

Data Collection and Preprocessing Phase

Date	13 september 2024
Team ID	166
Project Title	Deep learning techniques for breast cancer prediction
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>Clinical data Histopathological data Imaging data Genomic data Lifestyle data Laboratory results Survival and recurrence data</p> <p>The objective of this project is to develop an efficient and accurate breast cancer detection system using deep learning techniques, specifically the EfficientNetB3 model, to classify medical images into benign and malignant categories</p>

Data Collection Plan	From kaggle account
Raw Data Sources Identified	From medical reports and other datasources

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Google drive	https://drive.google.com/file/d/1AvRp8hOK-O76WIFvPj_rk10fAZbt8nWH/view?usp=sharing	CSV	82 B	Public
Dataset 2	Google drive	https://drive.google.com/file/d/1AvRp8hOK-O76WIFvPj_rk10fAZbt8nWH/view?usp=sharing	CSV	6.48kB	Private (with access)
...