

Data Collection and Preprocessing Phase

Date	11 July 2024
Team ID	SWTID1720012105
Project Title	WarLens: Transfer Learning for Event Classification in Conflict Zones
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Dataset: Kaggle Dataset with 84,151,603 images	Incorrect Labels	High	Perform manual and automated validation of labels. Use a subset of images for manual verification and employ a model trained on a smaller verified dataset to predict and cross-check labels. Mislabeled images will be corrected or removed.
.....(same dataset as above)	Imbalanced Classes	Moderate	Use techniques like data augmentation to balance the class distribution. This can involve

			generating new images for underrepresented classes through transformations such as rotation, flipping, and cropping.
--	--	--	--