

## Data Collection and Preprocessing Phase

Date	11 July 2024
Team ID	SWTID1720011518
Project Title	WCE curated Colon Disease Classification using Deep Learning
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	Healthcare professionals struggle with accurately diagnosing colon diseases due to inconsistent imaging data and complex medical records. Developing a deep learning model to analyze these data sources can enhance diagnostic precision, support early detection, and improve treatment planning and patient outcomes.
Data Collection Plan	The dataset which is used is a public dataset. It is available on Kaggle with the name “Curated Colon Dataset for Deep Learning”
Raw Data Sources Identified	The dataset contains medical imaging data in a curated manner. This initiative intends to accurately categorize diverse on these orders through the analysis of patient data and colonoscopy images therefore facilitating early identification treatment planning and improved patient outcomes

### Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle	WCE Curated Colon Disease Dataset Deep Learning	<a href="https://www.kaggle.com/datasets/francismon/curated-colon-dataset-for-deep-learning">https://www.kaggle.com/datasets/francismon/curated-colon-dataset-for-deep-learning</a>	Images	2 GB	Public