

Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720333657
Project Title	Wce Curated Colon Disease Classification Using Deep
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	Existing colonoscopy-based diagnostics, dependent on physician interpretation, exhibit limitations in accuracy and consistency. This can lead to missed diagnoses, unnecessary procedures, and workflow inefficiencies. We propose a novel deep learning system for colon disease classification to address these challenges, aiming to enhance diagnostic precision, streamline clinical workflows, and ultimately improve patient care and cost-effectiveness.
Data Collection Plan	<ul style="list-style-type: none"> • Existing Datasets: Look for Kaggle datasets with colonoscopy-related images (e.g., histology) or anonymized data (e.g., polyp labels). • Partnerships: Collaborate with medical institutions for access to anonymized data or synthetic images.

	<ul style="list-style-type: none"> • Public Repositories: Explore resources like TCIA for potentially relevant datasets.
Raw Data Sources Identified	<ul style="list-style-type: none"> • Kaggle Datasets: Look for anonymized colonoscopy data (e.g., polyp labels) or related images (e.g., histology). • Medical Collaborations: Partner with hospitals for access to anonymized colonoscopy data. (Benefits: relevant data, diverse cases. Challenges: approvals, anonymization) • Public Repositories: Explore TCIA ([Cancer Imaging Archive]) or NIH trials for potentially relevant datasets, keeping in mind data usage licenses and ethics.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	Data is comprised of colon disease images (Normal, Ulcerative Colitis, Polyps) from Kaggle.	https://www.kaggle.com/datasets/francismon/curated-colon-dataset-for-deep-learning/data	Image	2 GB	Public