

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 Collection Plan & Raw Data Sources Identification Report:

Data Collection Plan:

Section	Description
Project Overview	This machine learning project aims to classify mango varieties using image data. By collecting and preprocessing a dataset of labeled mango images, we train a convolutional neural network (CNN) to automatically identify the mango type based on its visual features. The objective is to develop a high-accuracy mango recognition system that can be deployed as a web application to assist in agricultural quality control and classification.
Data Collection Plan	<p>Search for mango image datasets containing multiple labeled varieties such as Langra, Sindhri, Anwar Ratool, etc.</p> <p>Prioritize datasets with:</p> <ul style="list-style-type: none"> • High-resolution images • Balanced class distributions (or use augmentation for imbalance)

	<ul style="list-style-type: none"> • Clear image labeling (folder names or metadata) <p>Perform exploratory data analysis to assess dataset quality and variability</p> <p>Resize images and apply normalization for model compatibility</p>
--	---

Raw Data Sources Report:

Source Name	Description	Location / URL	Format	Size	Access Permissions
Custom Kaggle Dataset / Open Mango Image Repository	Contains images of various mango varieties organized in folders for classification tasks. Classes include Langra, Sindhri, Chaunsa, Fajri, and others.	https://www.kaggle.com/datasets/saurabhshahane/mango-varieties-classification	JPG / PNG	~300 MB	Public