Data Collection and Preprocessing Phase

Date	5 Feb 2026
Student Name	Rounak Pratap Gajbar
Project Title	greenclassify: deep learning-based approach for vegetable image

Data Collection Plan & Raw Data Sources Identification Template

Data Collection Plan Template

Section	Description
Project Overview	The Greenclassify project aims to develop a deep learning model for accurate and efficient vegetable image classification. This will automate tasks currently performed manually in agricultural, processing, and retail settings, leading to improved efficiency, reduced costs, and minimized waste.
Data Collection Plan	The primary data source for this project will be a publicly available dataset found on Kaggle (linked below), supplemented by additional images collected to address any class imbalances or improve model robustness.
Raw Data Sources Identified	Kaggle Datasets

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Vegetable Dataset	A dataset containing images of 15 common vegetable types, split into training, validation, and testing sets.	https://www.kaggle.com/datasets/misrakahmed/vegetable-image-dataset	Image (JPG)	21,000 images	Public