

## RESEARCH METHODOLOGY

Two different approaches were compared in recognizing sign language on word-level in this study. The first approach uses Mediapipe (Mediapipe, n.d.) pose and hands landmark coordinates and the second approach uses sequence of video frames. The structure plan of both approaches is as shown in figure 1.

As this task involves multi-class classification, data balance was first checked. Since all data presented was in video MPEG-4 Part 14 (MP4) format and due to restrictive memory capacity available, the study was conducted using a subset of 15 labels from each dataset. The detail of this step is further discussed under section 3.1. After splitting the data into training and test set at 0.8:0.2 ratio, 20% of the remaining training data was further split for validation purpose. Then, data augmentation was performed include rotation, horizontal flip, vertical flip and frames cropping to increase the diversity and size of the training data.

### Method 1: Mediapipe Landmark Coordinates

Method 1 uses MediaPipe landmark coordinates. The videos were then processed with Mediapipe Hand and Pose Landmarker (Mediapipe, n.d.) to locate the key points coordinates. The output was then extracted and stored in Numpy array format. To maintain an equal sequence length, only information from frame 5<sup>th</sup> to 150<sup>th</sup> was extracted from all videos in WLASL dataset while frame 5<sup>th</sup> to 200<sup>th</sup> for all videos in LSA64 dataset.

### Method 2: Frames Sequence

Method 2 uses frames sequence. After applying MediaPipe Hand and Pose Landmarker on the videos using black background, the video was stored as MP4 file. Due to limited memory capacity available, only 8 frames with 12 frame steps apart were extracted from each video from LSA64, whereas 6 frames with 8 frame steps apart for WLASL dataset, padding with empty frames was applied for video with shorter length.

Next, two base models were constructed for each method separately. Based on the model fitness evaluation on training and validation set, model tuning was performed as required. Lastly, each final model was tested on the test data to evaluate the model performance using accuracy, precision, recall and f1-score metrics. Lastly, the results of both methods were compared and discussed to determine the best approach.

Figure 1 Research Methodology Flowchart

