Data Collection and Preprocessing Phase

Date	4 July 2025
TeamID	SWTID1750180871
Project Title	Mangonet: A Vgg16-Based Neural Network For Mango Classification
Maximum Marks	2 Marks

Data Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
Custom Mango Dataset (e.g., from Kaggle / field-collected)	Class imbalance: Some mango varieties (like 'Langra' or 'Fajri') have fewer images compared to others.	Moderate	Apply data augmentation techniques to balance the dataset.
Custom Mango Dataset	Missing or corrupted images that fail to load (due to format issues).	High	Use try-except to skip and log corrupted images.
Custom Mango Dataset	Inconsistent image sizes (varies widely across classes).	Moderate	Resize all images to a fixed dimension (224x224) during preprocessing.
Custom Mango Dataset	No image labels in metadata; class info only from folder names.	Low	Use folder names to assign labels while loading data.