DataCollectionandPreprocessingPhase

Date	16th may 2025
TeamID	LTVIP2025TMID60515
Project Title	Revolutionizing Liver Care: Predicting Liver CirrhosisUsingAdvancedMachineLearning Techniques.
Maximum Marks	2 Marks

DataCollectionPlan &RawData SourcesIdentification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuringmeticulousdatacurationandintegrityforinformeddecision-makingineveryanalysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	Themachinelearningprojectaimstopredictlivercirrhosisbased on patient information. Using a dataset with features such as age, gender, medical history, lab results, and lifestyle factors, the objectiveistobuildamodelthataccuratelypredictsthelikelihood of liver cirrhosis, facilitating early detection and informed decision-making in healthcare.
Data Collection Plan	• Searchfordatasetsrelatedtolivercirrhosis,patienthealth records, and medical history. • Prioritizedatasetswithdiversedemographicandmedical information.
RawDataSources Identified	Therawdatasourcesforthisprojectincludedatasetsobtainedfrom Kaggle, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as age, gender, medical history, lab results, and lifestyle factors for machine learning analysis.

RawDataSourcesTemplate

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises patient details(age,gender, etc.), medical metrics, and liver cirrhosis outcomes.	https://www.kaggle.com/datasets/bhavanipriya222/liver-cirrhosis-prediction	XLSX	47 kB	Public