Data Collection and Preprocessing Phase

Date	15 June 2025
Team ID	SWTID1750006853
Project Title	ASL- Alphabet Image Recognition
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
ASL Alphabet Dataset	Class imbalance (some classes have more images than others)	Moderate	Use ImageDataGenerator with balanced batches or apply oversampling techniques.
ASL Alphabet Dataset	Background clutter and variation in lighting	Moderate	Apply image augmentation (e.g., flipping, shifting, brightness adjustment) to improve generalization.
ASL Alphabet Dataset	Noisy or low-quality images	Low	Apply optional denoising filters like cv2.fastNlMeansDenoisingColored on input.

ASL Alphabet Dataset	Redundant or duplicate images	Low	Use hashing or similarity metrics (e.g., SSIM) to detect and optionally remove duplicates.
ASL Alphabet Dataset	Images stored in nested folders	Low	Standardize and flatten the directory structure during data loading.
ASL Alphabet Dataset	No explicit validation/test split provided	Moderate	Manually split the dataset using train_test_split while ensuring class stratification.
ASL Alphabet Dataset	Unclear file naming for test images	Low	Use image path structure and manual labeling to ensure consistent labels for test data.