

Data Collection and Preprocessing Phase

Date	08 July 2024
Team ID	SWTID1720201335
Project Title	Rice Type Classification Using CNN
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	This project focuses on classifying different types of rice grains using deep learning techniques. By leveraging a pre-trained MobileNetV2 model and applying various image preprocessing methods such as resizing, normalization, and data augmentation, we aim to accurately identify rice types from images. The dataset includes multiple rice varieties, and the model is trained to achieve high accuracy and robust performance in classification tasks
Data Collection Plan	<ul style="list-style-type: none"> Look for datasets concerning rice type classification Give priority that include images of different rice types.

Raw Data Sources Identified	<p>The primary data sources for this project are datasets acquired from kaggle which is renowned platforms for data science competitions and repositories.</p> <p>The provided sample data represents a subset of the collected information, which is classified into five subclasses,</p> <ol style="list-style-type: none"> 1. Arborio 2. Basmati 3. Ipsala 4. Jasmine 5. karacadag
-----------------------------	--

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle dataset	The dataset comprises, five classes of rice (arborio, basmati, ipsala, jasmine, karacadag)	https://www.kaggle.com/datasets/muratkokludataset/rice-image-dataset	JPG	230 MB	Public