Data Collection and Preprocessing Phase

Date	15 November 2024
Team ID	739653
Project Title	Fireguardian yolov8 Empowered wildfire smoke surveillance
Maximum Marks	2 Marks

Data Quality Report:

A Data Quality Report evaluates the dataset used in the project to ensure it meets the requirements for building a reliable and accurate system. Below is a structured report for assessing the quality of the data in this context

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missing or incomplete image metadata (e.g., geotags, timestamps).	Moderate	Use data augmentation or imputation techniques to generate missing metadata where possible.
Satellite Imagery	Variations in resolution and scale across image datasets.	High	Normalize image sizes by resizing them to a consistent resolution suitable for YOLOv8 model input (e.g., 640x640).
Drone Footage	Poor lighting or environmental conditions (e.g., fog, low visibility).	High	Apply image preprocessing techniques such as histogram equalization or contrast adjustments to enhance clarity.
Ground Camera Images	Presence of noisy or irrelevant data (e.g., non-smoke objects misclassified as smoke).	High	Perform manual or automated data cleaning to filter out noisy and irrelevant samples using domain-specific rules.
All Datasets	Imbalanced data distribution between smoke and non-smoke labels.	High	Use oversampling (e.g., SMOTE) for minority classes or undersampling for majority classes to balance the dataset.