Liver Cirrhosis Prediction Project

1. Introduction

This project focuses on predicting liver cirrhosis using machine learning techniques. Liver cirrhosis is a progressive liver disease that can lead to liver failure. Early prediction is crucial for timely intervention and treatment.

2. Dataset Description

The dataset includes features such as age, gender, total bilirubin, direct bilirubin, alkaline phosphatase, alanine aminotransferase (ALT), aspartate aminotransferase (AST), total proteins, albumin, and the albumin-globulin ratio. The target variable indicates whether the patient has liver cirrhosis or not.

3. Data Preprocessing

The preprocessing steps include handling missing values, encoding categorical variables, scaling numerical features, and splitting the data into training and test sets. Visualizations such as histograms and correlation heatmaps were used for data exploration.

4. Model Building & Evaluation

Various classification algorithms were applied, including Logistic Regression, Random Forest, Support Vector Machine, and XGBoost. The models were evaluated using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Cross-validation and hyperparameter tuning were also employed to improve model performance.

5. Results

The Random Forest classifier achieved the highest performance with an accuracy of around 85%. Feature importance analysis revealed that bilirubin levels and liver enzyme levels were key predictors.

6. Conclusion

Machine learning can be effectively used to predict liver cirrhosis, aiding in early diagnosis. Future improvements could include using more advanced models and incorporating additional clinical features. Interpretability tools like SHAP can also help in understanding model decisions, which is important in medical applications.