Revolutionizing Liver Care: Predicting Liver Cirrhosis Using Advanced Machine Learning

Abstract

Liver cirrhosis is the final phase of chronic liver disease, characterized by scarring and impaired liver function. Traditional diagnosis relies on invasive procedures and often comes too late. This project introduces a data-driven approach using machine learning (ML) to predict cirrhosis based on clinical and biochemical markers. The goal is to assist clinicians in early diagnosis, reducing dependence on costly or delayed lab tests. By leveraging publicly available datasets and applying multiple ML algorithms, this project aims to identify the most accurate predictive model, which can be used in real-time screening tools.

Objective

- Predict liver cirrhosis with high accuracy using clinical lab values.
- Determine the most influential features (biomarkers) using feature importance.
- Compare classification algorithms: Logistic Regression, SVM, Decision Tree, Random Forest, and XGBoost.
- Develop a lightweight web app for clinical environments.
- Advocate for data-driven decision-making in preventive hepatology.

Problem Statement

Liver diseases account for over 2 million deaths globally each year. Cirrhosis, being asymptomatic in early stages, often goes undetected. Conventional diagnostics like liver biopsy are invasive and expensive. There's a clear need for predictive models that can identify potential cirrhosis cases using basic lab reports. This project addresses that gap by building an ML-powered prediction system that is both interpretable and actionable for primary healthcare workers.

Dataset Description

- Dataset: Indian Liver Patient Dataset (ILPD) from UCI
- Instances: ~583 patient records
- Features: Age, Gender, Total/Direct Bilirubin, SGPT, SGOT, Alkaline Phosphatase, Total Protein, Albumin, A/G Ratio
- Target: Liver Disease Status (1 = Cirrhosis, 0 = Healthy)

Revolutionizing Liver Care: Predicting Liver Cirrhosis Using Advanced Machine Learning

Clinical Note: Elevated bilirubin, altered SGOT/SGPT ratio, and low albumin are strong cirrhosis indicators.

Technology Stack

- Language: Python
- Data Handling: Pandas, NumPy
- Modeling: Scikit-learn, XGBoost
- Visualization: Matplotlib, Seaborn
- Deployment: Streamlit (for web app)
- Version Control: GitHub

Methodology

1. Data Cleaning (null handling, encoding)
2. Exploratory Data Analysis (correlations, distributions)
3. Feature Engineering (ratios, transformations)
4. Model Building (training 5 models, cross-validation)
5. Evaluation (accuracy, precision, recall, AUC)
6. Deployment using Streamlit

Results

Model Performance:

- Logistic Regression: 76% accuracy, 0.80 AUC
- SVM: 78% accuracy, 0.82 AUC
- Random Forest: 87% accuracy, 0.90 AUC
- XGBoost: 86% accuracy, 0.89 AUC

Top Features:

1. Total Bilirubin
2. Direct Bilirubin

Revolutionizing Liver Care: Predicting Liver Cirrhosis Using Advanced Machine Learning

3. Albumin
4. Alkaline Phosphatase
5. Age

Conclusion

The Random Forest model proved most effective in predicting liver cirrhosis. This project confirms the feasibility of using simple blood test data to flag at-risk patients before symptoms appear, helping in early intervention and better prognosis.

Future Scope

- Include more diverse datasets
- Use deep learning (e.g., neural networks)
- Real-time integration with hospital systems
- Explainable AI using SHAP or LIME

References

- UCI Machine Learning Repository
- Kaggle Notebooks on Liver Disease
- Research papers on cirrhosis diagnosis
- Scikit-learn Documentation