

Data Collection and Preprocessing Phase

Date	19 July 2024
Team ID	SWTID1720110092
Project Title	Beneath the Waves: Unraveling Coral Mysteries Through Deep Learning
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	The coral image classification project aims to develop an AI model capable of accurately distinguishing between healthy and bleached coral reefs. This model will contribute to conservation efforts by providing insights into coral health based on image analysis.
Data Collection Plan	The dataset will be collected from publicly available sources on Kaggle, specifically focusing on images of healthy and bleached coral reefs.
Raw Data Sources Identified	Train and Test dataset from Kaggle contains images categorized into healthy and bleached corals used for training a machine learning model.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Train Dataset	Dataset containing training images of healthy and bleached coral reefs for machine learning model development.	https://www.kaggle.com/datasets/gauravduttakiit/corals-image-classification/data?select=train	JPEG images	85.29 kB	Public
Test Dataset	Dataset containing test images of healthy and bleached coral reefs for model evaluation.	https://www.kaggle.com/datasets/gauravduttakiit/corals-image-classification/data?select=test	JPEG images	21.42 kB	Public