**Problem Statement:**

* The project involves processing videos of 30 frames each for doing gesture recognition for smart TV
* Images are from two sources where the dimentsions are (160,120,3) and (360,360,3)
* Each video is tagged with a gesture name amongst the five gesture names
  + Thumbs up – Increase the volume
  + Thumbs down – decrease the volume
  + Left swipe – Jump 10 secs backward
  + Right swipe – Jump 10 secs forward
  + Stop – Pause the movie

The data can be downloaded from <https://drive.google.com/uc?id=1ehyrYBQ5rbQQe6yL4XbLWe3FMvuVUGiL>

**Understanding the Dataset:**

The training data consists of a few hundred videos categorized into one of the five classes. Each video is divided into a sequence of 30 frames(images).

**Objective:**

To train different models on the train folder to predict the action performed in each sequence or video and which performs well on the validation folder as well.

Two types of architectures are suggested for analysing videos using deep learning.

* 3D convolutional Neural Networks
* CNN + RNN architecture