Data Collection and Preprocessing Phase

Date	03-10-2024
Team ID	LTVIP2024TMID24892
Project Title	Liver Patient Identification – prediction of liver patient
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification :

A data collection plan for liver patient identification involves gathering relevant data from various sources to identify potential liver patients. Raw data sources can include medical claims data, laboratory results. Additionally, data from wearable devices, mobile apps, and patient-reported outcomes can also be used. The data collection plan should ensure that the data is accurate, complete, and relevant to the research question.

Data Collection Plan

Section	Description
Project Overview	The objective of this machine learning project is to identify the persons who have liver disease based on the laboratories records.
Data Collection Plan	The data will be searched from the kaggle and other websites and indian liver patient websites.
Raw Data Sources Identified	The dataset has been collected from kaggle and from indian liver patient website.

Raw Data Sources

Source Name	Description	Location/URL	Format	Size	Access Permissions
ILPD (Indian Liver Patient Dataset)	The Indian Liver Patient Dataset (ILPD) is a valuable collection of medical records that includes 583 entries. It features ten key attributes age and gender, as well as crucial laboratory measurements such as total bilirubin, alkaline phosphatase.	https://archive.ics.uci.edu/dataset/225/ilpd+indian+liver+patient+dataset	CSV	23.4 KB	Public
Kaggle dataset	This dataset also contains the same common values as	https://www.kaggle.com/datasets/abhi8923shriv/liver-disease-patient-dataset	CSV	1.2 MB	Public