Final Report

Liver Cirrhosis Prediction using Machine Learning

1. INTRODUCTION

1.1 Project Overview

Liver cirrhosis is a life-threatening chronic disease that gradually replaces healthy liver tissue with scar tissue, impairing liver function. The disease is often diagnosed at later stages due to vague symptoms or lack of awareness. This project aims to build a machine learning–based prediction model that can detect liver cirrhosis early using clinical and biochemical patient data.

1.2 Purpose

The purpose of this project is to assist healthcare providers in early diagnosis of liver cirrhosis using AI techniques, enabling timely intervention and better treatment outcomes.

2. IDEATION PHASE

2.1 Problem Statement

Early detection of liver cirrhosis remains a challenge due to limited access to diagnostics and vague early symptoms. A cost-effective, data-driven approach is necessary to identify high-risk individuals before irreversible damage occurs.

2.2 Empathy Map Canvas

Think: Concerned about worsening liver health without visible symptoms.
Feel: Anxiety over costly late-stage treatments.
Say: 'I wish I had known sooner.'
Do: Visit hospitals only when symptoms are severe.

2.3 Brainstorming

We explored multiple approaches including medical imaging, rule-based systems, and machine learning. ML was chosen due to its efficiency, accuracy, and ability to work with standard lab data.

3. REQUIREMENT ANALYSIS

3.1 Customer Journey Map

The patient provides lab test data, the system predicts the risk, and doctors can intervene early.

3.2 Solution Requirements

- Clean and labeled dataset
- Exploratory data analysis tools

- Classification ML algorithms
- Visualization tools

3.3 Data Flow Diagram

Patient Data → Preprocessing → ML Model → Prediction Output → Diagnosis Support

3.4 Technology Stack

- Programming Language: Python
- Libraries: pandas, sklearn, matplotlib, seaborn
- Jupyter Notebook

4. PROJECT DESIGN

4.1 Problem-Solution Fit

A predictive model based on patient clinical data addresses the issue of late diagnosis and enables data-driven decision-making.

4.2 Proposed Solution

Build a classification model using logistic regression, decision trees, and random forests to predict liver cirrhosis risk.

4.3 Solution Architecture

1. Load and clean dataset
2. Perform feature selection and EDA
3. Train model
4. Validate and test model
5. Deploy model

5. PROJECT PLANNING & SCHEDULING

5.1 Project Planning

Week 1: Data collection
Week 2: Preprocessing & EDA
Week 3: Model building
Week 4: Testing and validation
Week 5: Report

6. FUNCTIONAL AND PERFORMANCE TESTING

6.1 Performance Testing

The Random Forest model achieved a validation accuracy of approximately 85%. Fine-tuning improved performance slightly.

7. RESULTS

7.1 Output Screenshots

Include screenshots of the confusion matrix, accuracy graphs, and predictions. (To be attached manually)

8. ADVANTAGES & DISADVANTAGES

Advantages:
- Early detection
- Scalable
- Uses simple lab data

Disadvantages:
- Depends on data quality
- May need retraining for other populations

9. CONCLUSION

The liver cirrhosis prediction model provides an efficient tool for healthcare professionals to detect cirrhosis early. It has the potential to reduce late-stage diagnosis and improve patient outcomes.

10. FUTURE SCOPE

Future improvements can include deploying the model in clinical software, real-time prediction apps, and incorporating more advanced algorithms or larger datasets.

11. APPENDIX

Source Code: Available on GitHub
Dataset: https://www.kaggle.com/datasets/bhavanipriya222/liver-cirrhosis-prediction
GitHub: https://github.com/Chandrika591/Liver_Cirrohosis_