Project Report

Our project is to convert sign language images of letters to text.

Problem statement:

What is the problem?

The problem is not everyone knows sign language and would not be able to understand someone speaking in sign language.

How can this problem be solved?

By creating a code that can read and give out what sign language is to text.

How does AI make this possible?

AI makes this possible because through the use of image classification and a neural network, we can create a code that can convert sign language to text. This uses deep learning processes in order to train the machine learning model to identify each gesture in sign language and translate it to text so that a person who is unable to understand sign language can understand the conversation by text.

How is this helpful?

This would be helpful to anyone who uses sign language as a way of communication and they want to talk to anyone who doesn't know ASL.

Methodology:

So far we have collected the relevant data needed to train and test the model, so now we need to create the model.

To build a sign language to text model, we first need an algorithm that will act as the foundation from which the program can be coded.

There are many algorithms that we can choose from:

1. Neural Networks

2. Convolutional Neural Network (CNN)

3. Multi Layer Perceptron (MLP)

Neural Network:

A neural network is a code that is supposed to replicate the human brain's thought process, inspired by the neurons of the brain, hence the name. It is a set of algorithms to recognize and classify data given from information on a given database. This is a very common way of classifying images or text and has many different uses in many different industries. The way it works is a code has an input layer, where the data is given. Then there are hidden layers. There could be just one or there could be many. After going through these layers, an output is produced. This is structured to replicate the neurons in a brain.

CNN:

Neural networks are a part of machine learning and an especially important part of deep learning algorithms. Convolutional neural networks, or CNNs in particular are superior to regular neural networks in terms of processing performance of analysing any image, speech or audio inputs.

The CNN is also made up of three separate layers: a convolutional layer, a pooling layer and a fully connected (FC) layer.

This allows it to have three dimensions, which are height, width and depth of the input. With each layer in the CNN,

the complexity increases which allows it to identify larger portions of an image until the image is recognised.

MLP:

A multilayer perceptron is also an algorithm to help classification.

It is also a neural network and follows the same structure as one.

1. An input layer

2. A hidden layer

3. An output layer

In an MLP, there is only one hidden layer but in Neural Networks or CNN's, there can be many.

Choice of algorithm:

After reviewing each algorithm, we have decided to use the CNN. This is because of its superior prcessing power and multiple layers that allow for more complex processing. This would be helpful

realistically, as the hand signals made by the users may not always be perfectly clear. We believe that the CNN would be able to distinguish different hand signals more effectively.

Method:

1. First, we need to gather the data. This can be from the web or we can create it. We are choosing it from the web as it is a greater range of data.

2. Now we have to prepare the data so it can be read by our model. This is done by changing it and reordering in a proper matrix for it to read.

3. Choosing a model - We have chosen CNN but we will create 3 different codes for all three of them and using the accuracy and precision scores, decide which is best.

4. Now we have to train the model from the dataset.

5. After training, we can evaluate it using the precision and accuracy scores and the confusion matrix.

6. Then, if required, hyperparameter tuning is necessary. This would help give optimal parameters for the model to perform its best.

7. Now the model is ready to test and if it meets the required scores for accuracy and precision, it is ready.

Data:

In our dataset, our different columns with different values allow us to distinguish between the different pictures and hence different letters.

Our dataset comes from an image showing the signs for different letters and hence allows us to classify new images when scanning from the existing database.

| Sign language alphabet image dataset reference - <https://www.researchgate.net/figure/ASL-hand-gesture-dataset_fig1_335503031>  Other references include our previous codes worked on in class |
| --- |