Data Collection and Preprocessing Phase

Date	24 July 2024
Team ID	**SWUID20240034367**
Project Title	**Anemia-Sense-Leveraging-Machine-Learning-For-Precise-Anemia-Recognitions-using-python**
Maximum Marks	2 Marks

Data Quality Report

Data Source	Data Quality Issue	Severity	Resolution Plan
Dataset	Missing labels for some images.	High	Implement a manual review or crowdsource labeling for missing entries. Ensure all images have correct disease labels before training.
Dataset	Images with varying resolutions.	Moderate	Standardize image resolutions through preprocessing steps (e.g., resizing) to ensure uniform input for the model.
Dataset	Imbalanced dataset with some mismatched anemia disease labels.	High	Use data augmentation techniques to balance the dataset or collect additional images for mismatched labels.
Dataset	Presence of noisy or corrupted images.	Moderate	Apply image quality checks and remove or replace corrupted images from the dataset. Use denoising during preprocessing.

Dataset	Inconsistent image formats or color channels.	Low	Convert all images to a consistent format (e.g., RGB) and ensure uniform preprocessing steps.