

1. Frame Level Classification:

Model	Accuracy
CNN w/ pooling	55.35 / 91.3*
CNN w/o pooling	57.7 / 87.97*
Deep CNN	53.8 / 92.8*

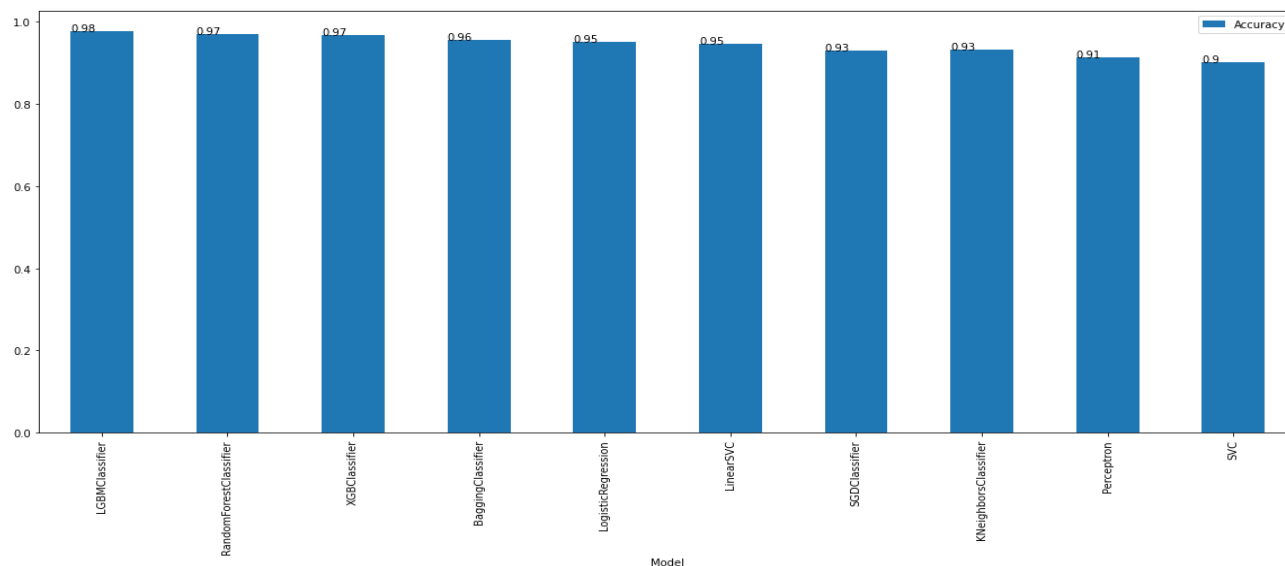
* When Model is train on frame level but classification is done on batch level

2. Sequence Model

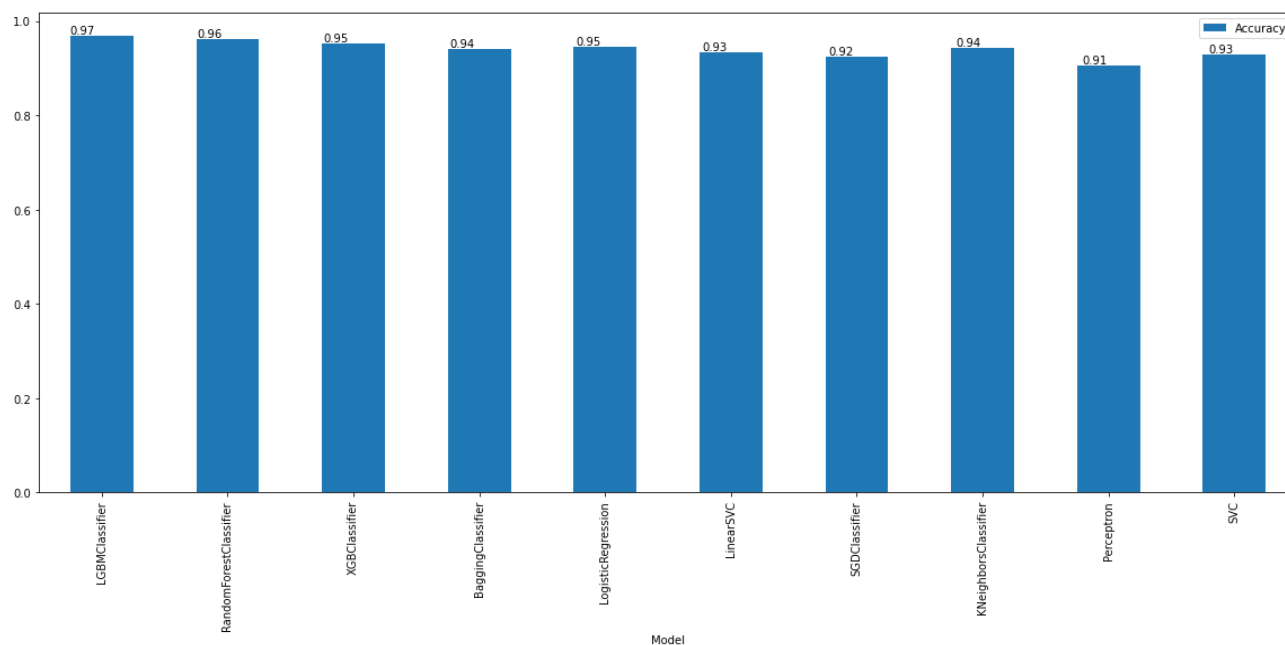
Model	Accuracy
RNN shallow	98.81
CNN+LSTM (end to end)	99.02
3D CNN	99.06

3. stacked Images

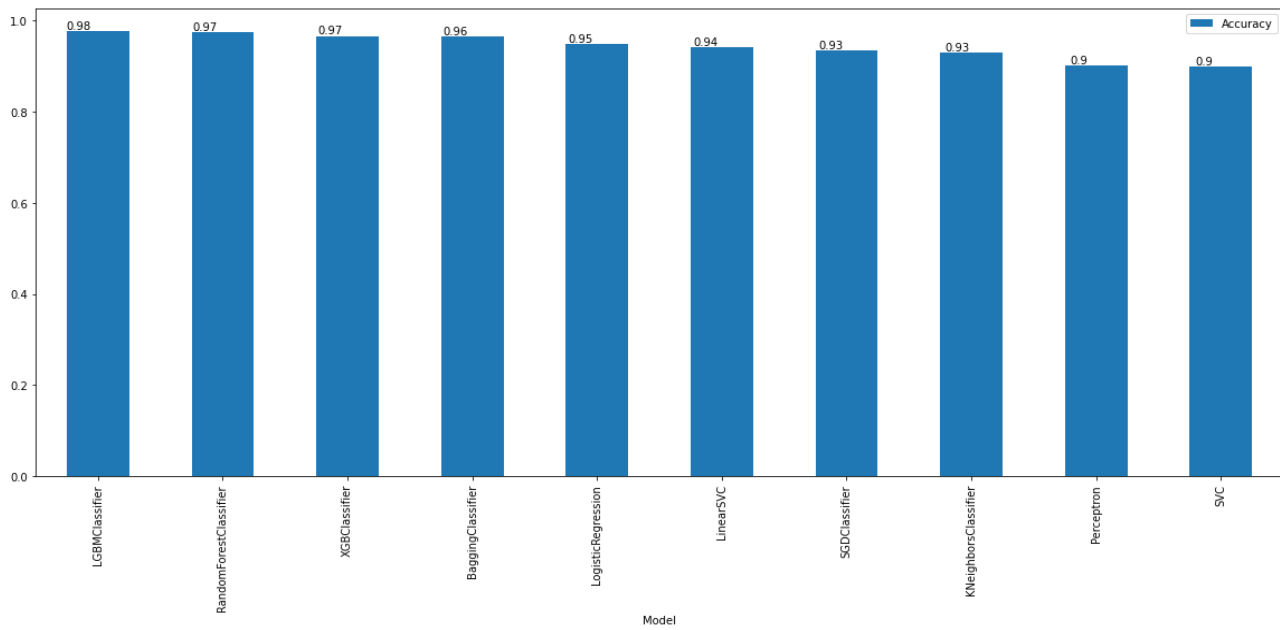
a. SUM IMAGES



b. MAX IMAGES

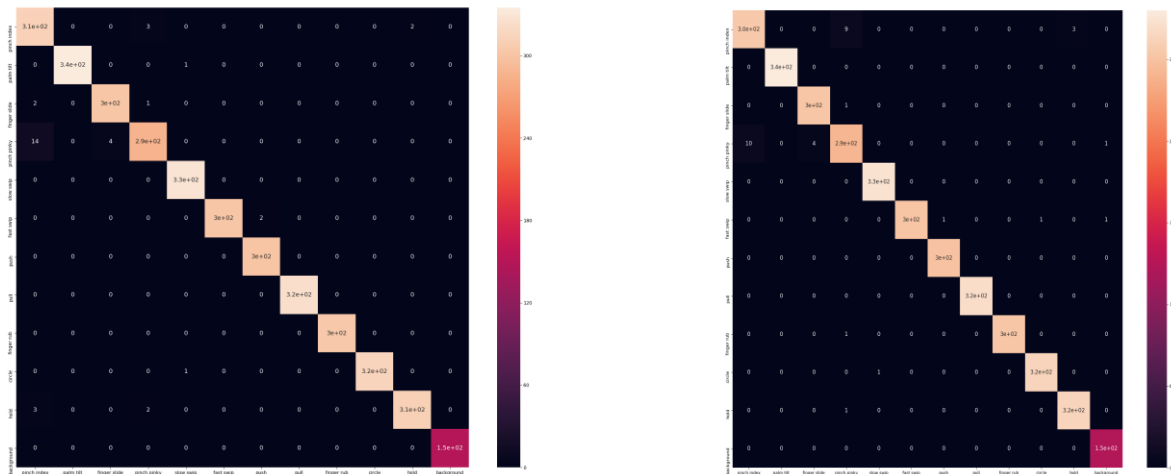


c. AVG IMAGES



MODEL : CNN	Accuracy(best across diff arch. of cnn)
Sum Images	98.3
Max images	97.2
Average images	97.7

Confusion Matrix (CNN LSTM and 3D CNN)



Remark: all models performs quite well in inter separable gesture (like hold, circle, pull, push, fast swipe, slow swipe) whereas the gestures like pich index(pinching thumb and index finger) and pinchpinky(pinching thumb and little finger) are generally get confused. finger rub and finger slide being micro gesture performs well.

[Github](#)