**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 24 March 2024 |
| Team ID | SWTID1750316859 |
| Project Title | ASL- Alphabet image Recognition |
| Maximum Marks | 2 Marks |

**Data Quality Report :**

This report outlines key data quality issues in the ASL Alphabet Dataset, including their severity and technical solutions. It helps improve the accuracy and reliability of the ASL recognition model. Ensuring data quality is crucial for effective sign language interpretation.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| ASL Alphabet Dataset | Class imbalance – certain letters (like 'J' and 'Z') have fewer samples than others. | Moderate | Apply data augmentation (rotation, flipping, zoom) to increase data for underrepresented classes. |
| ASL Alphabet Dataset | Variations in lighting and background across images. | High | Use preprocessing methods like normalization, background removal, and contrast adjustment to standardize image conditions. |
| Preprocessed Images | Incorrect or misaligned labels due to folder naming or data import issues. | High | mplement validation scripts to cross-check filenames and folder labels; manually inspect a sample for accuracy. |
| Validation/Test Dataset | Low variability – test set images are too similar to training data, reducing evaluation accuracy. | Moderate | Ensure a diverse and representative validation/test split by selecting images from different classes and lighting conditions. |
| Training Dataset | Overfitting due to overly similar images and lack of variation | Moderate | Introduce diverse augmentations during training and use dropout and regularization in the CNN model to improve generalization. |