Data Collection and Preprocessing Phase

Date	15 JUNE 2025
Team ID	SWTID1749902640
Project Title	CRIME VISION:ADVANCED CRIME CLASSIFICATION USING TRANSFER LEARNING
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
/content/Train, /content/Test - Class: RoadAccidents	Class imbalance – some crime categories have significantly fewer images than others.	Moderate	Use **data augmentation** to synthetically increase samples for underrepresented classes, or apply **class weights** during training.
/content/Train, /content/Test - Class:	Low-resolution images (64x64) may not capture enough features for accurate classification.	High	Increase resolution (e.g., 128x128), or use **pretrained models** with higher feature sensitivity (e.g., EfficientNet).

NormalVideos			
/content/Train, /content/Test - Class: Explosion, Shooting, Abuse	Noisy or mislabeled data in folders may affect training accuracy.	Moderate	**Manually verify** or run a script to detect anomalies in label-folder mismatches and remove or correct bad samples.
All classes (Train/Test)	Image format inconsistency (e.g., PNG, JPG with alpha channel).	High	**Normalize image loading by converting all to RGB and resizing using image.load_img(..., color_mode='rgb').**
All classes	Overfitting due to small dataset size.	Low	Use dropout layers, data augmentation, or early stopping to prevent overfitting during training.
All classes	Uneven distribution between train/test splits across classes.	Moderate	**Use stratified split if possible or verify each class has representation in both sets.**