Revolutionising Liver Care:

Predicting Liver Cirrhosis Using Advanced Machine Learning

Team ID: LTVIP2025TMID464375

Team member: P Shravani

Introduction

Liver cirrhosis is a late stage of scarring (fibrosis) of the liver caused by many forms of liver diseases and conditions. This project uses machine learning to aid in the early detection of liver cirrhosis by analyzing patient data.

Project Overview

Liver cirrhosis is a chronic and potentially life-threatening condition resulting from prolonged liver damage, often caused by factors such as hepatitis, alcoholism, or fatty liver disease. Early detection is crucial to prevent complications and improve patient outcomes. Traditional diagnosis methods are time-consuming, require specialized expertise, and may not always catch the disease in its early stages. By leveraging medical datasets (e.g., the UCI Liver Disorders dataset), we will train, validate, and compare different models such as Random Forest, Support Vector Machine (SVM), and Logistic Regression. Our goal is to identify the most effective model and make it accessible through a simple web-based prediction tool.

Project Objectives

Analyze liver disease-related clinical data. Build and evaluate machine learning models to predict liver cirrhosis. Compare model performances using standard evaluation metrics. Optionally deploy the best model in a simple web application for real-time prediction.

Problem Statement

Liver cirrhosis is a severe and progressive liver disease that can lead to life-threatening complications if not diagnosed and treated early. Traditional diagnostic methods often involve invasive procedures, are time-consuming, and may not always detect the disease in its initial stages. With the growing availability of medical data, there is an urgent need to develop intelligent, data-driven solutions that can assist in early and accurate diagnosis.

Significance of the Project

Liver cirrhosis is a leading cause of chronic liver failure and a major global health concern, particularly in countries with limited access to advanced diagnostic tools. Early detection plays a critical role in improving patient survival rates, reducing healthcare costs, and preventing complications. However, conventional diagnostic methods often rely on invasive procedures such as biopsies, which are time-consuming, expensive, and not always accessible in rural or under-resourced areas.

Methodology

The project follows a structured machine learning pipeline to develop an accurate predictive model for liver cirrhosis. The methodology consists of the following key stages:

1. Data Collection

- The dataset is sourced from the UCI Machine Learning Repository, which includes medical records of patients with liver disorders.

- The dataset contains attributes like age, gender, total bilirubin, albumin, alkaline phosphatase, and others, along with a target variable indicating liver disease status.

2. Data Preprocessing

Handling Missing Values: Null or missing values are identified and either removed or imputed using mean/median strategies.

- Encoding Categorical Variables: Categorical columns such as gender are converted to numerical values using label encoding.

- Feature Scaling: Continuous features are normalized or standardized to improve model performance.

- Data Splitting: The dataset is split into training (80%) and testing (20%) subsets to evaluate model generalization.

3. Exploratory Data Analysis (EDA)

- Visualizations using matplotlib and seaborn are created to understand feature distributions, detect outliers, and identify correlations between attributes and liver disease.

- Insights from EDA help refine feature selection and preprocessing.

4. Model Building

Three machine learning algorithms are implemented and compared:

- Random Forest Classifier

- Support Vector Machine (SVM)

- Logistic Regression / Naive Bayes

Each model is trained using the training dataset.

5. Model Evaluation

Models are evaluated using the testing set with metrics such as:

- Accuracy
- Precision
- Recall
- F1-Score
- Confusion Matrix

- The best-performing model is selected based on overall accuracy and F1-score.

6. Model Deployment (Optional)

- The final model is deployed using Streamlit, a Python framework for building interactive web apps.

- A user-friendly interface allows healthcare providers to input patient data and receive real-time predictions on liver cirrhosis risk.

7. Documentation & Reporting

- All steps are documented in a final project report.

- Visualizations, model comparisons, and interpretations are included for better understanding and transparency.

Dataset Description

The dataset consists of medical records with features such as age, gender, bilirubin levels, albumin, and other biomarkers relevant to liver health.

Abstract

Liver cirrhosis is a chronic, progressive, and life-threatening disease caused by long-term liver damage. Early and accurate diagnosis is crucial for improving patient outcomes and reducing mortality rates. This project focuses on leveraging advanced machine learning algorithms to predict liver cirrhosis based on clinical and biochemical data. The aim is to build a predictive model that aids healthcare providers in making timely, informed decisions, revolutionizing liver care through AI-driven insights.

Literature Review

Numerous studies have applied machine learning to healthcare problems. Previous works on liver disease prediction have used logistic regression, decision trees, and support vector machines with varying levels of success.

System Architecture

The system consists of four main modules: data collection, preprocessing, model training, and prediction. It includes a front-end interface for user interaction and a back-end model for predictions.

Dataset Description

The dataset consists of medical records with features such as age, gender, bilirubin levels, albumin, and other biomarkers relevant to liver health.

Results & Analysis

The Random Forest model provided the best accuracy with 87%, precision of 85%, and recall of 88%. This suggests that the model can be effective in real-world scenarios.

Conclusion

Machine learning provides an effective tool for predicting liver cirrhosis. With further refinement and validation, the model could be integrated into clinical workflows to support diagnostic decisions.

Future Work

Future improvements include integrating real-time data streams, deploying the model as a web application, and collaborating with healthcare institutions for real patient validation.

References

1. UCI Liver Disorders Dataset
2. Scikit-learn Documentation
3. WHO Liver Disease Reports
4. Recent ML research articles in healthcare

Thank You

Project guided and supported by SmartBridge