Data Collection and Preprocessing Phase

Date	17 July 2024
Team ID	SWTID1720011518
Project Title	WCE Curated Colon Disease Classification
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

The model requires labelled data consisting of Colonoscopy Images and associated labels for each image indicating the presence or absence of colon disease and the specific disease type. Such data can be collected from medical institutions, public databases and research collaborators.

Data Collection Plan Template

Section	Description
Project Overview	The project aims to classify the WCE Curated colon diseases based on Wireless Capsule Endoscopy (WCE) images. The project aims to reduce time in detection and classification of the disease.
Data Collection Plan	Data will be collected from an already existing dataset on Kaggle.
Raw Data Sources Identified	Data can be collected from medical institutions, public databases and research collaborators.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
WCE Curated Colon Disease Dataset Deep Learning	The dataset is taken from J. Silva, A. Histace, O. Romain, X. Dray and B. Granado, "Toward embedded detection of polyps in WCE images for early diagnosis of colorectal cancer", International Journal of Computer Assisted Radiology and Surgery, vol. 9, no. 2, pp. 283-293, 2013. DOI:10.1007/s11548-013-0926-3.	https://www.kaggle.com/datasets/francismon/curated-colon-dataset-for-deep-learning	Images	2 GB	Public