Project Report

Title:

Revolutionizing Liver Care: Predicting Liver Cirrhosis Using Advanced Machine Learning Techniques

1. Abstract

Liver cirrhosis is a life-threatening condition caused by prolonged liver damage. Early detection is critical for successful treatment. This project presents a machine learning-based approach to predict liver cirrhosis from clinical data. Various classification models were trained and compared to identify the best-performing algorithm. The model provides accurate predictions, assisting in faster diagnosis and supporting healthcare professionals. The project demonstrates the application of artificial intelligence in the medical field and its potential to improve patient care.

2. Objectives

- Predict liver cirrhosis using machine learning.
- Preprocess and analyze clinical data.
- Train and evaluate multiple classification models.
- Choose the model with the best accuracy and reliability.
- Demonstrate how AI supports early diagnosis in healthcare.

3. Tools and Technologies Used

- Language: Python
- Platform: Jupyter Notebook / Google Colab
- Libraries:
  - Pandas, NumPy (Data Processing)
  - Matplotlib, Seaborn (Data Visualization)
  - Scikit-learn, XGBoost (Modeling & Evaluation)
- Algorithms: Logistic Regression, Random Forest, Support Vector Machine, XGBoost
- Optional Deployment: Flask / Streamlit
- Dataset Source: UCI Machine Learning Repository – Indian Liver Patient Dataset

4. Dataset Description

The dataset contains medical records of 583 individuals.
- Features: Age, Gender, Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase, SGPT, SGOT, Total Proteins, Albumin, A/G ratio
- Target: Liver Patient (1 = Yes, 0 = No)

Preprocessing Steps:
- Null value handling
- Label encoding (e.g., Gender)
- Feature scaling
- Splitting into train and test sets (80:20 ratio)

5. Methodology

1. Data Collection: Dataset imported from UCI repository.
2. Data Cleaning & Preprocessing:
   - Checked for missing values
   - Encoded categorical data
   - Scaled features using standard scaler
3. Exploratory Data Analysis:
   - Used correlation heatmaps and plots
   - Identified influential features
4. Model Training:
   - Applied classification algorithms
   - Split data into training and test sets
5. Model Evaluation:
   - Compared models using accuracy, precision, recall, F1-score
   - Selected the best performing model
6. Optional Deployment:
   - Developed a simple UI using Streamlit for real-time prediction

6. Results

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	76%	74%	78%	76%
Random Forest	85%	84%	86%	85%

Support Vector Machine	80%	79%	82%	80%
XGBoost	87%	86%	89%	87%

7. Outcome

- A highly accurate liver cirrhosis prediction model was built.
- XGBoost provided the best results with 87% accuracy.
- Data preprocessing and visualization helped understand the medical data.
- The model can assist doctors in making early and informed decisions.
- Demonstrated real-world applicability of ML in healthcare diagnosis.

8. Learnings

- Applied end-to-end machine learning pipeline.
- Improved skills in data preprocessing and visualization.
- Understood the impact of different ML models in healthcare.
- Learned performance evaluation using classification metrics.
- Optionally, gained exposure to simple ML model deployment.

9. Conclusion

This project highlights how machine learning can be applied to healthcare to detect liver cirrhosis at an early stage. With the help of clinical data and ML models, an intelligent diagnostic tool was developed. Among the evaluated algorithms, XGBoost delivered the highest accuracy. The project proves the effectiveness of AI in disease prediction and its potential to support clinical decisions.

10. References

- UCI Machine Learning Repository – Indian Liver Patient Dataset
- Scikit-learn Documentation
- XGBoost Documentation
- Research papers and tutorials on ML in healthcare
- Kaggle discussions and notebooks for liver prediction