

Data Collection and Preprocessing Phase

Date	14 June 2025
Team ID	SWTID1749876754
Project Title	SynapseScan: AI Driven Classification of Ovarian Cancer Variants
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report

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

Section	Description
Project Overview	SynapseScan represents a groundbreaking initiative focused on the AI-driven classification of ovarian cancer variants, utilizing the power of transfer learning. This project aims to enhance the accuracy and efficiency of ovarian cancer diagnosis by leveraging pre-trained models and transfer learning techniques to analyze medical imaging data, such as histopathological images or radiological scans, facilitating early detection, personalized treatment plans, and accelerated research in ovarian cancer.
Data Collection Plan	<ul style="list-style-type: none">Search for datasets containing histopathological images of ovarian cancer tissues and variants.

	<ul style="list-style-type: none">• Prioritize datasets with diverse ovarian cancer types and staging information.• Collect radiological imaging data (CT scans, MRI, ultrasound) related to ovarian cancer.• Gather datasets with molecular and genetic profiling information for ovarian cancer variants.• Ensure datasets include proper medical annotations and expert classifications.
Raw Data Sources Identified	The raw data sources for this project include medical imaging datasets obtained from Kaggle and specialized medical databases. The datasets encompass histopathological images, radiological scans, and clinical data related to ovarian cancer variants. These sources provide comprehensive information including tissue morphology, imaging characteristics, patient demographics, and cancer staging details essential for AI-driven classification and transfer learning implementation.

Raw Data Sources Report

Source Name	Description	Location/URL	Format	Size	Access Permissions
Ovarian Cancer Subtype Classification	The dataset comprises histopathological images of ovarian cancer tissue samples with different cancer variants.	https://www.kaggle.com/datasets/sunilthite/ovarian-cancer-classification-dataset	PNG	4.04 GB	Public