**Method**

To analyze the gestures using computer vision, we tried to implement a classification model that a video to

**VGG:**

* No added dense layer: 70 minutes (1fps, 60% on biased test set)
* 2 labels
  + 2 min for processing data
  + 50 seconds per epoch (batch size of 50)
  + Time (min) 23.398771381378175249

Best Epoch:

Epoch 26/30

28/28 [==============================] - 39s 1s/step - loss: 6.6008e-05 - accuracy: 1.0000 - val\_loss: 1.9669 - val\_accuracy: 0.6920

Chart

Description automatically generated

Chart, line chart

Description automatically generated

* 4 labels
  + 4 min for processing data
  + Time: 33.6666 min

Best Epoch

Epoch 20/30

50/50 [==============================] - 67s 1s/step - loss: 4.6015e-05 - accuracy: 1.0000 - val\_loss: 2.8212 - val\_accuracy: 0.5853

Chart

Description automatically generatedChart, histogram

Description automatically generated

**VGG Model Architecture:**

Table

Description automatically generated

**Resnet:**

* 2 labels
  + 2 min for processing data
  + 40 seconds per epoch (batch size of 50)
  + Time: 29.152676479021707223 min

Best Epoch:

Epoch 27/30

28/28 [==============================] - 34s 1s/step - loss: 6.4692e-04 - accuracy: 1.0000 - val\_loss: 0.8601 - val\_accuracy: 0.8616

Chart, line chart

Description automatically generatedChart, line chart, histogram

Description automatically generated

* 4 labels
  + 4 min for processing data
  + Time: 29.152676479021707223 min

Best Epoch:

Epoch 25/30

50/50 [==============================] - 82s 2s/step - loss: 0.0013 - accuracy: 1.0000 - val\_loss: 2.4657 - val\_accuracy: 0.6221

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

**ResNet Model Architecture:**

Table

Description automatically generated

**EfficientNet (from b0 to 7):**

* 2 labels
  + 2 min for processing data
  + 50 seconds per epoch (batch size of 50)
* 4 labels

**MediaPipe:**