MAIS 202 - Project Deliverable 1 - Tracy - Annaliese - Avyan

Deliverable Description (Content of your Deliverable):

1. **Choice of dataset:**

This dataset contains over 27000 images of the alphabet signed in American Sign Language with each image sized at 512 x 512 pixels. Moreover, this dataset has a variety of data, including images of different shades, skin tone, hand size and background that distinguish it from the other dataset found online. It also has a training set of 900 examples and a testing set of 100 examples and they have a very high usability of 9.38 on Kaggle. Furthermore, we plan on merging it with this numeric American Sign Language dataset.

2. **Methodology:**

    a. Data Preprocessing:

Our selected datasets are highly feasible for our project. Their variety of diverse data with all the letters of the English alphabet aligns well with our objective of translating letter-by-letter in American Sign Language (ASL). However, one consideration is the absence of the "space" label. So, we plan to preprocess the dataset by converting the blank label into a space label, thus ensuring that our ASL translation model can be spelled out into a sentence.

These datasets are already cleaned and mostly preprocessed. All images within the datasets are sized at 512 x 512 pixels and positioning all hands at the center of each image.

    b. Machine learning model:

Our initial idea for this model was to use linear regression in order to classify images by interpreting what the sign displayed is, translating that sign from ASL to English, and displaying that translation. However, our TPM explained to us that since the data is not easily linearly separable (images tend to be relatively similar to each other), it is better to use neural networks to determine the most reliable interpretation of an image. For now, we expect the main hurdle to be training the neural network with a feasible amount of appropriate data. Training may also be computationally intensive depending on the size of the data.

c. Evaluation metric: Confusion matrix

   i. Using a normalized confusion matrix will allow us to see the actual sign and the predicted sign, which will show us overall accuracy (weighted F1-score) and also any trends in misclassification. The matrix will be rather large (28x28 for alphabet datasets + delete and space), so we could also use one-vs-all matrices to check performance for specific signs.
   ii. Baseline results to beat: we should aim to beat the ZeroR classifier, which defaults to predicting whatever class has the largest prevalence in the dataset, or a random guessing classifier. Essentially, we want to beat a "dummy classifier" using any strategy that isn't really learning from the data and improving over time.
   iii. For usability and communicating about the project to users, we can aim for a specific percentage (e.g. 80%), since just because it beats a dummy classifier does not guarantee that it would actually be helpful to users.

3. Application:

Since our model is an image classification and processing model meant to recognize human movements, we believe the best mode of input would be either a stream of images or a webcam. In the latter case, the model would process each frame coming from the video stream. The output that will be displayed should be text corresponding to the translation of the sign from ASL to English, as either a stream of continuous text or as a simple caption to the video or image.