Data Collection and Preprocessing Phase

Date	6 Feb 2026
Student Name	Amulya Shamkumar Tikare
Project Title	Uncovering The Hidden Treasures Of The Mushroom Kingdom: A Classification Analysis
Maximum Marks	2 Marks

Data Collection Plan

Section	Description
Project Overview	This deep learning project focuses on classifying images of three types of mushrooms—Boletus, Lactarius, and Russula—using Convolutional Neural Networks (CNNs). The objective is to uncover hidden patterns and visual cues that distinguish each type, contributing to better mushroom identification in the wild.
Data Collection Plan	The dataset has been sourced from a ZIP file provided by the SmartInternz, which includes categorized images in subdirectories named after each mushroom type. Additional reference images were accessed from publicly available sources such as Wikimedia and Kaggle 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
SmartInterz Provided Dataset	Curated image dataset provided by SmartInternz, containing Boletus, Lactarius, and Russula images in separate subdirectories.	https://drive.google.com/drive/folders/1WHjhoYnyrltQWJ_TYI5xM_5dTyIzByo3	ZIP File	~ 175 MB	Public

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?search=Mushrooms+images++Boletus%2C+Lactarius+%26+Russula	JPG	~ 111 MB	Public