

Problem Planning and Understanding:

1. Problem Statement:

Liver cirrhosis is a progressive condition that often remains undetected until its advanced stages, leading to severe health complications and fatalities. The lack of early diagnosis and predictive tools limits timely intervention. This project aims to build an intelligent ML-based system to predict liver cirrhosis early, enabling better clinical decisions and improved patient outcomes.

2. Project Objectives:

- To analyze patient health data and identify significant predictors of liver cirrhosis.
- To develop a machine learning model for accurate prediction of liver cirrhosis.
- To evaluate and compare various ML algorithms for performance and accuracy.
- To deploy the best-performing model through a user-friendly interface for healthcare professionals.

3. Project Timeline & Phases:

Phase	Time (Duration)	Activities
Phase 1: Research & Planning	Week 1	Literature review, finalize dataset, define evaluation metrics
Phase 2: Data Handling	Week 2	Data collection, cleaning, preprocessing
Phase 3: Exploratory Analysis	Week 3	EDA, visualization, correlation analysis
Phase 4: Model Development	Weeks 4–5	Train/test split, try models like Logistic Regression, Random Forest, XGBoost, SVM
Phase 5: Model Evaluation	Week 6	Evaluate models using accuracy, precision, recall, F1-score, ROC-AUC
Phase 6: Optimization & Tuning	Week 7	Hyperparameter tuning, cross-validation
Phase 7: Final Integration	Week 8	Build interface (Streamlit, Flask), write final report, documentation

4. Dataset:

- **Name:** Indian Liver Patient Dataset (ILPD)
- **Source:** UCI Machine Learning Repository or Kaggle
- **Attributes:** Age, Gender, Total Bilirubin, Direct Bilirubin, Alkaline Phosphatase, SGPT, SGOT, Total Proteins, Albumin, etc.

5. Tools & Technologies:

- **Languages:** Python
- **Libraries:** pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost
- **ML Models:** Logistic Regression, Random Forest, XGBoost, SVM
- **Deployment:** Streamlit or Flask (optional)
- **Version Control:** Git & GitHub
- **Documentation:** Jupyter Notebooks, PDF Report

6. Evaluation Metrics:

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

7. Risk Management:

Risk	Mitigation Strategy
Imbalanced dataset	Use SMOTE or class weighting
Overfitting	Regularization, cross-validation
Data quality issues	Robust preprocessing and outlier handling

8. Expected Outcome:

- A high-accuracy ML model that can predict liver cirrhosis.
- Insights into critical features influencing liver disease.
- A basic interface for healthcare professionals to input values and get predictions (optional).