Forming a Project Idea

- Project idea: Building an AI that recognizes sign language and translate it to english text. It uses a webcam to take pictures of signs that don't require movement of the hand, such as the alphabet and the numbers.

1. Choice of dataset

- The dataset for the alphabet: https://www.kaggle.com/datasets/grassknoted/asl-alphabet
- Why we chose this dataset : it has a total of 87,000 images of the ASL alphabet and it seems to have different color of hands, which mean our model won't be too much biased.

- The dataset for the numbers : https://www.kaggle.com/datasets/lexset/synthetic-asl-numbers
- Why we chose this dataset : it has a total of 11,000 images of the ASL alphabet and it seems to have different color of hands, which mean our model won't be too much biased.

2. Methodology

a) We will use the kaggle datasets to train our model to differentiate between the background and the hand. The dataset we will use is feasible as it is already labeled, as we will be doing supervised learning. The test data in the dataset has high contrast and thus will be helpful to convert the targeted pixels into the required unit vectors. All the images in the dataset are of same dimensions, thus very little preprocessing will be required, and we can easily convert all the images into our required dimensions. The dataset also has a very diverse range (images of hands from different ethnicities), thus decreasing chances of any unintentional bias. Additionally, we will add images of asl with different backgrounds to improve the generalization capability of our model.
For preprocessing, we will be checking the image dimensions, and we will be changing the dimensions of the images to 64x64 pixels.
We will create various classes (36 classes- for 26 alphabets, and 10 numbers). Then we will train it to to differentiate between the different hand postures. After preprocessing the data, we will include appropriately scaled features to add non-linearity. During usage we will use OpenCV to get live cam feed, where it will capture the frame where we want to detect the hand in gestures.

b) **Machine learning model**

We will be using a classification algorithm such as the CNN algorithm, which is a popular algorithm used to process images in the form of grid pattern. For our model, we would use the loss function categorical cross entropy, since a lot of work has already been made by the community using this loss function. In the following research paper: https://core.ac.uk/download/pdf/55626122.pdf, ANN have been determined to be the best among 3 other algorithm for static hand movement. After looking at various code and articles about it, the best classification algorithm seems to be the CNN

algorithm. The other model that we considered is the spiking neural network (SNN) model, which is referenced in this article : https://arxiv.org/abs/2207.12559.

c) **Evaluation Metric**

Since we have a classification algorithm, we will use a confusion matrix to analyse how good is our model, .This allow us the see in a tabular way if our model classified each data correctly or not. This confusion matrix also gives us three other variables that help us know how good our model is, which are the accuracy, the precision and recall.

3. Application

We will integrate our model in a web-app. In our application, the user will input either a ASL letter in the alphabet or a number in ASL. For this to happen, we will have a webcam that will take a picture of the user's hand. As output, the model will tell the user what letter he signed. For example, if the user makes an "A" in ASL, the model will output the letter A in text.

An idea for a more fun and interactive application would be to have a learning application for ASL alphabet. Our application would give a countdown for the user to sign the letter and take the picture at the end of the countdown. As output, it would then translate what the user inputed and tell them if this answer was right or wrong. For example, if our application asks the user to sign an A, and our model detects that he actually signed an M, the application will then say that the user made an error, that he inputed M instead.