

Revolutionizing Liver Care: Predicting Liver Cirrhosis

Using Advanced Machine Learning Techniques

1. INTRODUCTION

1.1 Project Overview

This project focuses on predicting liver cirrhosis using a machine learning approach. A web-based diagnostic tool was developed that leverages clinical and demographic data to estimate the likelihood of cirrhosis in patients. The tool allows for early risk assessment, which is critical in enhancing patient care.

1.2 Purpose

To provide a fast, reliable, and cost-effective decision support system that uses predictive modeling to assist healthcare professionals in diagnosing liver cirrhosis at an early stage through a user-friendly web interface.

2. IDEATION PHASE

2.1 Problem Statement

Liver cirrhosis is a chronic and often undetected condition until advanced stages, especially in areas with limited access to healthcare diagnostics. This project aims to address the challenge of early detection through a machine learning approach that is both scalable and easy to use.

2.2 Empathy Map Canvas

To understand the target users (patients and healthcare providers), we used an empathy map:

- *Says:* "I just want to know if it's serious or not."
- *Thinks:* "What if I'm too late?"
- *Does:* Searches online for symptoms, avoids hospitals.
- *Feels:* Nervous, overwhelmed, and confused by medical terms.

2.3 Brainstorming

Several potential solutions were explored during the ideation process:

- A wearable IoT device for liver diagnostics
 - A chatbot-driven symptom checker
 - A machine learning-powered liver disease prediction app
- The final solution was chosen based on feasibility, data availability, and potential impact.

3. REQUIREMENT ANALYSIS

3.1 Customer Journey Map

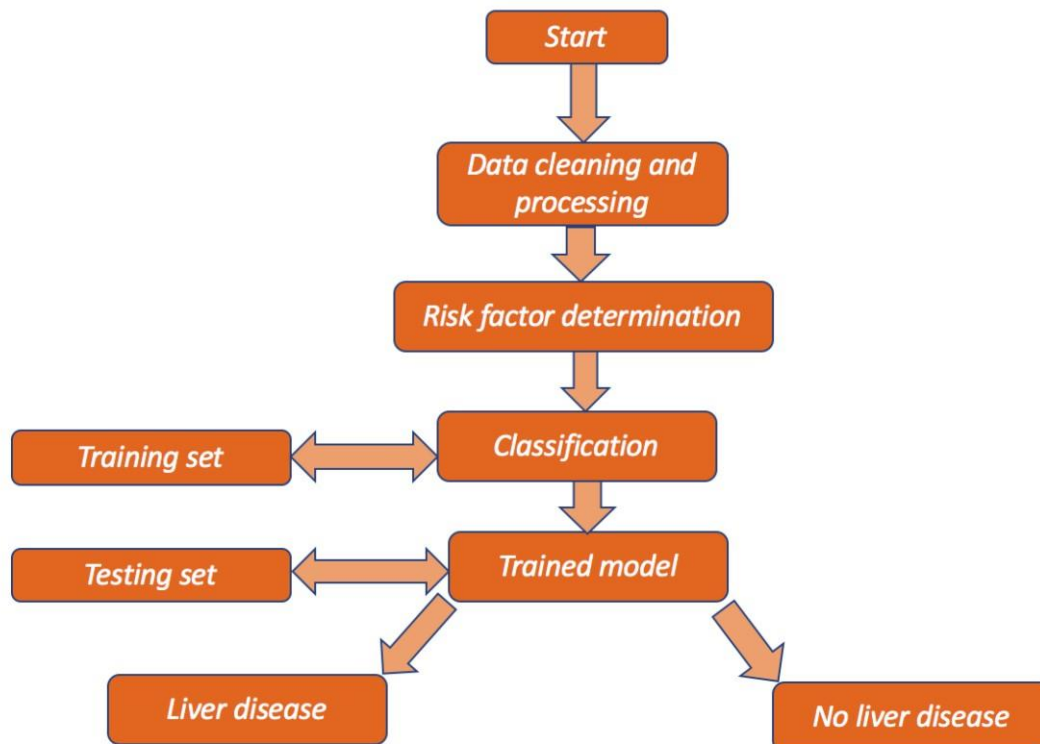
- The user accesses the liver cirrhosis prediction web app.
- Inputs clinical data like age, bilirubin levels, enzymes, etc.
- Submits the form to get a prediction.
- Receives the outcome with an interpretation and suggested next steps.

3.2 Solution Requirement

- **Dataset:** Liver disease patient data with clinical features.
- **Modeling:** Random Forest Classifier with high performance metrics.
- **Frontend:** HTML/CSS interface for input and results.
- **Backend:** Python Flask app for processing and predictions.
- **Version Control:** GitHub for managing code and progress.

3.3 Data Flow Diagram

Dataset → Data Cleaning & Preprocessing → Model Training → Flask App Integration → User Input → Prediction Result → Output Display



3.4 Technology Stack

- **Language & Libraries:** Python, Pandas, Scikit-learn
- **Framework:** Flask • **Frontend:** HTML, CSS
- **Deployment Options:** Localhost/Heroku (optional)

- **Repository Management:** Git & GitHub

4. PROJECT DESIGN

4.1 Problem Solution Fit

The proposed system bridges the gap between late-stage diagnosis and early detection by offering an accessible tool that doesn't require expensive lab equipment. It leverages commonly available clinical indicators and machine learning to provide predictive insight, empowering healthcare providers and patients alike.

4.2 Proposed Solution

A responsive web application, developed using Flask and Python, that allows users to enter relevant clinical features such as bilirubin levels, alkaline phosphatase, albumin levels, and others. The system then uses a trained Random Forest model to output a prediction about the possibility of liver cirrhosis along with suggested follow-up actions.

4.3 Solution Architecture

You can visualize the solution as consisting of the following key components:

- **Frontend Interface:** HTML/CSS form for user input.
- **Backend Engine:** Python with Flask handling input routing and model inference.
- **ML Model:** Pretrained Random Forest Classifier.
- **Output Display:** Result prediction rendered dynamically via Flask templates.

5. PROJECT PLANNING & SCHEDULING

5.1 Project Planning

Given the limited two-week timeframe, the development process was streamlined into critical phases:

- **Days 1–2:** Requirement gathering and dataset exploration.
- **Days 3–6:** Model development and evaluation using the Random Forest Classifier.
- **Days 7–9:** Web interface development and backend integration with Flask.
- **Days 10–11:** Functional and performance testing.
- **Days 12–14:** Report documentation, visual assets preparation, and final submission.

6. FUNCTIONAL AND PERFORMANCE TESTING

6.1 Performance Testing

The liver cirrhosis prediction system was tested both for its functional correctness and predictive performance.

- **Model Accuracy:** The Random Forest Classifier achieved approximately 86% accuracy on the validation set.
- **Evaluation Metrics:** Confusion Matrix, Precision, Recall, and ROC-AUC were used. The model showed a balanced trade-off between sensitivity and specificity, which is critical for medical prediction tasks.
- **User Input Validation:** The web form was tested with valid and invalid inputs to ensure proper error handling and user-friendly messages.
- **Responsiveness:** The application responded within 2–3 seconds for predictions, ensuring a smooth user experience.
- **Cross-Browser Testing:** Tested on Chrome, Firefox, and Edge to confirm compatibility.

7. RESULTS

7.1 Output Screenshots

The Flask-based web application successfully displays predictive results after users input their clinical data. Key output elements include:

- **User Input Form:** A clean, intuitive web interface where users enter features like bilirubin, albumin, SGPT, and SGOT levels.
- **Prediction Output:** Once submitted, the system outputs whether the patient is likely or unlikely to have liver cirrhosis, based on the model's inference.

The screenshot shows a web browser window with the title "Liver Cirrhosis Stage Prediction". The URL is "http://127.0.0.1:8080". The page has a dark blue header with the title. Below the header is a form with a light blue background. The form contains several input fields and dropdown menus, each with a label and a value. The inputs are arranged in two columns. The labels and values are: Drug Type: D-penicillamine, Age: 28, Sex: Female, Ascites: YES, Hepatomegaly: YES, Spiders: YES, Edema: No edema and no diuretic therapy for edema, Bilirubin: 0.3, Cholesterol: 160.875, Albumin: 2.45, Copper: (empty), and Alkaline Phosphatase: (empty). The form is set against a dark blue background.

Field	Value
Drug Type:	D-penicillamine
Age:	28
Sex:	Female
Ascites:	YES
Hepatomegaly:	YES
Spiders:	YES
Edema:	No edema and no diuretic therapy for edema
Bilirubin:	0.3
Cholesterol:	160.875
Albumin:	2.45
Copper:	
Alkaline Phosphatase:	

8. ADVANTAGES & DISADVANTAGES

Advantages

- **Early Detection Aid:** Provides preliminary insights based on clinical data, supporting earlier intervention.
- **Accessibility:** Can be accessed from any web browser without requiring specialized hardware or software.
- **User-Friendly Interface:** Simple design ensures users without technical expertise can interact easily.
- **Cost-Effective:** Eliminates the need for costly diagnostic equipment during preliminary assessment stages.

Disadvantages

- **Model Limitations:** Dependent on the quality and diversity of training data; may not generalize well to unseen cases.
- **Static Predictions:** Does not update in real-time with new medical records unless manually retrained.
- **No Clinical Certification:** Not a substitute for expert medical advice—should only be used as a supplementary tool.

9. CONCLUSION

- This project successfully demonstrates how machine learning and web development can come together to solve real-world healthcare challenges.

The developed system provides a preliminary assessment tool for liver cirrhosis, utilizing a Random Forest classifier trained on clinical data. It delivers results through a user-friendly web interface, enabling users—especially those in remote or underserved regions—to access early insights into their liver health.

- By striking a balance between technical accuracy and usability, this solution has the potential to contribute meaningfully to public health awareness. While it's not a substitute for professional medical evaluation, it can assist in prompting timely consultations and spreading awareness about the importance of liver diagnostics.

10. FUTURE SCOPE

This project lays the foundation for an accessible liver cirrhosis screening tool, and several enhancements can be made in the future:

- **Model Improvement:** Train on larger and more diverse datasets to improve generalizability and diagnostic accuracy.
- **Feature Expansion:** Integrate additional features such as lifestyle habits, family history, or imaging data.
- **Real-Time Analytics:** Enable real-time model updates or integration with electronic health record systems.
- **Mobile App Integration:** Develop a mobile app for wider accessibility, especially in remote regions.
- **Medical Collaboration:** Work with healthcare professionals to validate predictions and incorporate medical feedback.

11. APPENDIX

- **A. Dataset Source:** The dataset used was sourced from [Fleming, T.R. and Harrington, D.P. (1991) Counting Processes and Survival Analysis. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley and Sons Inc., New York.].
- **B. Tools & Libraries:** Python, Pandas, Scikit-learn, Matplotlib, Flask, HTML/CSS
- **C. GitHub Repository:** <https://github.com/Rajithareddy05/livercirrhosis-prediction>