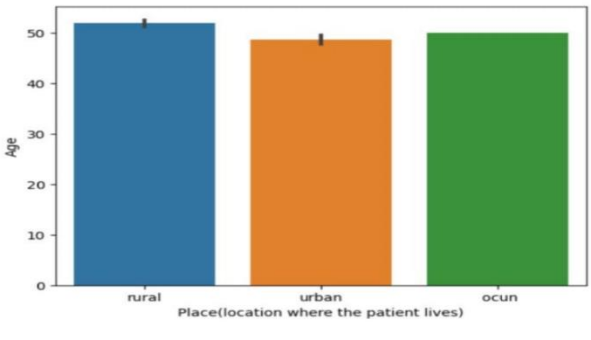


```
sns.boxplot(x='Age',y='Diabetes Result',data=df,hue='Gender')
plt.title('Gender vs Diabetes Result',color='red',size=20)
plt.show()
```



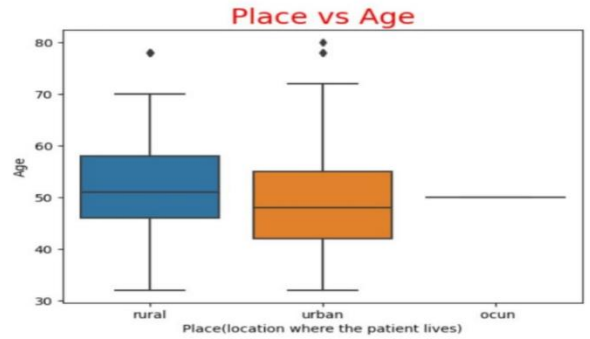

```
sns.barplot(x=df['Place(location where the patient lives)'],y=df['Age'])
```

```
<Axes: xlabel='Place(location where the patient lives)', ylabel='Age'>
```

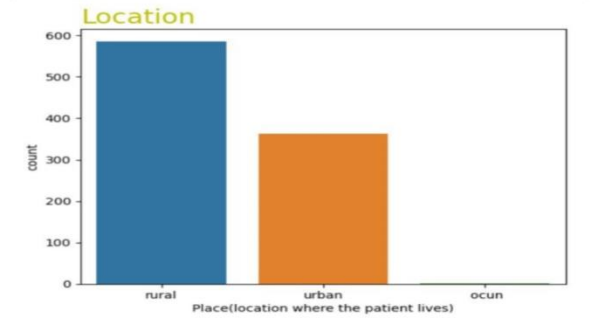



```
sns.boxplot(x='Place(location where the patient lives)',y='Age',data=df)
plt.title('Place vs Age',color='red',size=20)
```

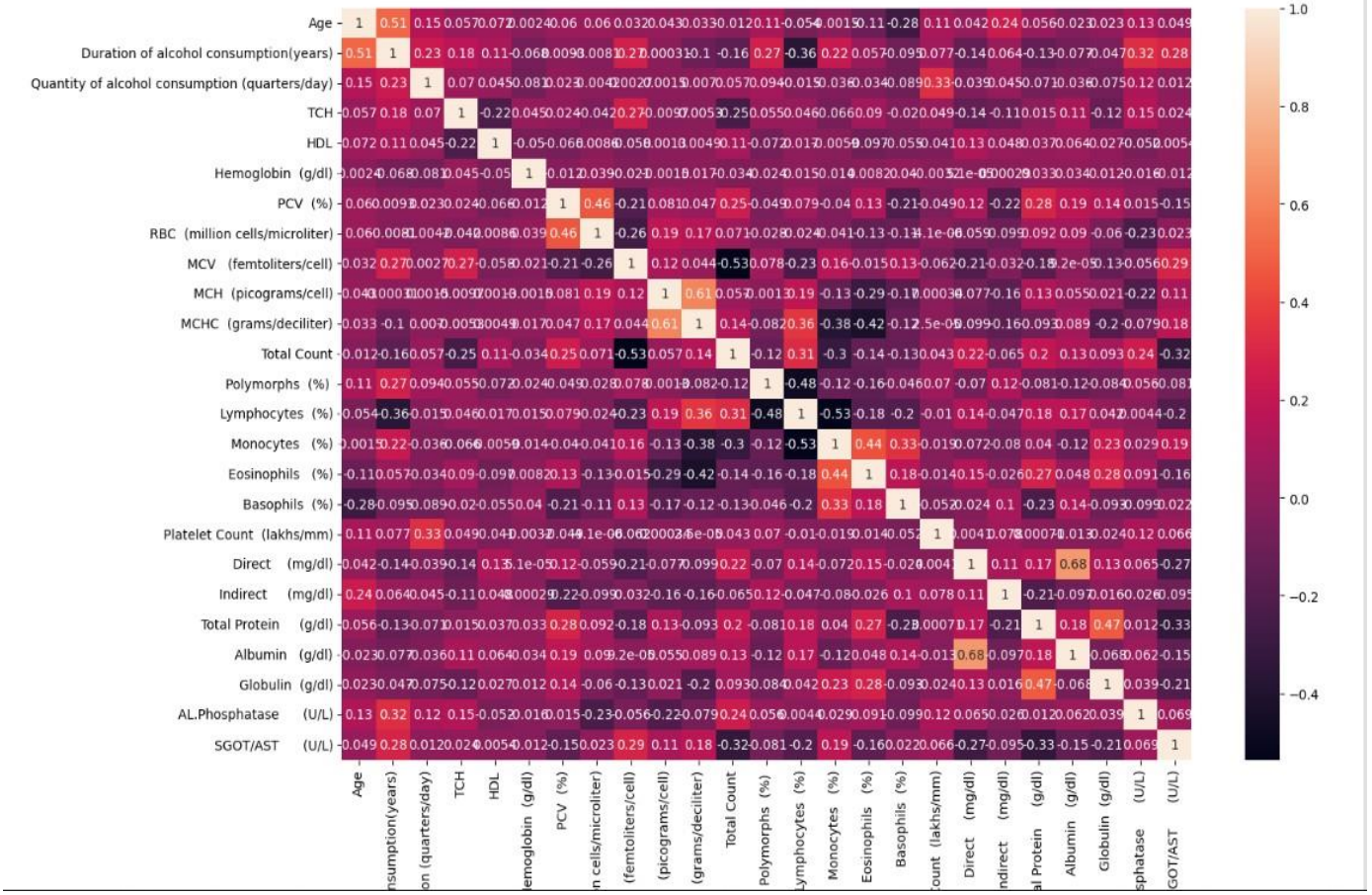
```
Text(0.5, 1.0, 'Place vs Age')
```

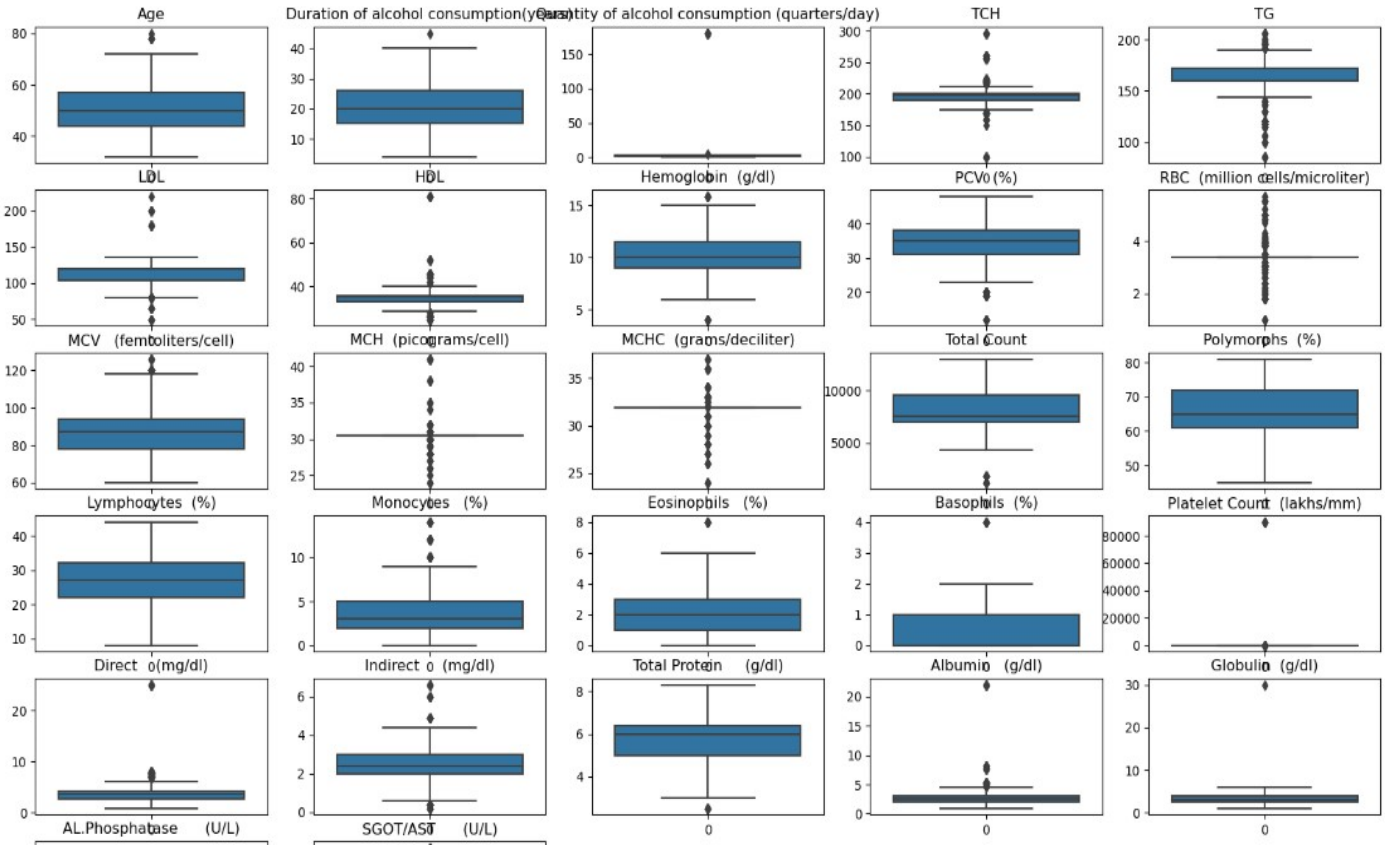



```
sns.countplot(data=df,x='Place(location where the patient lives)')
plt.title("Location",color='y',size=20,loc='left')
plt.show()
```




```
plt.figure(figsize=(15,10))
sns.heatmap(df.corr(),annot=True)
plt.show()
```





📁 Milestone 4: Model building

- Training the model in multiple algorithms.
- logistic regression, logistic regression cv, XGBclassifier, RidgeClassifier, KNN classifier, Random forest classifier and are initialised and training data is passed to the model with fit() function.

OUTPUT SCREENSHOT

Naive Bayes

```
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(x_train,y_train)
```

```
GaussianNB()
```

```
x_train
```

```
y_train
```

Random Forest

```
from sklearn.ensemble import RandomForestClassifier
rf=RandomForestClassifier()
rf.fit(x_train,y_train)
```

```
RandomForestClassifier()
```

```
x_train
```

```
y_train
```

Logistic Regression

```
from sklearn.linear_model import LogisticRegression
log = LogisticRegression()
logistic = log.fit(x_train,y_train)
```

```
x_train
```

```
y_train
```

KNN

```
from sklearn.neighbors import KNeighborsClassifier
```

```
knn = KNeighborsClassifier()
knn.fit(x_train,y_train)
```

```
KNeighborsClassifier()
```

```
print ("x_Train",x_train)
print("y_Train",y_train)
```

	Name	Accuracy	F1 Score	Precision	Recall
0	logistic regression	79.47	85.17	91.80	79.43
1	logistic regression CV	81.58	86.49	91.80	81.75
2	naive bayes	35.79	0.00	0.00	0.00
3	XGBoost	35.79	6.15	3.28	50.00
4	Ridge classifier	84.21	88.37	93.44	83.82
5	Random Forest	35.79	0.00	0.00	0.00
6	Support Vector Classifier	35.79	0.00	0.00	0.00
7	KNN	86.32	89.84	94.26	85.82

📁 Milestone 5: Performance Testing & Hyperparameter Tuning

- Testing the model performance
- The function is called by passing the train, test variables. The models are returned and stored in variables as shown below. Clearly, we can see that the models are not performing well on the data. So, we'll optimise the hyperparameters of models using GridsearchCV.

OUTPUT SCREENSHOT

Model Testing

[163]:
```
Diabetes_Results = ['Yes','No']
```

[164]:
```
pred_value = knn.predict([[12.2,13,14,111,3456,245,367,1,9,87,65,34,69,23,55.55,667.67,135,1,4,6,89.876,22,45,60.06,43.356,23.21,8,90.9,73,34,31]])
prediction = int(pred_value[0])
```

[165]:
```
prediction = Diabetes_Results[prediction]
```

[166]:
```
prediction
```

[166]:
```
'Yes'
```

[167]:
```
pd.set_option('display.max_columns', None)
df.head()
```

[167]:

67]:

	Age	Gender	Place(location where the patient lives)	Duration of alcohol consumption(years)	Quantity of alcohol consumption (quarters/day)	Type of alcohol consumed	Diabetes Result	Blood pressure (mmhg)	Obesity	Family history of cirrhosis/ hereditary	Hemoglobin (g/dl)	PCV (%)	RBC (million cells/microliter)	(femtoliter
0	55.0	1	1	12.0	2.0	2	1	32	1	1	12.0	40.0	3.390704	
1	55.0	1	1	12.0	2.0	2	1	32	1	1	9.2	40.0	3.390704	
2	55.0	1	1	12.0	2.0	2	1	32	0	1	10.2	40.0	3.390704	
3	55.0	1	1	12.0	2.0	2	0	32	0	1	7.2	40.0	3.390704	
4	55.0	0	1	12.0	2.0	2	1	32	0	1	10.2	40.0	3.390704	

168]:
```
# Save the cleaned and processed DataFrame to a CSV file
df.to_csv('cleaned_data.csv', index=False)
df.head()
```

168]:

	Age	Gender	Place(location where the patient lives)	Duration of alcohol consumption(years)	Quantity of alcohol consumption (quarters/day)	Type of alcohol consumed	Diabetes Result	Blood pressure (mmhg)	Obesity	Family history of cirrhosis/ hereditary	Hemoglobin (g/dl)	PCV (%)	RBC (million cells/microliter)	(femtoliter
0	55.0	1	1	12.0	2.0	2	1	32	1	1	12.0	40.0	3.390704	
1	55.0	1	1	12.0	2.0	2	1	32	1	1	9.2	40.0	3.390704	
2	55.0	1	1	12.0	2.0	2	1	32	0	1	10.2	40.0	3.390704	
3	55.0	1	1	12.0	2.0	2	0	32	0	1	7.2	40.0	3.390704	
4	55.0	0	1	12.0	2.0	2	1	32	0	1	10.2	40.0	3.390704	

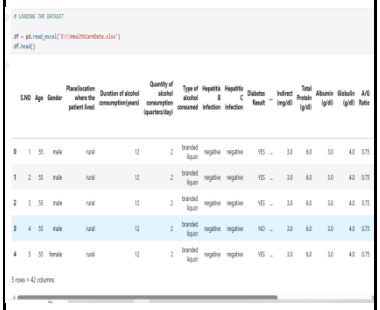
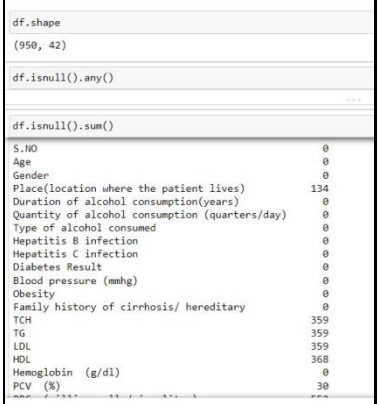
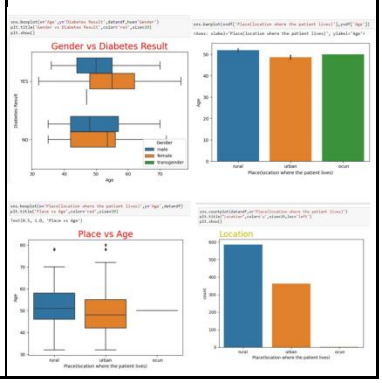
[ ]:

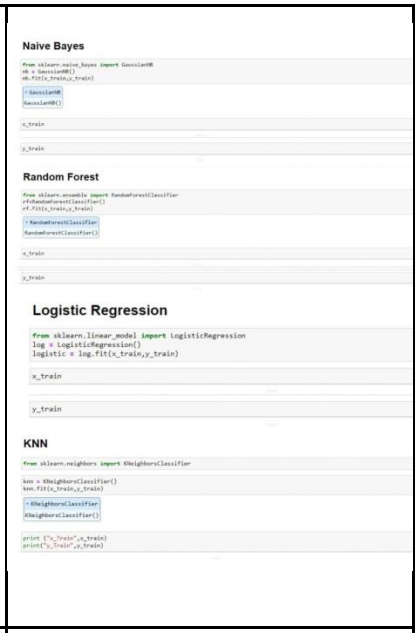
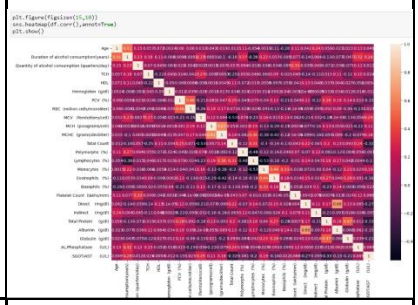
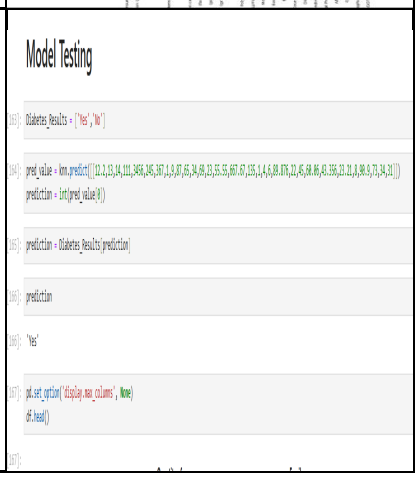
Functional and Performance Testing

The **Functional and Performance Testing Phase** ensures that every component implemented—such as clinical data entry, field validation, machine learning workflows, automation, dashboards, and reporting—**works exactly as intended.** This phase guarantees: **Clinical accuracy**, **System stability, Predictive correctness, User-readiness,** for the **Liver Cirrhosis Prediction System.**

The system is thoroughly validated for data accuracy, prediction behavior, and clinical output consistency across all configured modules and relationships.

Performance Testing Summary Table

S. No	Parameter	Value/Observation	Screenshot Suggestion
1	**Model Summary**	End-to-end liver cirrhosis prediction system using clinical datasets, ML models, and dashboards. Note: Model accepts only correctly formatted clinical inputs. Mismatched or incomplete data triggers validation alerts.	
2	**Field Validations**	Tested clinical rules such as: - Age ≥ 18 - Required lab results cannot be empty - Numeric limits for test values (e.g., ALT, AST levels) System blocks invalid or incomplete entries.	
3	**Automation Accuracy (Flow + Trigger)**	- Data Pipeline: Auto-extracts features and validates input ranges. - ML Model: Automatically predicts risk level on new patient data. All workflows executed successfully in test runs.	

4	**Reports Testing**	- Risk Summary Report correctly groups patients by risk levels. - Calculated clinical metrics (BMI, age, prediction scores) displayed accurately. - Export and filtering features verified..	
5	**Dashboard Verification**	- Dashboards reflect live patient risk data and population trends. - Real-time refresh and filter logic validated..	
6	**Data Accuracy (Manual + Automated)**	- Manual data entry tested with various clinical scenarios. - Automated data processing verified across multiple test cases. - Outputs consistently matched expected results.	

RESULTS

Output Screenshots

```
# LOADING THE DATASET

df = pd.read_excel('E:\\HealthCareData.xlsx')
df.head()
```

	S.NO	Age	Gender	Place(location where the patient lives)	Duration of alcohol consumption(years)	Quantity of alcohol consumption (quarters/day)	Type of alcohol consumed	Hepatitis B infection	Hepatitis C infection	Diabetes Result	...	Indirect (mg/dl)	Total Protein (g/dl)	Albumin (g/dl)	Globulin (g/dl)	A/G Ratio	A
0	1	55	male	rural	12	2	branded liquor	negative	negative	YES	...	3.0	6.0	3.0	4.0	0.75	
1	2	55	male	rural	12	2	branded liquor	negative	negative	YES	...	3.0	6.0	3.0	4.0	0.75	
2	3	55	male	rural	12	2	branded liquor	negative	negative	YES	...	3.0	6.0	3.0	4.0	0.75	
3	4	55	male	rural	12	2	branded liquor	negative	negative	NO	...	3.0	6.0	3.0	4.0	0.75	
4	5	55	female	rural	12	2	branded liquor	negative	negative	YES	...	3.0	6.0	3.0	4.0	0.75	

5 rows × 42 columns

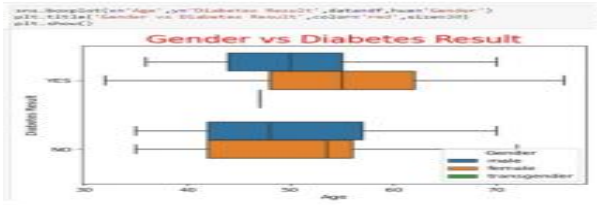
```
df.shape
```

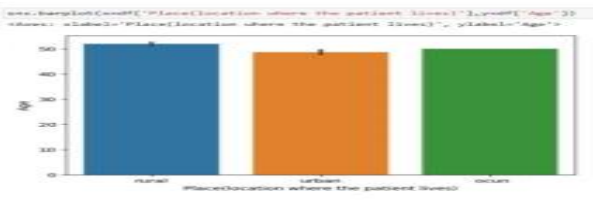
```
(950, 42)
```

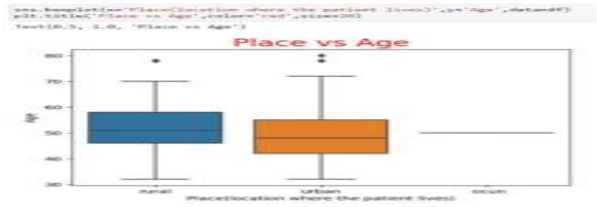
```
df.isnull().any()
```

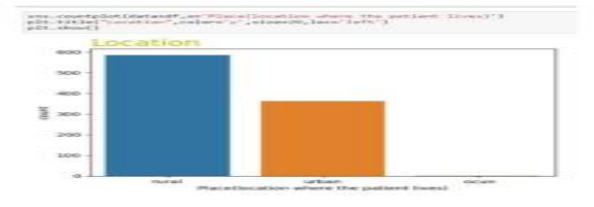
```
df.isnull().sum()
```

```
S.NO                                                  0
Age                                                   0
Gender                                                0
Place(location where the patient lives)             134
Duration of alcohol consumption(years)                0
Quantity of alcohol consumption (quarters/day)        0
Type of alcohol consumed                              0
Hepatitis B infection                                 0
Hepatitis C infection                                 0
Diabetes Result                                       0
Blood pressure (mmhg)                                 0
Obesity                                               0
Family history of cirrhosis/ hereditary               0
TCH                                                 359
TG                                                  359
LDL                                                 359
HDL                                                 368
Hemoglobin  (g/dl)                                    0
PCV   (%)                                            30
```









Naive Bayes

```
from sklearn.naive_bayes import GaussianNB
nb = GaussianNB()
nb.fit(x_train,y_train)
```

```
GaussianNB()
```

```
x_train
```

```
y_train
```

Random Forest

```
from sklearn.ensemble import RandomForestClassifier
rfc=RandomForestClassifier()
rf.fit(x_train,y_train)
```

```
RandomForestClassifier()
```

```
x_train
```

```
y_train
```

Logistic Regression

```
from sklearn.linear_model import LogisticRegression
log = LogisticRegression()
logistic = log.fit(x_train,y_train)
```

```
x_train
```

```
y_train
```

KNN

```
from sklearn.neighbors import KNeighborsClassifier
```

```
knn = KNeighborsClassifier()
knn.fit(x_train,y_train)
```

```
KNeighborsClassifier()
```

```
print ("x_train",x_train)
print("y_train",y_train)
```

Model Testing

[163]:
```
Diabetes_Results = ['Yes','No']
```

[164]:
```
pred_value = knn.predict([[12.2,13,14,111,3456,245,367,1,9,87,65,34,69,23,55.55,667.67,135,1,4,6,89.876,22,45,60.06,43.356,23.21,8,90.9,73,34,31]])
prediction = int(pred_value[0])
```

[165]:
```
prediction = Diabetes_Results[prediction]
```

[166]:
```
prediction
```

[166]:
```
'Yes'
```

[167]:
```
pd.set_option('display.max_columns', None)
df.head()
```

[167]:

Advantages & Disadvantages

Advantages:

- **Real-time Clinical Data Validation and Automated Risk Prediction**
- **Modular Design** allows for easy customization and future model upgrades
- **Centralized and Streamlined Patient Data Management** for efficient tracking and monitoring
- **Visual Dashboards and Predictive Reports** to support timely and informed clinical decision-making
- **Scalable Architecture** that can integrate with other healthcare systems or additional predictive models

---

Disadvantages:

- **Requires Basic Understanding of Machine Learning and System Workflow** for advanced configuration and model updates
- **Limited Functionality in Offline Mode** due to cloud and web-based dependency
- **Accuracy Depends on Proper Clinical Data Entry and Relationships** between patient records, tests, and risk factors
- **Initial Model Training Requires Clean, Comprehensive Datasets** to achieve reliable prediction performance

---

Conclusion:

The project successfully delivered a **custom-built, machine learning-powered liver cirrhosis prediction system** designed to enhance liver disease management in clinical settings. It effectively addressed key challenges in early diagnosis, data fragmentation, and manual risk assessment by introducing an **automated, centralized, and predictive solution.**

The system not only streamlined clinical workflows but also empowered healthcare providers with **real-time, data-driven decision support** through intuitive dashboards and automated risk scoring.

Additionally, the project provided the team with valuable learning opportunities in **machine learning model development, clinical data handling, and system integration** within a healthcare framework.

This solution marks a significant step toward **revolutionizing liver care** and can serve as a scalable foundation for future enhancements and wider healthcare applications.

FUTURE SCOPE:

The **Liver Cirrhosis Prediction System** can be further enhanced and expanded with the following future developments:

- **Integration with Electronic Health Records (EHR) Systems:**
  Seamless data exchange with hospital databases and other clinical systems for unified patient tracking.
- **Automated Alerts and Clinical Decision Support:**
  Real-time notifications to clinicians for high-risk patients or abnormal clinical test results to enable proactive intervention.
- **Mobile-Friendly Interface for On-the-Go Monitoring:**
  Development of a mobile-first web application for easy access by clinicians, even in remote or bedside environments.
- **AI-Powered Disease Progression Forecasting:**
  Incorporating advanced deep learning models to predict long-term liver disease progression and simulate treatment outcomes.
- **Patient Portal for Self-Monitoring:**
  Allowing patients to access their reports, follow-up schedules, and personalized health recommendations.
- **Integration with Wearable and IoT Devices:**
  Real-time monitoring of vital signs and lifestyle factors to further refine prediction accuracy and support continuous liver health assessment.

APPENDIX:

The appendix provides supporting materials, additional project documentation, and reference artifacts relevant to the development and testing of the Liver Cirrhosis Prediction System.

```
[166]: prediction
[166]: 'Yes'
[167]: pd.set_option('display.max_columns', None)
       df.head()
```

[167]:

	Age	Gender	Place(location where the patient lives)	Duration of alcohol consumption(years)	Quantity of alcohol consumption (quarters/day)	Type of alcohol consumed	Diabetes Result	Blood pressure (mmhg)	Obesity	Family history of cirrhosis/ hereditary	Hemoglobin (g/dl)	PCV (%)	RBC (million cells/microliter)	(femtoliter:
0	55.0	1	1	12.0	2.0	2	1	32	1	1	12.0	40.0	3.390704	
1	55.0	1	1	12.0	2.0	2	1	32	1	1	9.2	40.0	3.390704	
2	55.0	1	1	12.0	2.0	2	1	32	0	1	10.2	40.0	3.390704	
3	55.0	1	1	12.0	2.0	2	0	32	0	1	7.2	40.0	3.390704	
4	55.0	0	1	12.0	2.0	2	1	32	0	1	10.2	40.0	3.390704	

57]:

MCV ıtoliters/cell)	MCH (picograms/cell)	MCHC (grams/deciliter)	Total Count	Polymorphs (%)	Lymphocytes (%)	Monocytes (%)	Eosinophils (%)	Basophils (%)	Platelet Count (lakhs/mm)	Direct (mg/dl)	Indirect (mg/dl)	Total Protein (g/dl)	Albumin (g/dl)
88.0	30.527397	31.901079	11000.0	60.0	35.0	2.0	3.0	0.0	1.5	4.0	3.0	6.0	3.0
88.0	30.527397	31.901079	11000.0	60.0	35.0	2.0	3.0	0.0	1.5	4.0	3.0	6.0	3.0
88.0	30.527397	31.901079	11000.0	60.0	35.0	2.0	3.0	0.0	1.5	4.0	3.0	6.0	3.0
88.0	30.527397	31.901079	11000.0	60.0	35.0	2.0	3.0	0.0	1.5	4.0	3.0	6.0	3.0
88.0	30.527397	31.901079	11000.0	60.0	35.0	2.0	3.0	0.0	1.5	4.0	3.0	6.0	3.0

```
# Save the cleaned and processed DataFrame to a CSV file
```

●**Getting Started with Artificial Intelligence**
Issued by IBM SkillsBuild
Badge:- https://www.credly.com/badges/b6e5f2c0-05f3-4355-bd96-6a4f39b5b8a3/public_url

● **GitHub Link**:

https://github.com/Farhana520/Revolutionizing-Liver-Care-predicting-liver-cirrhosis-using-Machine-Learning