**Model Details :**

**Data Set**: Custom data. 4 Person, 2 male 2 female. 750 samples per person, each 4 sec long. All in Bengali language.   
  
**Feature Extraction Methods:**

MFCC: Number of MFCCs extracted from each clip 20

LFCC: Number of LFCCs extracted from each clip 13  
LPC: Number of LPCs extracted from each clip 13

RPLP: Number of RPLPs extracted from each clip 13

**Model 1:**

Input = MFCC

**Model Summary:**

Model: "sequential"

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

conv2d (Conv2D) (None, 248, 248, 128) 3584

max\_pooling2d (MaxPooling2D) (None, 124, 124, 128) 0

conv2d\_1 (Conv2D) (None, 122, 122, 128) 147584

max\_pooling2d\_1 (MaxPooling 2D) (None, 61, 61, 128) 0

conv2d\_2 (Conv2D) (None, 59, 59, 64) 73792

max\_pooling2d\_2 (MaxPooling 2D) (None, 29, 29, 64) 0

conv2d\_3 (Conv2D) (None, 27, 27, 64) 36928

max\_pooling2d\_3 (MaxPooling 2D) (None, 13, 13, 64) 0

flatten (Flatten) (None, 10816) 0

dropout (Dropout) (None, 10816) 0

dense (Dense) (None, 32) 346144

dense\_1 (Dense) (None, 5) 165

=================================================================

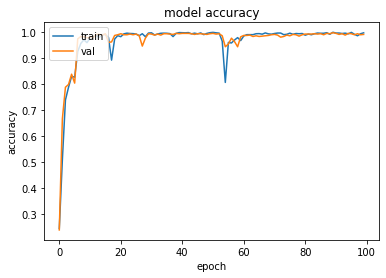
Total params: 608,197

Trainable params: 608,197

Non-trainable params: 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Output:**

Training Accuracy- 99.69%  
Validation Accuracy-99.08%  
  


**Model 2:**

Input = LFCC

Model Summary:

Model: "sequential\_2"

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

conv2d\_6 (Conv2D) (None, 248, 248, 64) 1792

max\_pooling2d\_6 (MaxPooling 2D) (None, 124, 124, 64) 0

conv2d\_7 (Conv2D) (None, 122, 122, 64) 36928

max\_pooling2d\_7 (MaxPooling 2D) (None, 61, 61, 64) 0

conv2d\_8 (Conv2D) (None, 59, 59, 32) 18464

max\_pooling2d\_8 (MaxPooling 2D) (None, 29, 29, 32) 0

flatten\_2 (Flatten) (None, 26912) 0

dropout\_2 (Dropout) (None, 26912) 0

dense\_4 (Dense) (None, 32) 861216

dense\_5 (Dense) (None, 5) 165

=================================================================

Total params: 918,565

Trainable params: 918,565

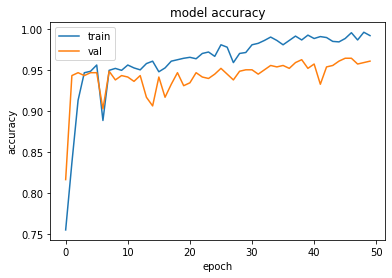
Non-trainable params: 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Output:**

Training Accuracy :99.24%

Validation Accuracy:96.13%



**Model 3:**

Input = LPC

**Model Summary:**

Model: "sequential"

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

conv2d (Conv2D) (None, 248, 248, 64) 1792

max\_pooling2d (MaxPooling2D) (None, 124, 124, 64) 0

conv2d\_1 (Conv2D) (None, 122, 122, 64) 36928

max\_pooling2d\_1 (MaxPooling2D) (None, 61, 61, 64) 0

conv2d\_2 (Conv2D) (None, 59, 59, 32) 18464

max\_pooling2d\_2 (MaxPooling2D) (None, 29, 29, 32) 0

flatten (Flatten) (None, 26912) 0

dropout (Dropout) (None, 26912) 0

dense (Dense) (None, 32) 861216

dense\_1 (Dense) (None, 5) 165

=================================================================

Total params: 918,565

Trainable params: 918,565

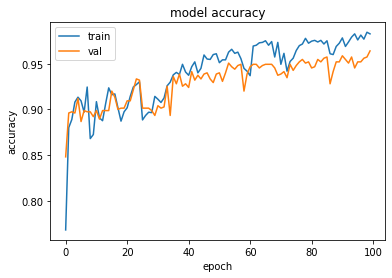
Non-trainable params: 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Output**:

Training Accuracy: 98.27%

Validation Accuracy: 96.40%



**Model 4:**

Input = RPLP

**Model Summary :**

Model: "sequential"

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

=================================================================

conv2d (Conv2D) (None, 248, 248, 64) 1792

max\_pooling2d (MaxPooling2D) (None, 124, 124, 64) 0

conv2d\_1 (Conv2D) (None, 122, 122, 64) 36928

max\_pooling2d\_1 (MaxPooling2D) (None, 61, 61, 64) 0

conv2d\_2 (Conv2D) (None, 59, 59, 32) 18464

max\_pooling2d\_2 (MaxPooling2D) (None, 29, 29, 32) 0

flatten (Flatten) (None, 26912) 0

dropout (Dropout) (None, 26912) 0

dense (Dense) (None, 32) 861216

dense\_1 (Dense) (None, 5) 165

=================================================================

Total params: 918,565

Trainable params: 918,565

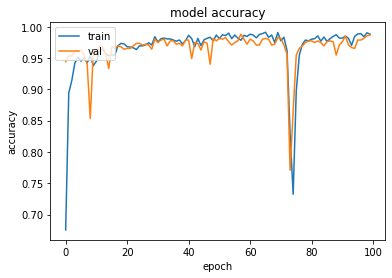
Non-trainable params: 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Output:**

Training Accuracy: 98.80%

Validation Accuracy: 98.67%



| **Feature** | **Feature Spectrogram** | **Epochs** | **Training Accuracy** | **Validation Accuracy** |
| --- | --- | --- | --- | --- |
| MFCC |  | 100 | 99.69% | 99.08% |
| LFCC |  | 50 | 99.24% | 96.13% |
| LPC |  | 100 | 98.27% | 96.40% |
| RPLP |  | 100 | 98.80% | 98.67% |