

Data Collection and Preprocessing Phase

Date	05 July 2025
Team ID	SWTID1749835721
Project Title	HematoVision - Blood Cell Classification using Transfer Learning
Maximum Marks	2 Marks

Data Quality Report:

The Data Quality Report summarizes issues encountered during loading and preparing the blood cell image dataset. It outlines the nature and severity of the problems and the applied solutions to ensure data integrity for effective training and evaluation.

Data Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
Google drive upload	Missing class folders (e.g., NEUTROPHIL not found)	High	Skipped folders with warnings; ensured core classes exist
Google drive upload	Corrupted or unreadable image files	Moderate	Verified with PIL; excluded from dataset.
Google drive upload	Class imbalance across blood cell categories	Moderate	Addressed through augmentation and careful sampling.

Google drive upload	Uniform image sizes making bivariate analysis flat	Low	Noted in analysis section; visualized using alternative plots.
Google drive upload	Images lack embedded labels or metadata	Low	Used directory names for label inference.

This report ensures that despite limitations in the dataset, appropriate preprocessing and validation steps were taken to maintain model input quality.