

Data Collection and Preprocessing Phase

Date	10/07/2024
Team ID	team-739866
Project Title	Revolutionizing Liver care : Predicting Liver cirrhosis using Advanced machine learning Techniques
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	<p>Project: Predicting Liver Cirrhosis Disease using Machine Learning</p> <p>Objective:</p> <ol style="list-style-type: none"> 1. Develop a predictive model that can accurately identify patients at risk of developing liver cirrhosis. 2. Utilize a combination of demographic, medical history, laboratory results, and ultrasound findings to predict the likelihood of liver cirrhosis.

	<p>3. Improve early detection and diagnosis of liver cirrhosis, enabling timely interventions and potentially improving patient outcomes.</p> <p>4. Explore the relationships between various risk factors and liver cirrhosis to inform future research and clinical decision-making.</p>
Data Collection Plan	Collected the data from kaggle HealthCare dataset
Raw Data Sources Identified	<p>1. Electronic Health Records (EHRs)</p> <ul style="list-style-type: none"> - Source: Healthcare providers, hospitals, and clinics - Content: Patient demographics, medical history, medications, lab results, and clinical notes <p>2. Hospital Information Systems (HIS)</p> <ul style="list-style-type: none"> - Source: Hospital databases - Content: Patient admissions, diagnoses, treatments, outcomes, and discharge summaries <p>3. Laboratory Information Systems (LIS)</p> <ul style="list-style-type: none"> - Source: Laboratory databases - Content: Lab test results, including liver function tests, complete blood counts, and other relevant parameters <p>4. Radiology Information Systems (RIS)</p> <ul style="list-style-type: none"> - Source: Radiology departments

	<ul style="list-style-type: none"> - Content: Imaging reports, including liver scans, biopsies, and other relevant radiology results <p>5. Patient Questionnaires and Surveys</p> <ul style="list-style-type: none"> - Source: Patient self-reports - Content: Lifestyle habits, alcohol consumption, family history, and other relevant factors <p>6. Public Health Databases</p> <ul style="list-style-type: none"> - Source: National and international health organizations - Content: Aggregated data on liver disease prevalence, risk factors, and outcomes <p>7. Research Studies and Clinical Trials</p> <ul style="list-style-type: none"> - Source: Academic institutions, research centers, and clinical trial databases - Content: Data from studies on liver cirrhosis, including genetic and molecular markers <p>8. Healthcare Claims Data</p> <ul style="list-style-type: none"> - Source: Insurance companies and healthcare payers - Content: Patient outcomes, treatments, and resource utilization data
--	---

	These raw data sources will be used to extract relevant features and build a comprehensive dataset for training and validating the machine learning model.
--	--

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
HealthCare	Collection of patient details	Link: https://www.kaggle.com/datasets/bhavanipriya222/liver-cirrhosis-prediction	CSV	XX GB	Public
...