

Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1719937289
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	WCE Curated Colon Disease Classification Using Deep Learning analysis colonoscopy images and patient records, this project aims to accurately classify various colon diseases, aiding in early detection, treatment planning, and improving patient outcomes. It assists healthcare professionals in diagnosing colon diseases. Enhances diagnostic accuracy, streamline treatment decisions, and improve patient care by providing timely and accurate disease classification.
Data Collection Plan	<ul style="list-style-type: none"> Look for datasets concerning curated colon disease prediction Give priority to datasets that include diverse demographic information.
Raw Data Sources Identified	<p>The primary data sources for this project are datasets acquired from Kaggle which is renowned platforms for data science competitions and repositories.</p> <p>The provided sample data represents a subset of the collected information, which is classified into 4 subclasses,</p> <ol style="list-style-type: none"> Normal

	<ol style="list-style-type: none"> 2. Ulcerative 3. Esophagitis 4. Colon Polyps
--	--

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle dataset	The dataset comprises, four classes of colon disease (normal, ulcerative, esophagitis, polyps)	https://www.kaggle.com/datasets/francismon/curated-colon-dataset-for-deep-learning	JPG images	2GB	Public