**Data Preprocessing**

Data preprocessing is a step of cleaning, transforming and aggregating data before it can work with algorithms. In this project we need to preprocess live video frames before they can be fed to a model using OpenCV we can detect and extract face(s) from frames. OpenCV provides three different algorithms for detecting faces in a stream of images viz., Cascade Classifier and MTCNN (MultiTask Convolution Neural Network) using Haar Basis functions with AdaBoost as its core component first created by Viola-Jones and HOG (Histogram of Oriented Gradients) descriptor and object detector by Navneet Dalal and Bill Triggs. Cascade Classifier can process 25 images/second with precision of 95.24% and recall of 82.60%, MTCNN can process 3 images/second with precision of 98.02% and recall of 89.85% this reading is estimated using CPU it can be increased using better CPU or GPU. MTCNN can deal with scale and orientation of the face where Cascade Classifier cannot. For our purpose we will be using MTCNN for training face recognition algorithms and Cascade Classifier for real time face detection (MTCNN can also be used for real time face detection if GPU is in use).