Data Collection and Preprocessing Phase

Date	20 April 2025
Student Name	Hrituraj Shashikant Narvekar
Project Title	GreenSnap: A Vegetable Classifier
Maximum Marks	2 Marks

Data Collection Plan

Section	Description
Project Overview	This deep learning project focuses on classifying images of 15 types of vegetables using Convolutional Neural Networks (CNNs) .The objective is to uncover hidden patterns and visual cues that distinguish each type, contributing to better vegetable identification in kitchens and farms.
Data Collection Plan	The dataset has been sourced from a ZIP file provided by SmartInternz, which includes categorized images in subdirectories named after: each vegetable type. Additional reference images were accessed from publicly available sources such as Kaggle and Open Food Facts to enhance variability and robustness.
Raw Data Sources Identified	The raw data includes SmartInternz provided images saved in structured subdirectories, supplemented by publicly available datasets for training and validation purposes.

Raw Data Sources

Source Name	Description	Location/URL	Format	Size	Access Permissions
Field Captured Images	Manually photographed images taken in natural environments, used for supplementing the dataset.	Local Storage	JPG/PNG	~100 MB	Private
Kaggle - Mushroom Image Dataset	Supplementary dataset with additional labeled mushroom images.	https://www.kaggle.com/datasets/misrakahmed/vegetable-image-dataset/data	JPG	~ 111 MB	Public
Wikimedia	The images in this dataset were collected by us from vegetable farm and market for a project.	https://en.wikipedia.org/w/index.php?search=vegetables&title=Special%3ASearch&profile=advanced&fulltext=1&ns0=1	JPG/PNG	~10 MB	Public