**Gesture Recognition**

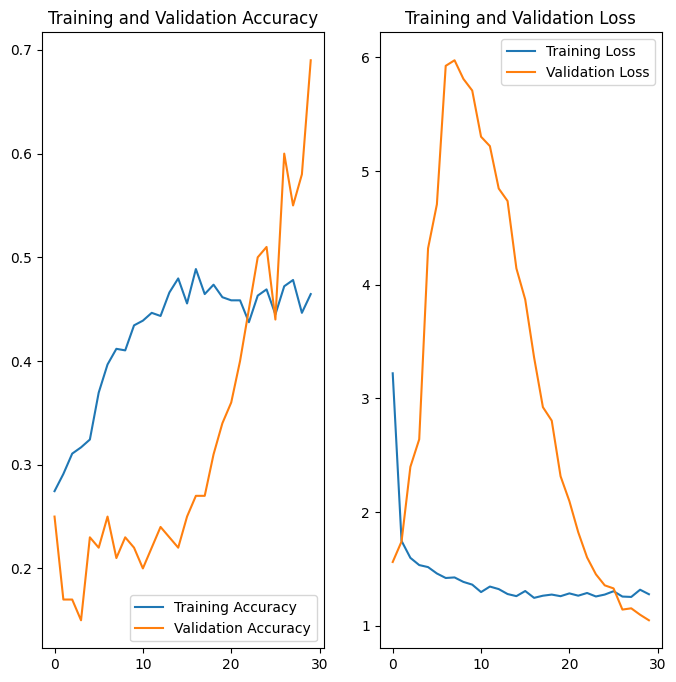
This is the write-up required for evaluation on this RNN assignment on Gesture Recognition.

**Common Key Points:**

* We’re taking only a subset of 15 images from the 30 images available from the video sample, taken after considering computational restrictions and without losing information from it.
* Data available has images in two different dimensions (360x360, 120x160). And we need to feed in same size images for training. So, all images have been resized to 120x120 (sufficient detail possible to capture from both)
* Batch size taking value of 32 (A batch size of 32 means that 32 samples from the training dataset will be used to estimate the error gradient before the model weights are updated)
* All models trained for 30 epochs
* Optimizer Adam for all models

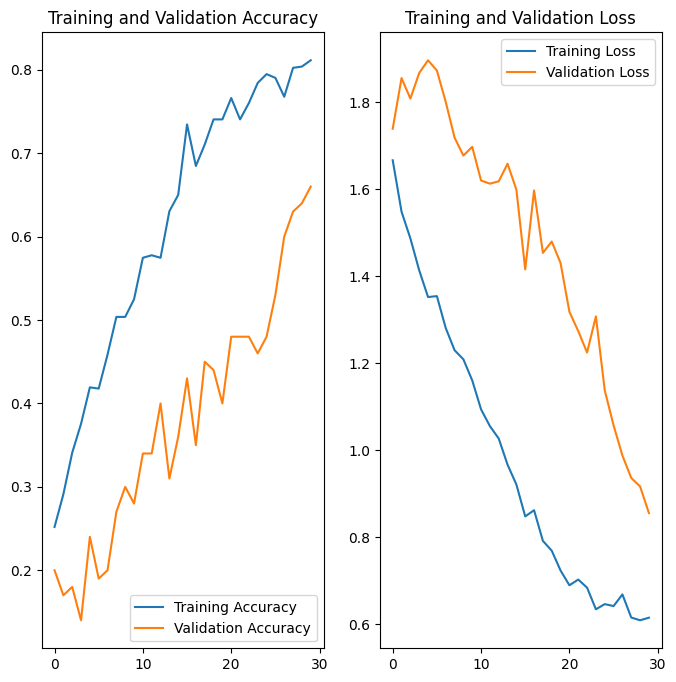
**Model 1 : Conv3D**

* No overfitting noticed but accuracy very low
* Training accuracy not even exceeding 50% for any of the epoch.
* Taking one optimum model value: (epoch 25)
  + Categorical accuracy: 46.91%
  + Validation Categorical accuracy: 51%



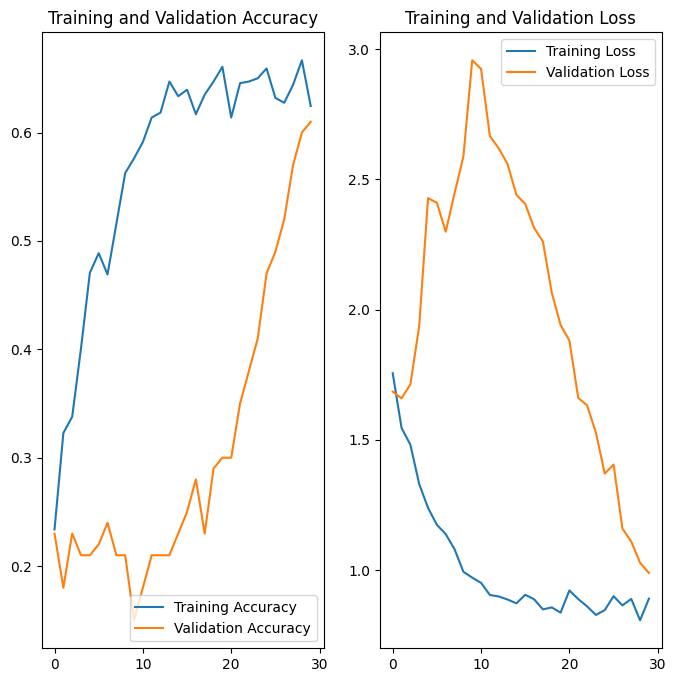
**Model 2 : Conv2D + LSTM**

* Better performance when compared against vanilla Conv3D model
* Not overfitting but validation accuracy slightly lesser compared against training
* Choosing one optimum model values for this experiment:
  + Categorical accuracy: 0.8115
  + Validation Categorical accuracy: 0.6600



**Model 3 : Conv2D + GRU**

* Less overall accuracy when compared against LSTM counterpart but gap between training and validation accuracy reduced drastically
* Choosing one optimum model values from this experiment:
  + Categorical accuracy: 0.6667
  + Validation categorical accuracy: 0.6000



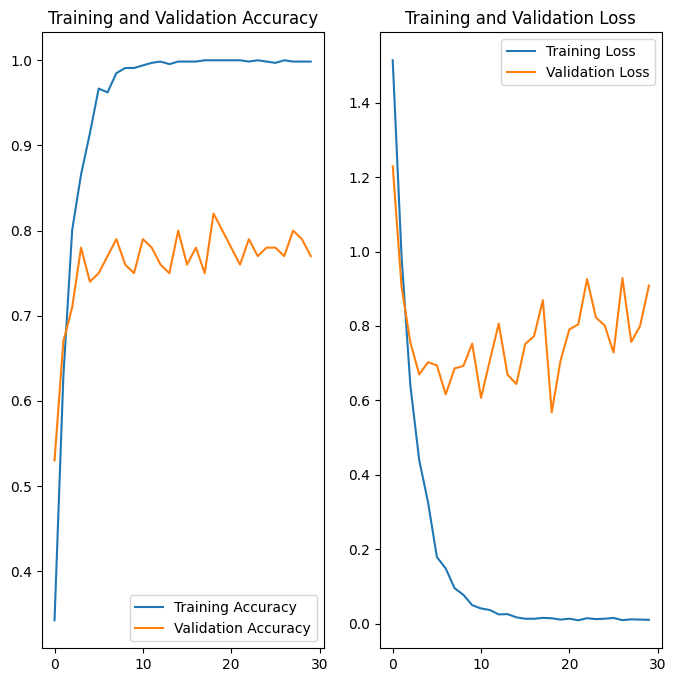
**Model 4: MobileNet Based LSTM**

* Using MobileNet architecture to implement transfer learning in conjunction with LSTM
* Choosing one optimum model values from this experiment:
  + Categorical accuracy: 0.9970
  + Validation Categorical accuracy: 0.8300



**Model 5 : MobileNet Based GRU**

* Using MobileNet architecture to implement transfer learning in conjunction with LSTM
* Choosing one optimum model values from this experiment:
  + Categorical accuracy: 1.0000
  + Validation Categorical accuracy: 0.8000

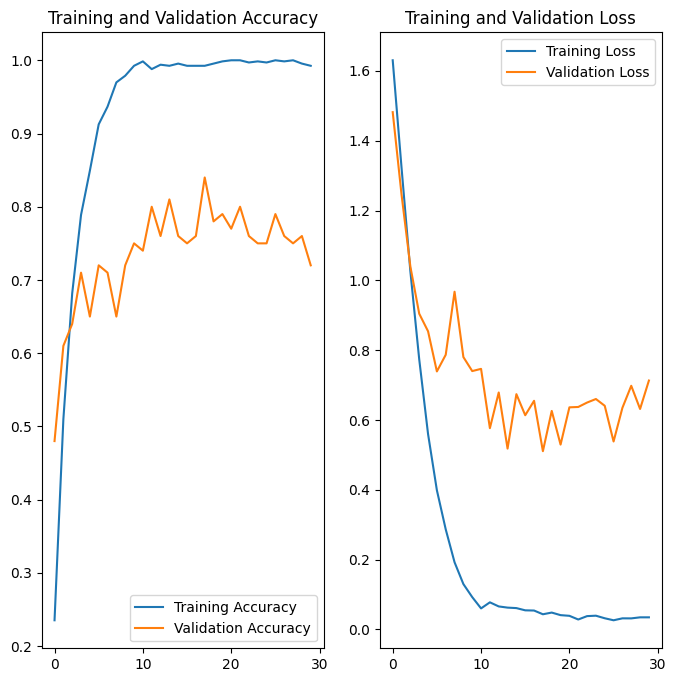


**Choosing the best model:**

* Both the LSTM and GRU models developed using pre-trained MobileNet architecture outperform the other models.
* LSTM transfer learning model performs better than the GRU one since the validation accuracy for it is higher.

**Model 6 : MobileNet Based LSTM (Experiment II)**

* So far, we’ve concluded the MobileNet LSTM model is the best to choose from all above. The results were produced keeping the batch size of 32.
* This time, we’re training the same LSTM model with the batch size of 64 to compare performance against its earlier iteration.
* The values for main performance indicators for this version are:
  + Categorical accuracy: 99.25%
  + Validation Categorical accuracy: 84%



**Conclusion:**

MobileNet Based LSTM (with batch size of 64) performs better than its counterpart. And thus, choosing this as the final model for the gesture recognition problem.