Data Collection and Preprocessing Phase

Date	24 June 2025
Team ID	SWTID1750316859
Project Title	ASL- Alphabet image Recognition
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification:

Elevate your ASL recognition strategy with a detailed Data Collection Plan and Raw Data Sources Report, ensuring accurate curation of sign language image data for precise model training and informed decision-making throughout the application lifecycle.

Data Collection Plan Template

Section	Description
Project Overview	This project focuses on developing a system that accurately recognizes and interprets American Sign Language (ASL) alphabet signs from images using deep learning. The aim is to facilitate communication for individuals who are deaf or hard of hearing. The system is deployed using a Flask web application for real-time predictions.
Data Collection Plan	The dataset used for training the ASL image recognition model was collected from Kaggle, specifically the "ASL Alphabet" dataset published by grassknoted. The dataset includes labeled images of ASL alphabet hand gestures. Google Colab was used to access and process the dataset using Kaggle API credentials.
Raw Data Sources Identified	The raw data source is: ASL Alphabet Dataset on Kaggle. It contains labeled image data for each ASL alphabet sign (A–Z, plus "Nothing," "Space," and "Delete"), organized in separate folders.

	Images were downloaded, unzipped, and stored in the working directory for preprocessing and training.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
ASL Dataset	Contains labeled images of American Sign Language (ASL) alphabets (A–Z, space, delete, nothing) hand signs.	https://www.kaggle.com/datasets/grassknoted/asl-alphabet	Image	3 GB	Public
ASL Test Dataset	A separate dataset of test images used to validate the trained model's accuracy and predictions.	https://www.kaggle.com/datasets/grassknoted/asl-alphabet/test	Image	300 MB	Public