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 Exploration and Preprocessing Report

The Data Quality Report will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Data imbalance of the number of pictures in each class.	Moderate	Used image multiplication in Python to have same number of images in each class.
Kaggle Dataset	There were some non-numeric characters in the csv file which made it difficult to process the data.	Moderate	Had to manually remove those non-numeric characters from the csv files.