Liver Cirrhosis Prediction Report

1. Abstract

This project applies machine learning techniques to predict liver cirrhosis based on clinical data. Using features such as bilirubin levels, liver enzymes, and albumin, the model aims to identify patients at risk of liver cirrhosis early. A variety of models were tested, with Random Forest achieving the highest accuracy.

2. Introduction

Liver cirrhosis is a progressive and often fatal condition that damages liver tissues. Early detection is crucial for effective treatment. This project uses machine learning algorithms to predict the presence of liver cirrhosis from patient medical records, improving the chances of timely diagnosis.

3. Dataset Description

The dataset consists of features such as age, gender, total and direct bilirubin, liver enzyme levels (ALT, AST), total proteins, albumin, and the albumin-globulin ratio. The target variable indicates whether liver cirrhosis is present.

4. Data Preprocessing

The data underwent cleaning processes including handling missing values, encoding categorical variables, and normalization. The dataset was then split into training and test sets for model evaluation.

5. Model Building

Several models were trained including Logistic Regression, Random Forest, Support Vector Machine, and XGBoost. Each model was evaluated using cross-validation and performance metrics to determine its predictive power.

6. Evaluation Metrics

Model performance was assessed using accuracy, precision, recall, F1-score, and ROC-AUC. These metrics helped in identifying the most effective model for predicting liver cirrhosis.

7. Results and Discussion

The Random Forest model achieved the highest performance with an accuracy of 85%. Important features included bilirubin levels and liver enzyme values. Limitations include the size and quality of the dataset.

8. Conclusion

This project demonstrates that machine learning can successfully predict liver cirrhosis from clinical data. Further improvements could include more complex models, larger datasets, and integration with clinical decision systems.

9. References

1. UCI Machine Learning Repository: Indian Liver Patient Dataset

2. Scikit-learn Documentation

3. Kaggle Datasets

10. Appendix

Code, plots, and additional results can be found in the attached notebook or GitHub repository.