

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

1. Introduction

1.1. Project Overview

This project aims to build a machine learning model to predict liver cirrhosis based on various patient features. By analysing these features, the model will classify patients into risk categories, aiding in early diagnosis and treatment.

1.2. Purpose

This project aims to develop a machine learning-based predictive model for early detection of liver cirrhosis using clinical and biochemical data. By applying advanced algorithms and performance tuning, the model identifies patients at risk with high accuracy. The best-performing model is deployed using Flask, enabling integration into real-time diagnostic systems. This approach supports faster, data-driven decisions in liver healthcare and improves patient outcomes through early intervention.

2. Ideation Phase

2.1. Problem Statement

The goal is to classify patients' risk levels for liver cirrhosis based on their medical data. Accurate prediction will support better management and early intervention for liver health.

2.2. Project Proposal (Proposed Solution)

The solution involves developing several machine learning models to predict liver cirrhosis. We will select and optimize the best model based on performance metrics to achieve the highest accuracy.

2.3. Initial Project Planning

Initial planning included setting up the project environment, defining objectives, and outlining the workflow for data collection, preprocessing, model development, and evaluation.

3. Data Collection and Preprocessing Phase

3.1. Data Collection Plan and Raw Data Sources Identified

The dataset for this project was sourced from Kaggle, containing patient data relevant to liver cirrhosis prediction (Dataset link: <https://www.kaggle.com/datasets/bhavanipriya222/liver-cirrhosis-prediction>).

3.2. DataQualityReport

- **DataShape:** Thedatasetinitiallycomprised[numberofrows,numberofcolumns]rows and columns.
- **MissingValues:**Handledbydroppingrowswithmissingvalues.

3.3. DataExplorationandPreprocessing

- **UnivariateAnalysis:**Histogramswereplottedfornumericalfeatures.
- **BivariateAnalysis:**Scatterplotsandpairplotsexploredrelationshipsbetweenfeatures.
- **OutlierHandling:**OutliersweredetectedandmanagedusingtheIQRmethod.

4. ModelDevelopmentPhase

4.1. FeatureSelectionReport

Featuresrelevanttolivercirrhosispredictionwereselected,anddatascalingwasappliedto standardize the input.

4.2. ModelSelectionReport

- **ModelsTested:**NaiveBayes,RandomForest,LogisticRegression,RidgeClassifier,Support Vector Classifier, KNN, XG Boost.
- **EvaluationMetrics:**Accuracy,ConfusionMatrix,ClassificationReport.

4.3. InitialModelTrainingCode,ModelValidationandEvaluationReport

- **Code:**Modeltrainingandevaluationstepswereimplementedforeachalgorithm.
- **Validation:**Modelswerevalidatedusingatestset,withperformancemetricsrecorded.The KNN model achieved the highest accuracy of 86.32%.

5. ModelOptimizationandTuningPhase

5.1. HyperparameterTuningDocumentation

- **KNN:**Optimizedbytuningthenumberofneighborsanddistancemetrics.
- **XGBoost:**Hyperparameterstunedforlearningrate,maxdepth,andn_estimators.

5.2. PerformanceMetricsComparisonReport

Model	Accuracy	Precision	Recall	F1-Score
NaiveBayes	35.79%	0.00	0.00	0.00
RandomForest	35.79	0.00	0.00	0.00
LogisticRegressionCV	81.58%	91.80	79.43%	86.49
RidgeClassifier	84.21%	93.44	83.82	88.37
SupportVectorClassifier	35.79%	0.00	0.00	0.00
LogisticRegression	79.47%	91.80	79.43	85.58
KNN	86.32%	94.26	85.82	89.84
XGBoost	35.79%	3.28	50.00	6.15

5.3. Final Model Selection Justification

The K-Nearest Neighbors (KNN) model was selected as the final model due to its superior accuracy of 86.32%. KNN excelled in handling complex data relationships and demonstrated the best performance in terms of precision, recall, and F1 score. This makes it a robust choice for predicting liver cirrhosis, aligning well with the project's goals.

6. Results

6.1. Output Screenshots

The source code and output screenshots are available in the accompanying files.

7. Advantages & Disadvantages

- **Advantages:** High accuracy, effective at handling local data variations, robust performance.
- **Disadvantages:** Can be computationally intensive, requires careful tuning.

8. Conclusion

The project successfully developed a machine learning model to predict liver cirrhosis with high accuracy. The KNN model, after hyperparameter tuning, provided the best results and was chosen for its robustness.

9. Future Scope

- Further data collection to include more features and increased dataset size.

- Exploration of additional features and engineering techniques.
- Experimentation with deep learning models to potentially outperform traditional models.
- Integration with a real-time predictions system for practical deployment.

10. Appendix

10.1. Source Code

Code File: model.ipynb

10.2. GitHub & Project Demo Link

GitHub Repository:

https://github.com/SURYA22508/Liver_Cirrhosis_prediction/tree/main

Demo Link:

https://github.com/SURYA22508/Liver_Cirrhosis_prediction/tree/main/demo%20video