Data Collection and Preprocessing Phase

Date	03 August 2025
Student Name	Neha Ranjit Nalage
Project Title	Uncovering The Hidden Treasures Of The Mushroom Kingdom: A Classification Analysis
Maximum Marks	2 Marks

Data Source	Data Quality Issue	Severity	Resolution Plan
Dataset	Image Variation	High	Collect images from diverse sources (different cameras, lighting conditions, angles). Implement data augmentation techniques (rotation, scaling, cropping) during preprocessing.
Dataset	Occlusion	Moderate	Include images with partial occlusion, and/or train the model to be robust to it.

Dataset	Insufficient Resolution	Moderate	Establish a minimum resolution threshold for images. Use superresolution techniques, if feasible, to enhance the resolution of some images.

Data Quality Report Template

Dataset	Unbalanced Classes	High	Employ stratified sampling to ensure proportional representation of each mushroom species. Use data augmentation for minority classes. Explore the use of weight loss functions during training.