Investigation on Facial Emotion Recognition

Under the guidance of
Dr. Anshu S. Anand &
Dr. Ramesh Kumar Bhukya

Group Members:
Priyamvada Priyadarshini IIB2020037
Shalini Agrawal IIT2020236
Prakhar Chitravanshi IIT2020235
Anshuman Jain IIT2020239
Siddhant Agrawal IIT2020228

INTRODUCTION

Definition: Facial expression recognition is sentiment analysis method to detect faces, code facial expressions, and recognize emotional states.

Different Facial expressions: happiness, sadness, anger, surprise, fear, neutral and disgust.

Objective: To recognize the facial expressions based on the images to find out in which mood the person is currently?

Application: Useful for human-human as well as human-computer interactions

Datasets: FER2013, CK+48

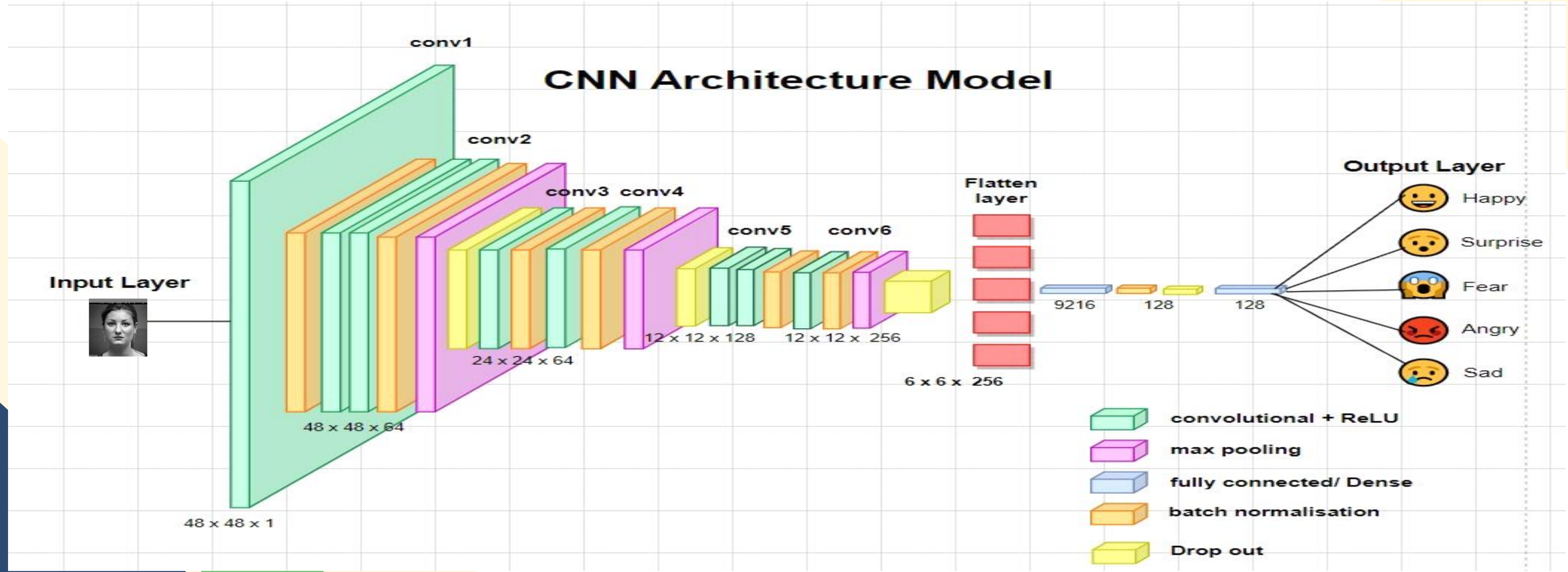
Machine learning methods: CNN, CNN-LSTM, CNN-BiLSTM

Transfer learning methods: VGG16, VGG19, and ResNet50

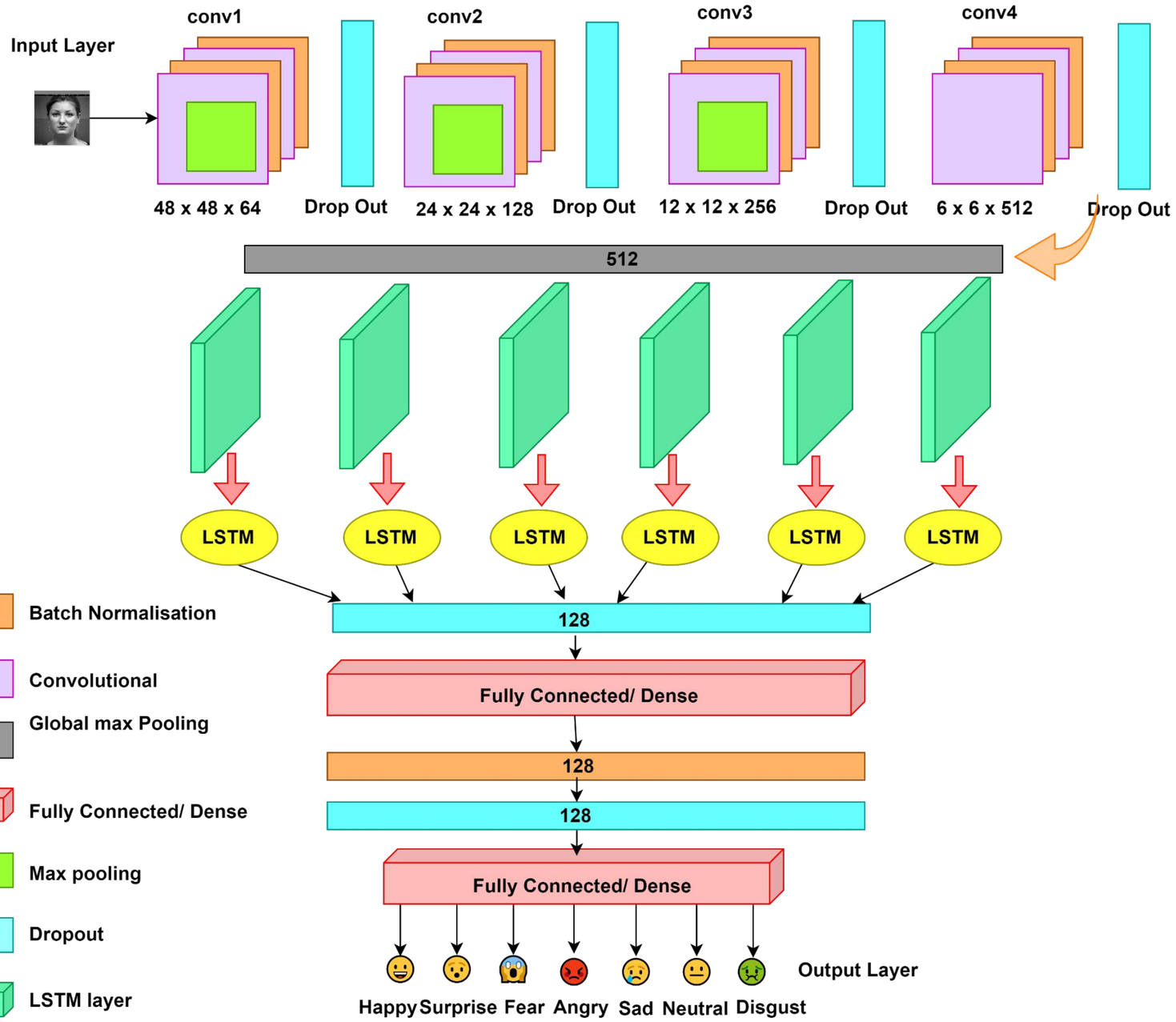
CNN:

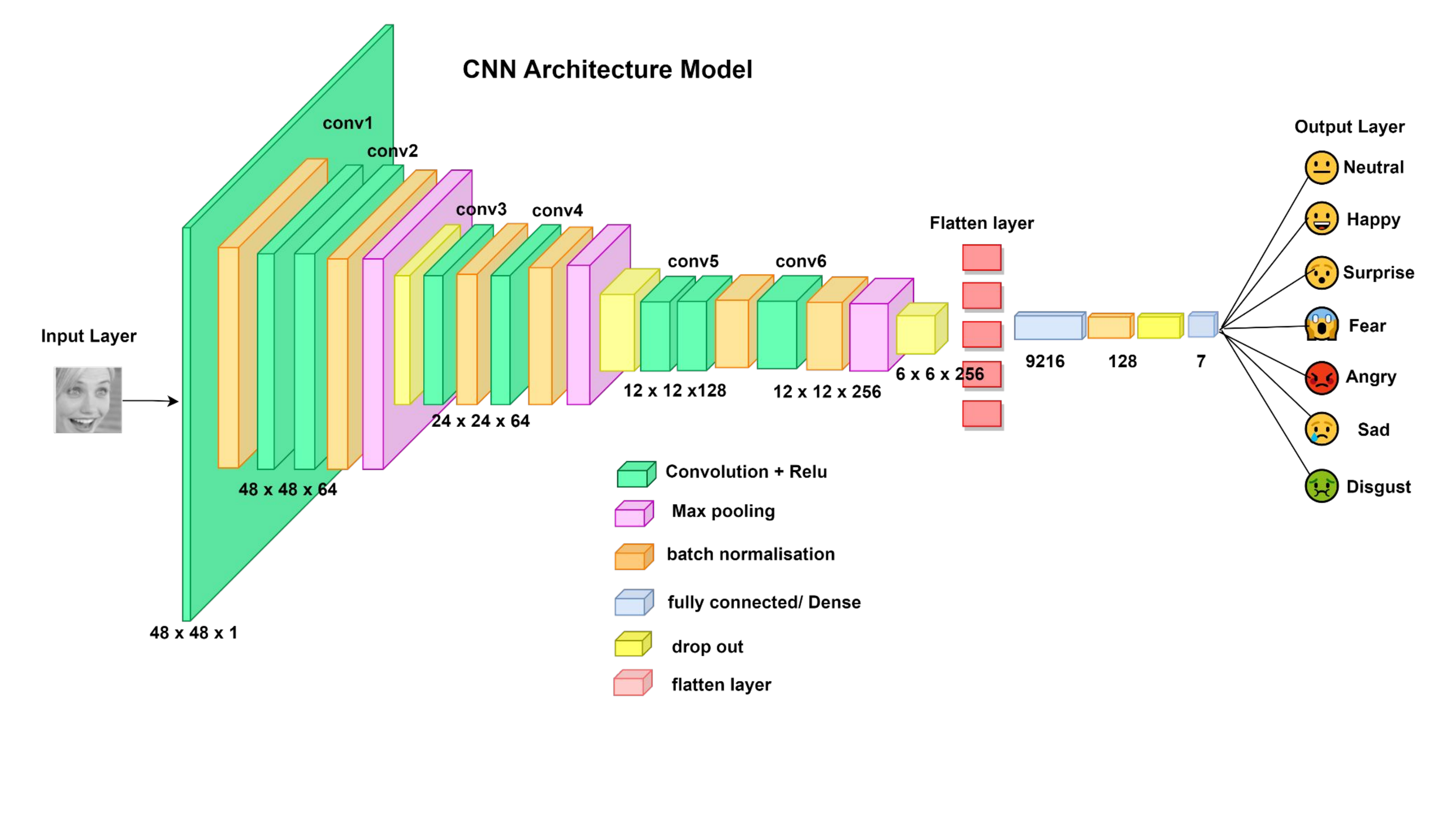
CNN Steps include:

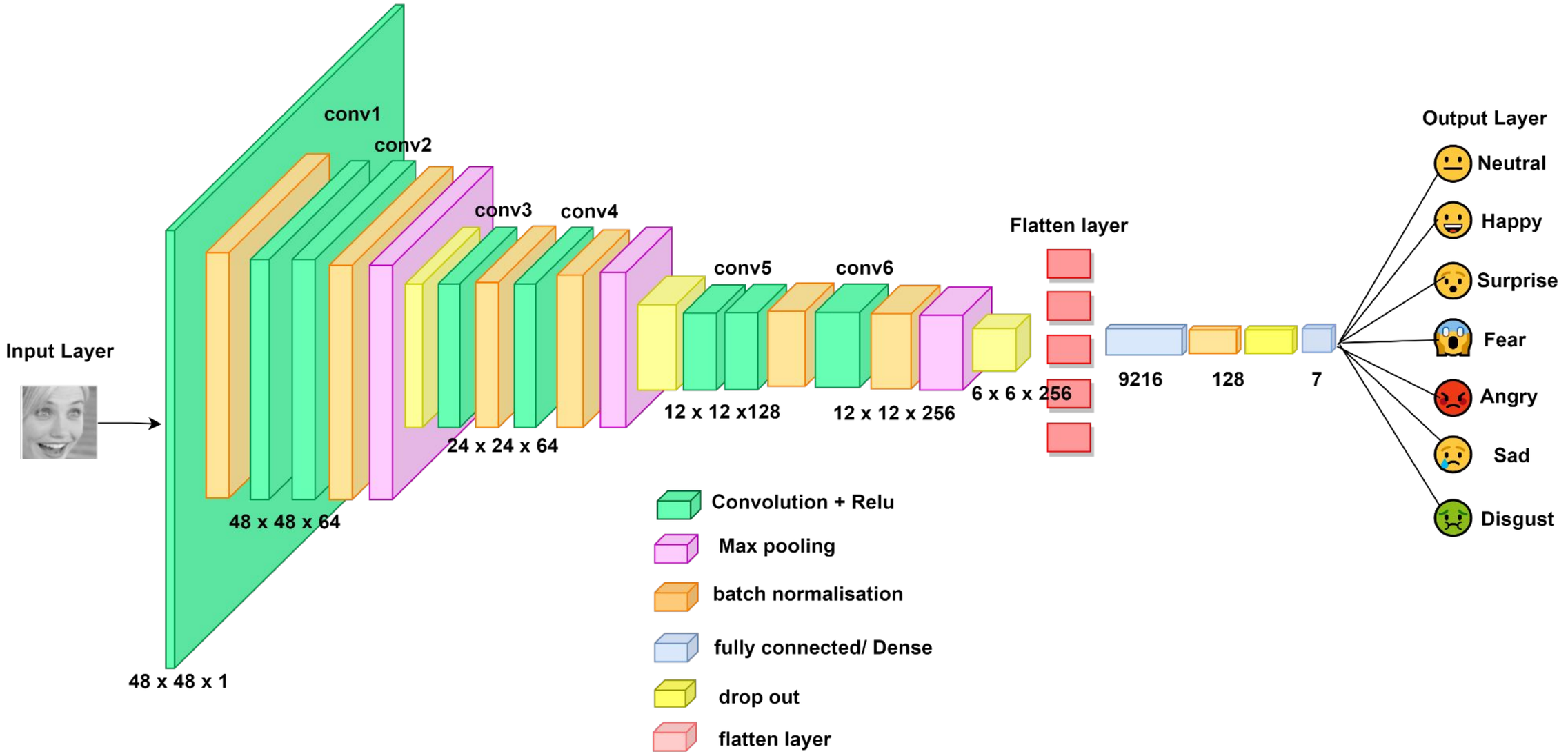
1. Convolution: Apply filters to generate feature maps
2. Non Linearity: ReLU
3. Pooling: Down sampling operation on each feature map. It is used for reducing dimensionality and making spatial invariance.
4. Fully Connected Layer: Classify the image based on the features extracted.



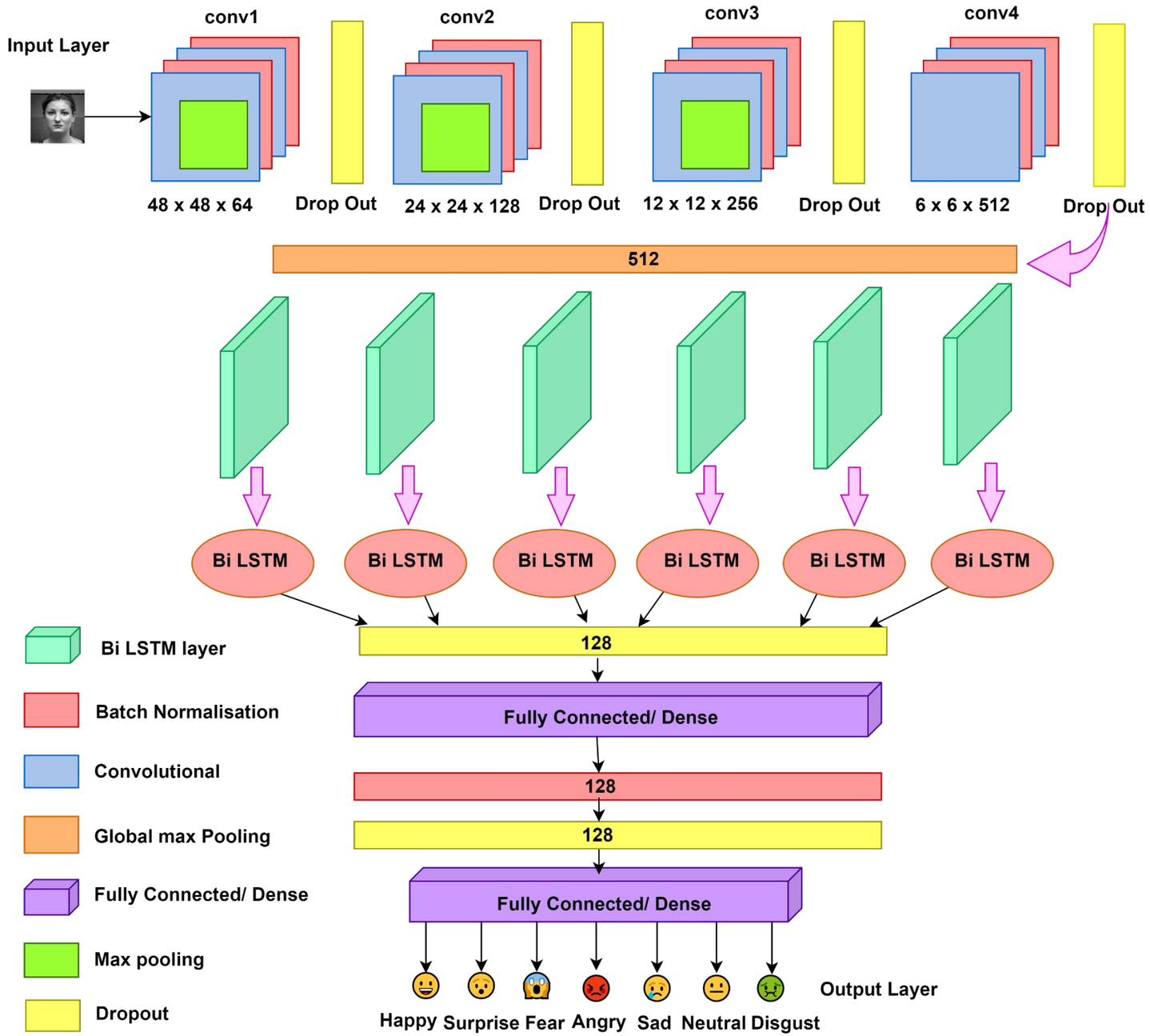
CNN+LSTM Architecture Model




CNN Architecture Model
Input Layer
48 x 48 x 1
conv1
conv2
48 x 48 x 64
conv3
conv4
24 x 24 x 64
conv5
12 x 12 x128
conv6
12 x 12 x 256
Flatten layer
6 x 6 x 256
9216
128
7
Output Layer
Neutral
Happy
Surprise
Fear
Angry
Sad
Disgust
Convolution + Relu
Max pooling
batch normalisation
fully connected/ Dense
drop out
flatten layer


Input Layer
48 x 48 x 1
conv1
conv2
48 x 48 x 64
conv3
conv4
24 x 24 x 64
conv5
12 x 12 x128
conv6
12 x 12 x 256
6 x 6 x 256
Flatten layer
9216
128
7
Output Layer
Neutral
Happy
Surprise
Fear
Angry
Sad
Disgust
Convolution + Relu
Max pooling
batch normalisation
fully connected/ Dense
drop out
flatten layer

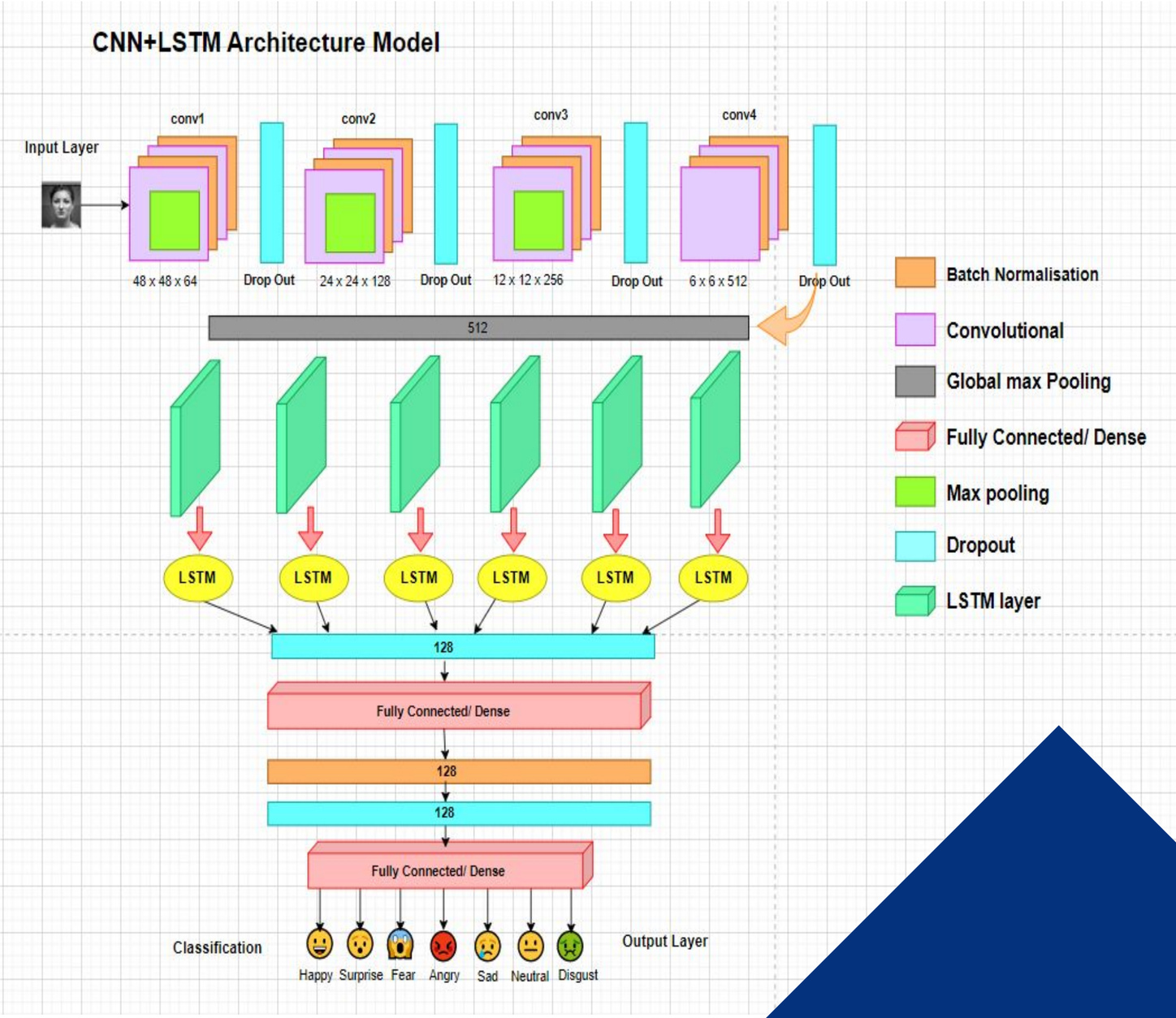
CNN+BiLSTM Architecture Model



CNN LSTM architecture :Hochreiter & Schmidhuber

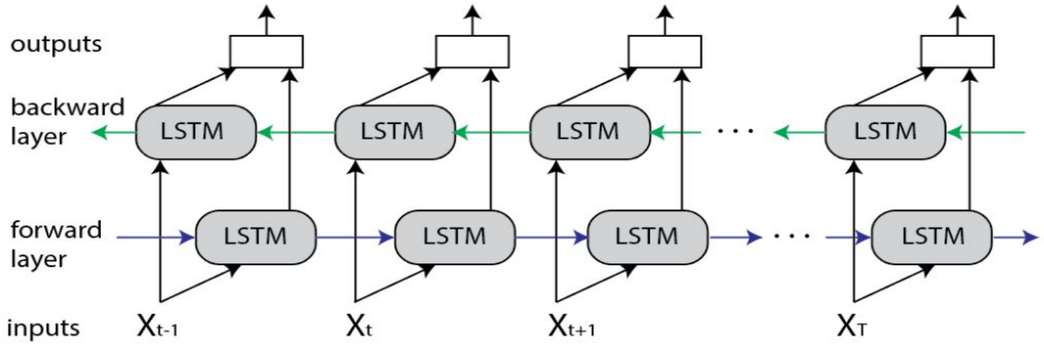
Recurrent Neural Network that can learn order dependence: **Long short Term Memory**

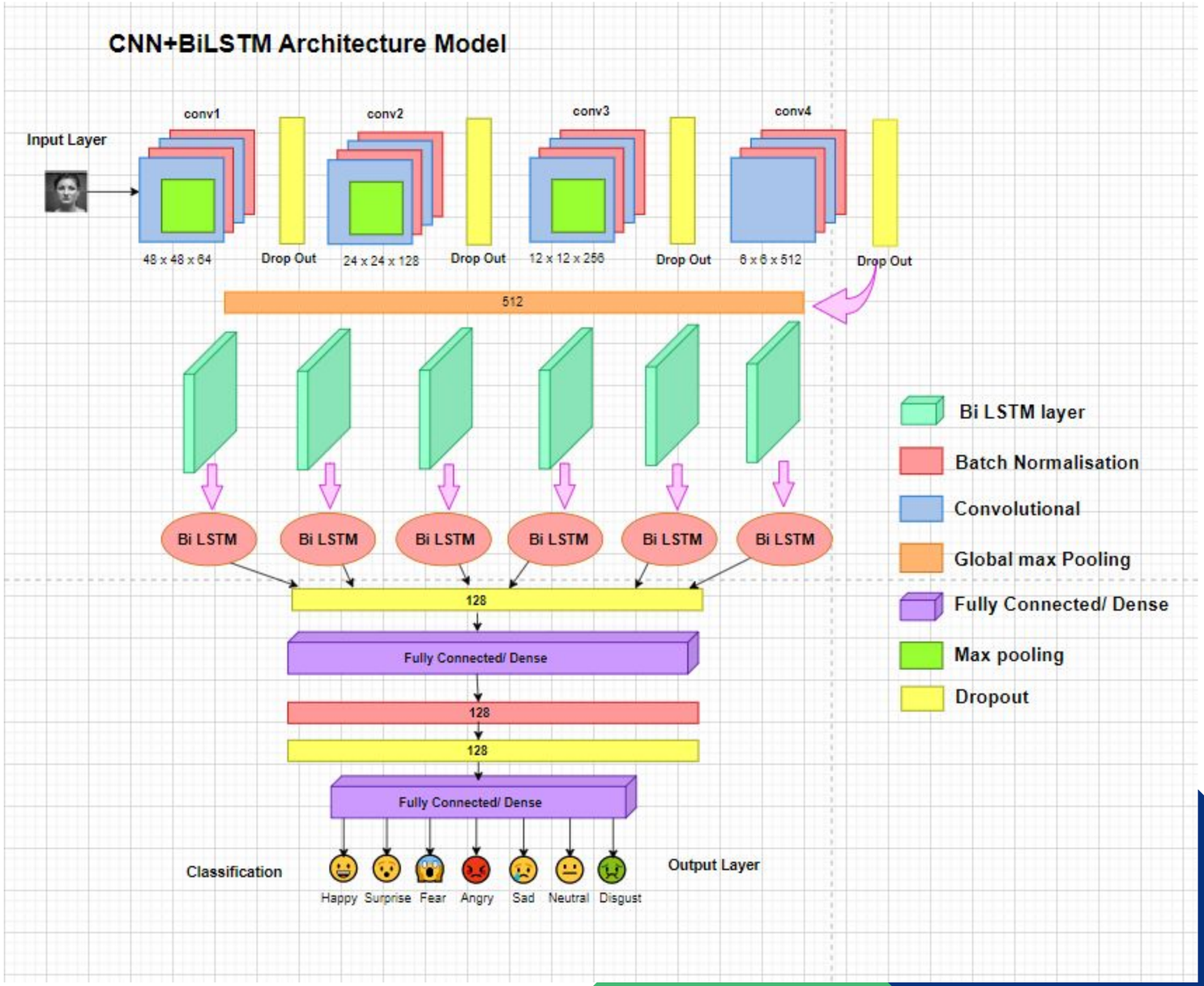
- Unfolded RNN
- Numerous memory blocks called cell form chain structure
- Learns from long-term dependencies.
- LSTM's ability to forget, remember and update the information pushes it one step ahead of RNNs.



CNN Bi-LSTM architecture

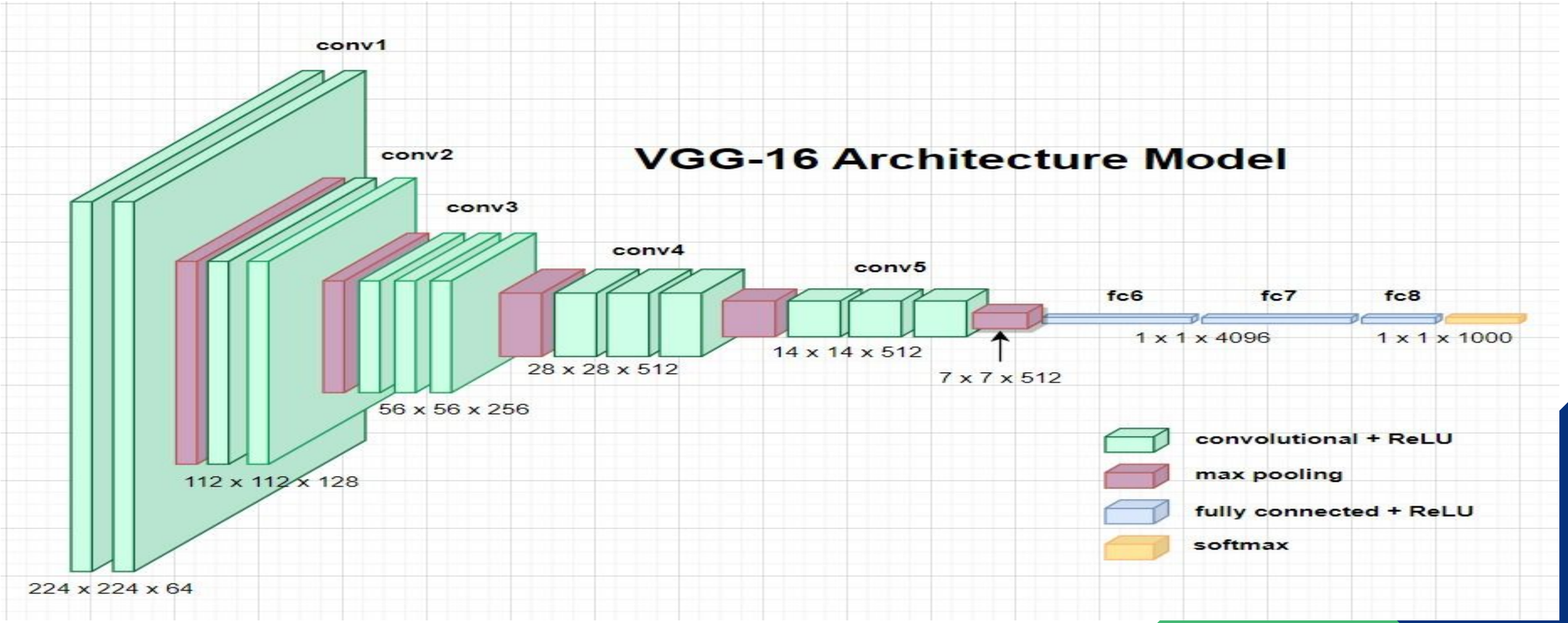
- Each training sequence is presented forwards and backwards to two independent recurrent nets, coupled to the same output layer in Bidirectional RNN (BRNN)
- It data in both ways with two hidden layers that feed-forward to the same output layer.
- BRNN + LSTM = LSTM that can access long-range context in both input directions.





VGG-16(Visual Geometry Group)

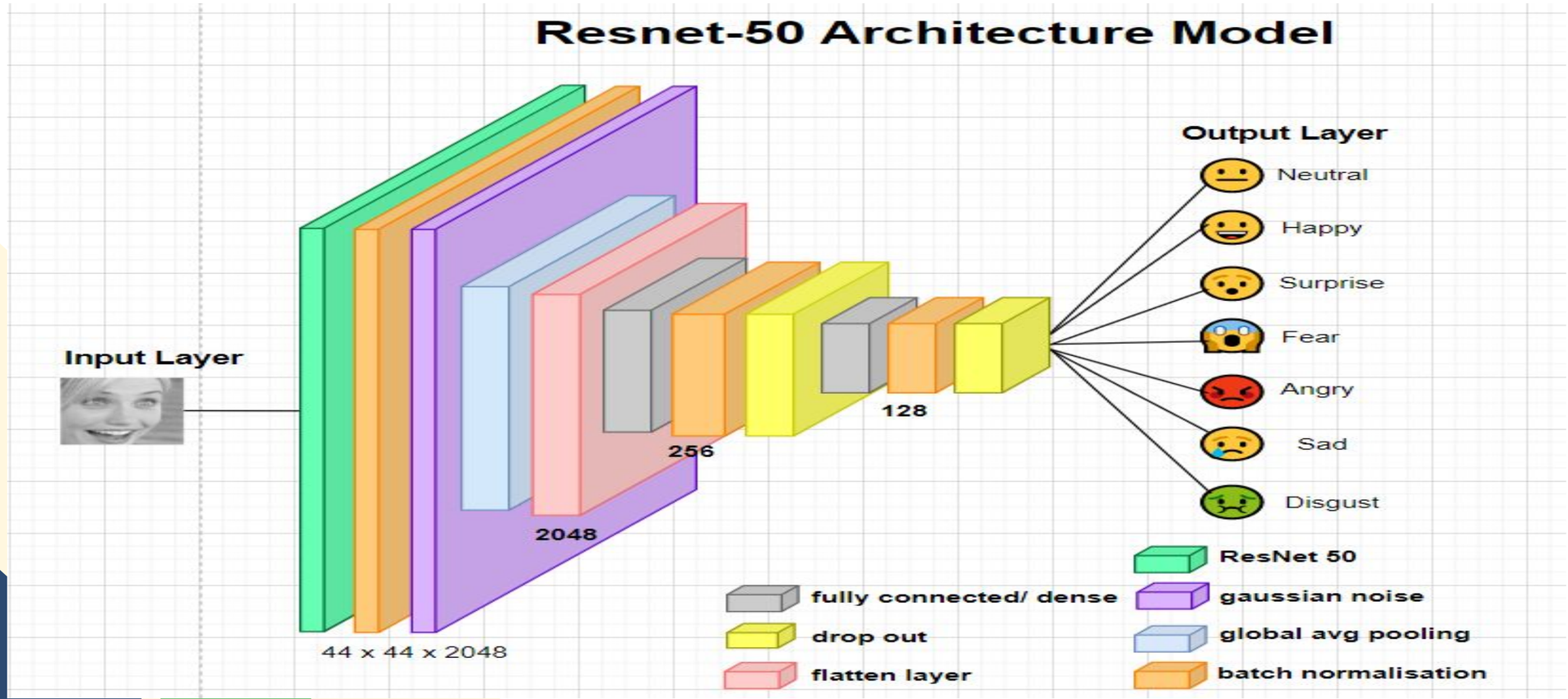
- VGG16 is object detection and classification algorithm.
- there are total 21 layers with 16 weight layers.
- Conv-1 Layer has 64 number of filters, Conv-2 has 128 filters, Conv-3 has 256 filters, Conv 4 and Conv 5 has 512 filters.
- VGG16 takes input tensor size as 224, 244 with 3 RGB channel



RESNET50

The 50-layer ResNet architecture:

- Residual Blocks. In this network, we use a technique called skip connections
- The skip connection connects activations of a layer to further layers by skipping some layers in between. This forms a residual block. Resnets are made by stacking these residual blocks together.



DATASETS

FER2013

35887 different images

7 different emotions:

Anger : 958 Sadness : 1247

Disgust : 111 Surprise : 831

Fear : 1024 Happy : 1774

Neutral : 1233

Training dataset: 28,709

Validation Dataset: 3,589

FER2018

35887 different images

7 different emotions:

Anger : 958 Sadness : 1247

Disgust : 111 Surprise : 831

Fear : 1024 Happy : 1774

Neutral : 1233

Training dataset: 28,709

Validation Dataset: 3,589

CK+48

750 different images.

5 different emotions:

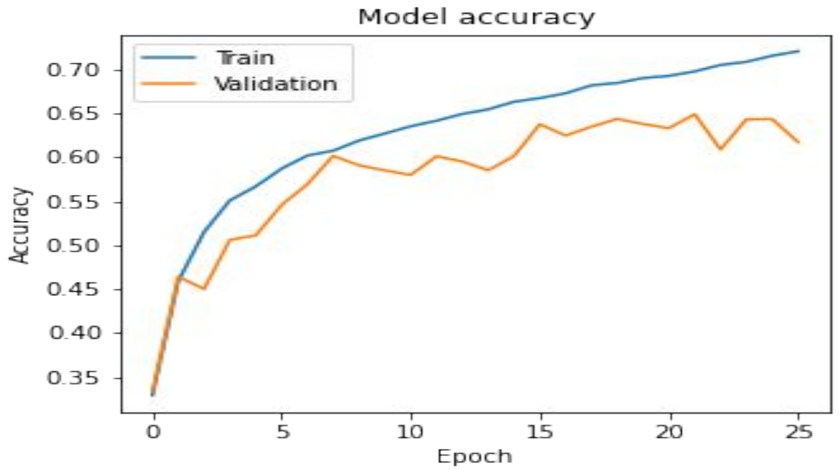
Anger : 135 Sadness : 84

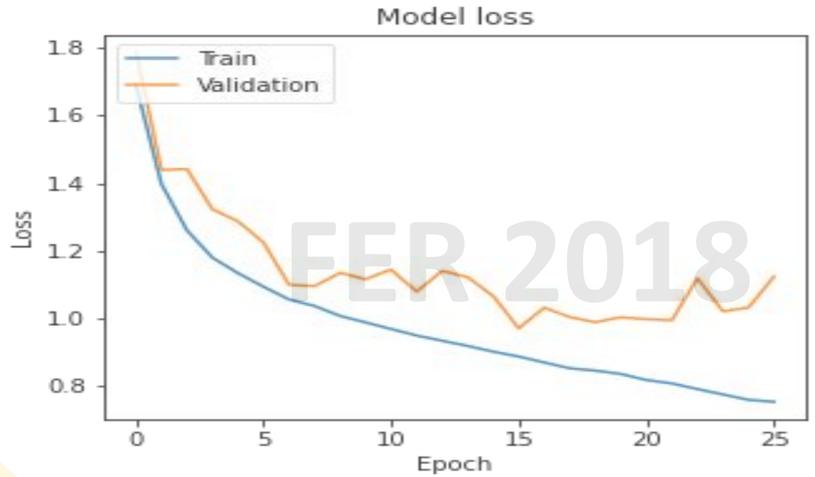
Happy : 207 Surprise : 249

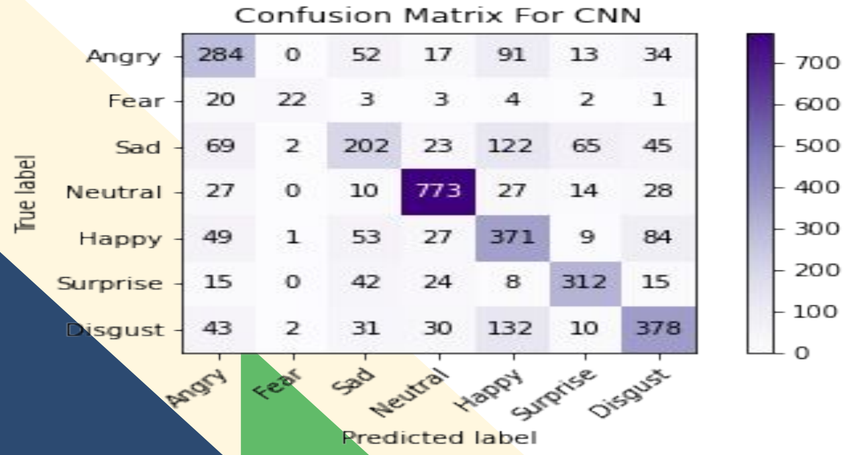
Fear : 75

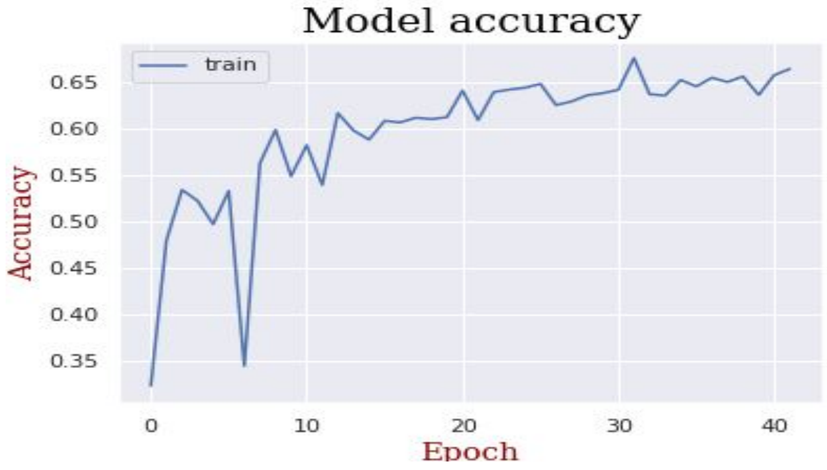
Training dataset:525

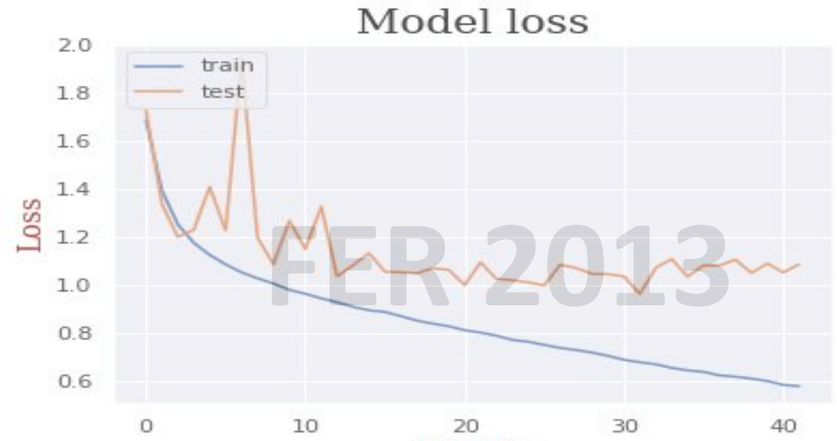
Validation Dataset:225







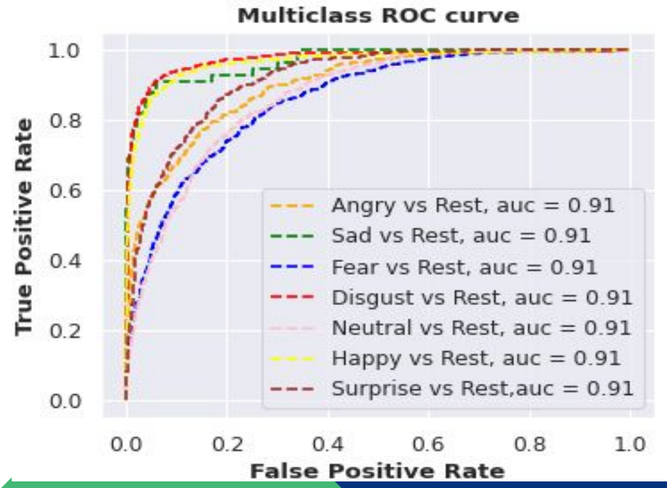


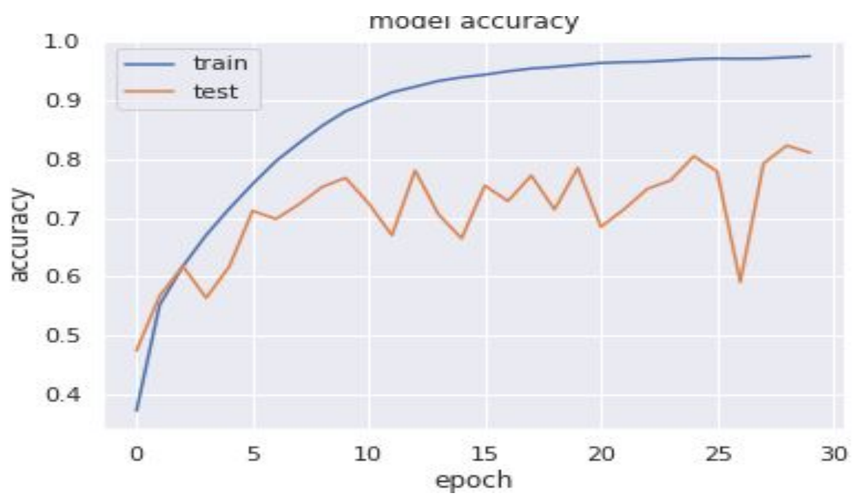


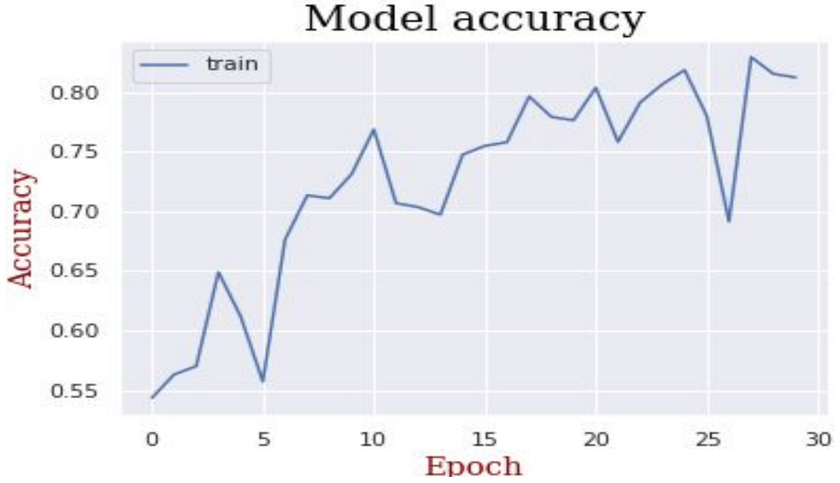
CNN:

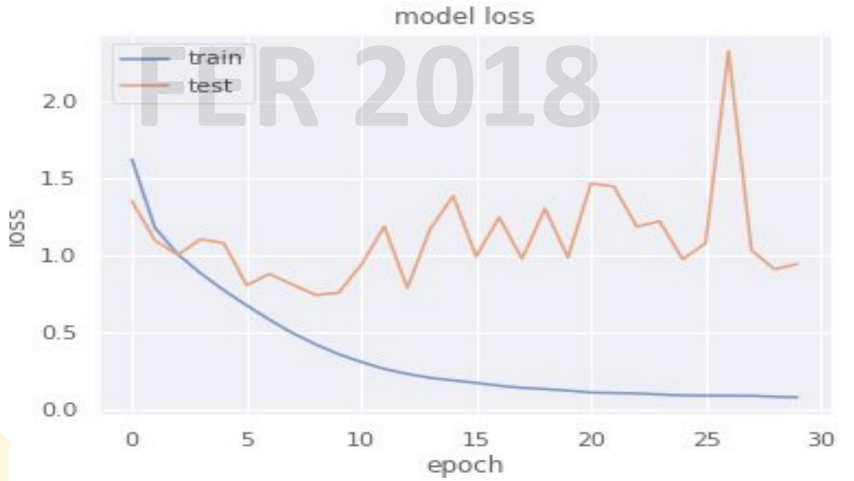
FER 2013-
Training Accuracy: 76.4%
Validation Accuracy: 66.9%
F1 Score: 0.69

FER 2018-
Training Accuracy: 70%
Validation Accuracy: 64%
F1 Score: 0.68



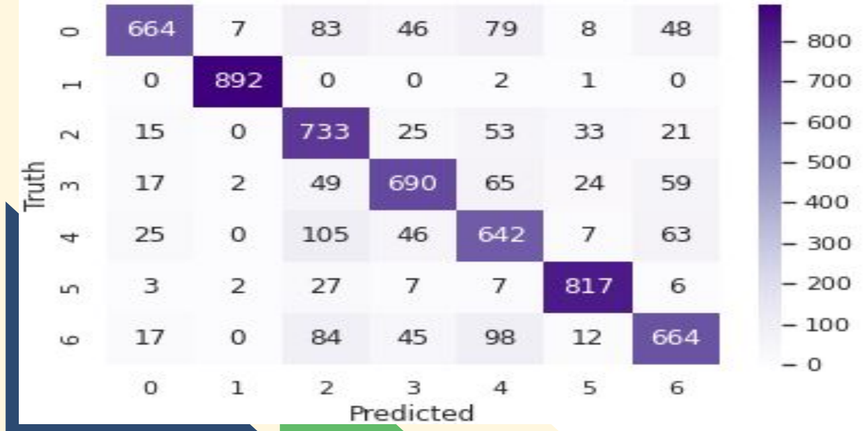


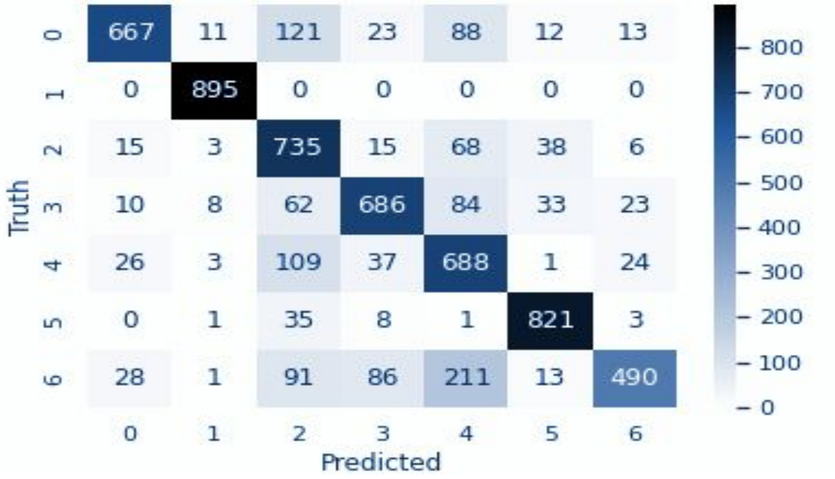




FER 2018

FER 2013

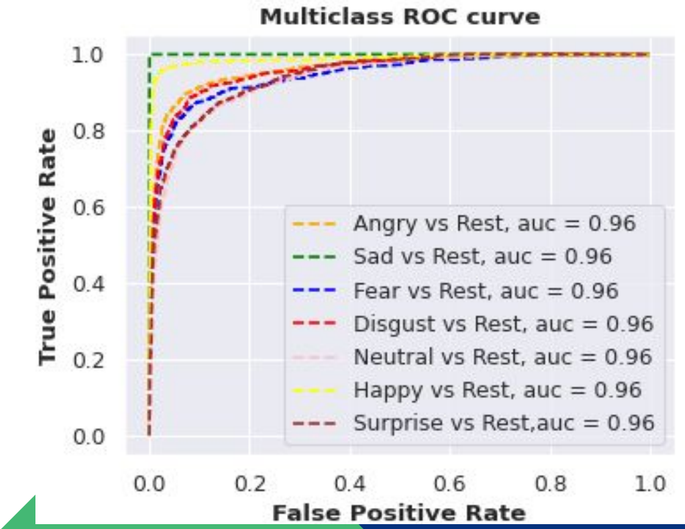




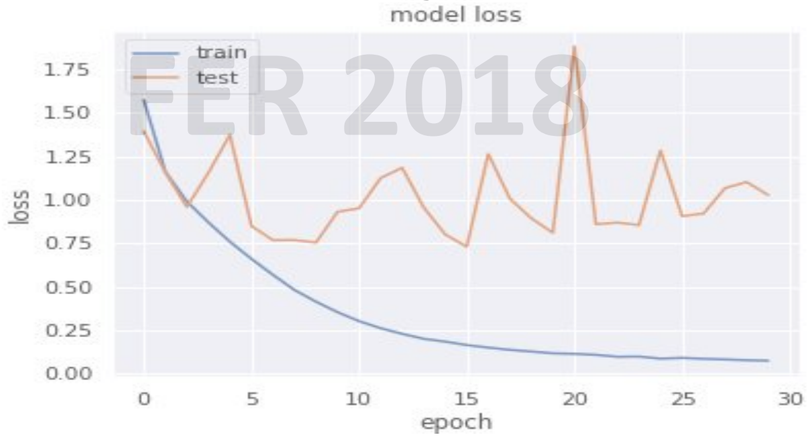
CNN+LSTM:

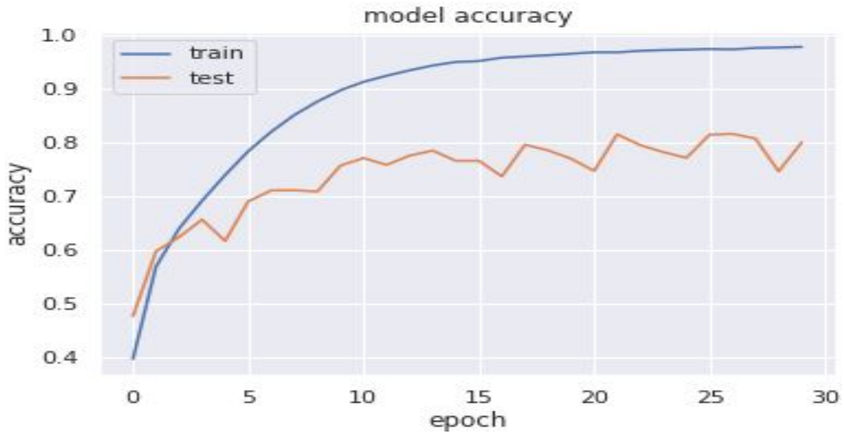
FER 2013-
Training Accuracy: 97.4%
Validation Accuracy: 79.7%
F1 Score: 0.79

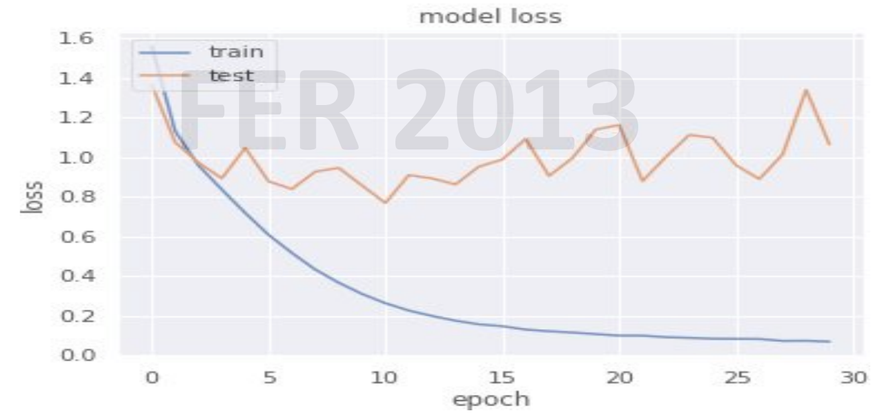
FER 2018-
Training Accuracy: 97.33%
Validation Accuracy: 78.33%
F1 Score: 0.78



CNN+BiLSTM:

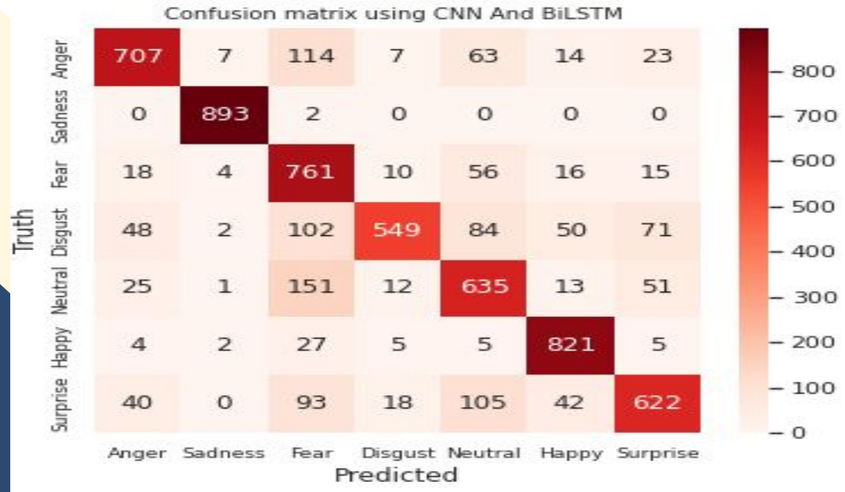


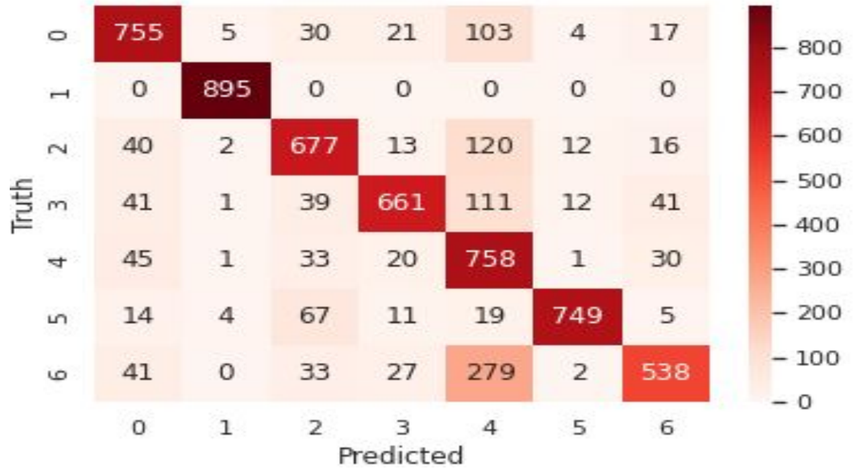


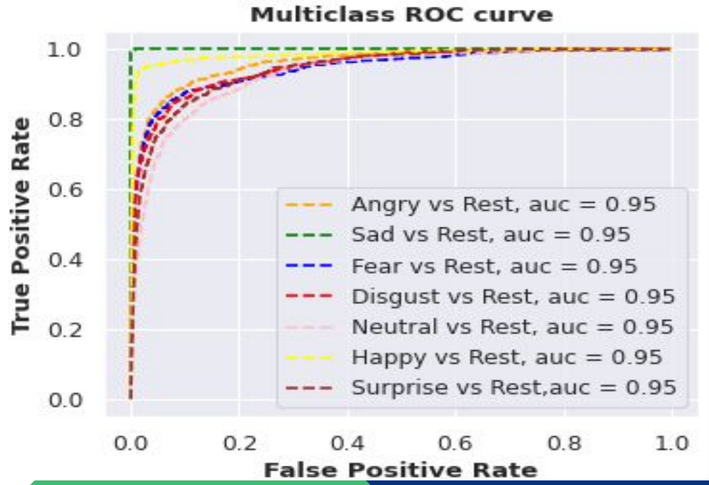


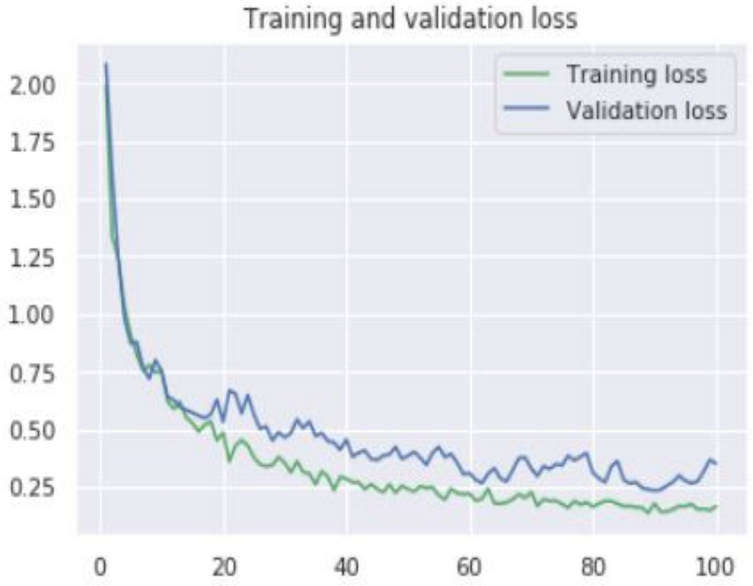
FER 2013-
Training Accuracy: 97.77%
Validation Accuracy: 79.97%
F1 Score: 0.80

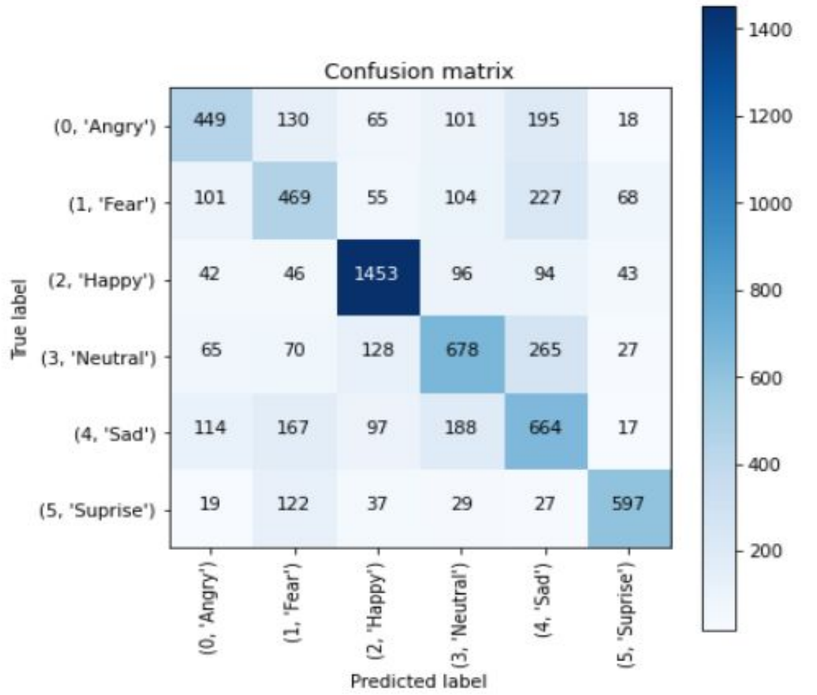
FER 2018-
Training Accuracy: 97.54%
Validation Accuracy: 80.1%
F1 Score: 0.81

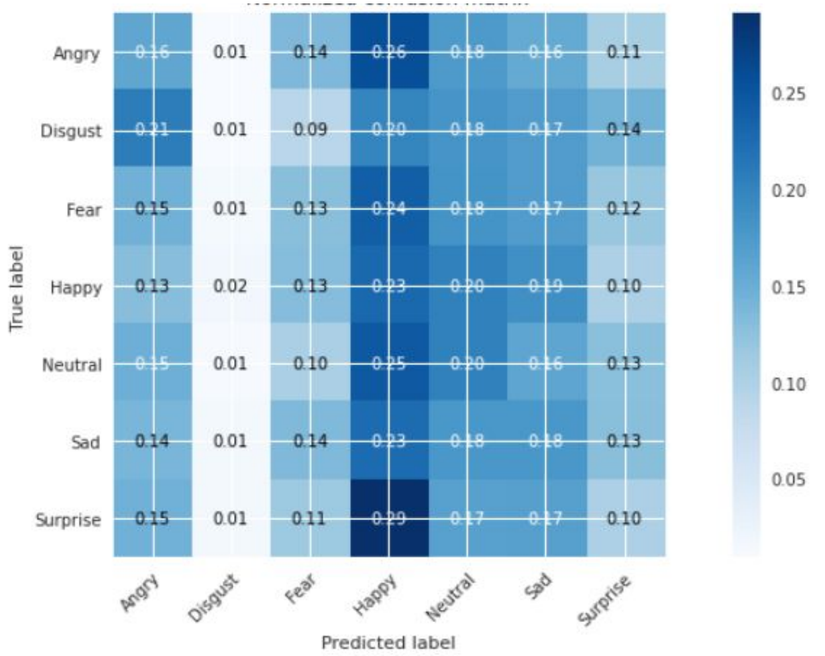








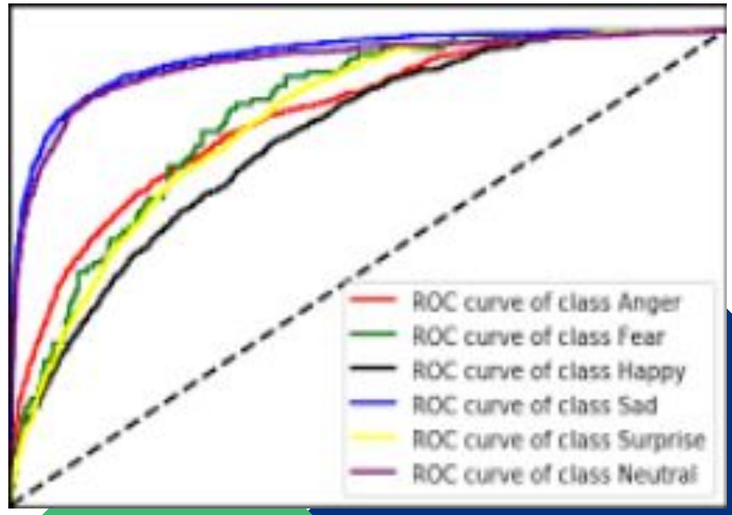


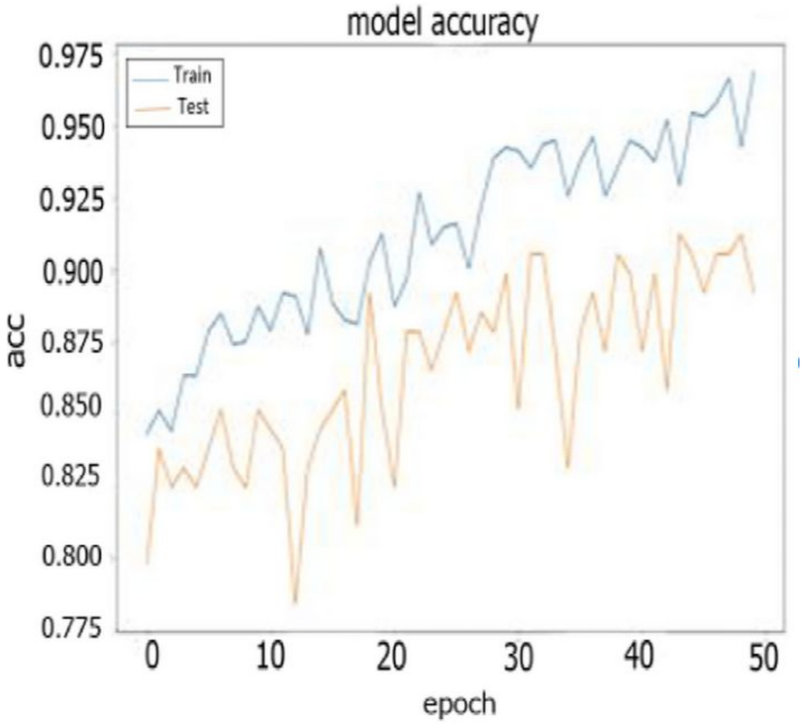


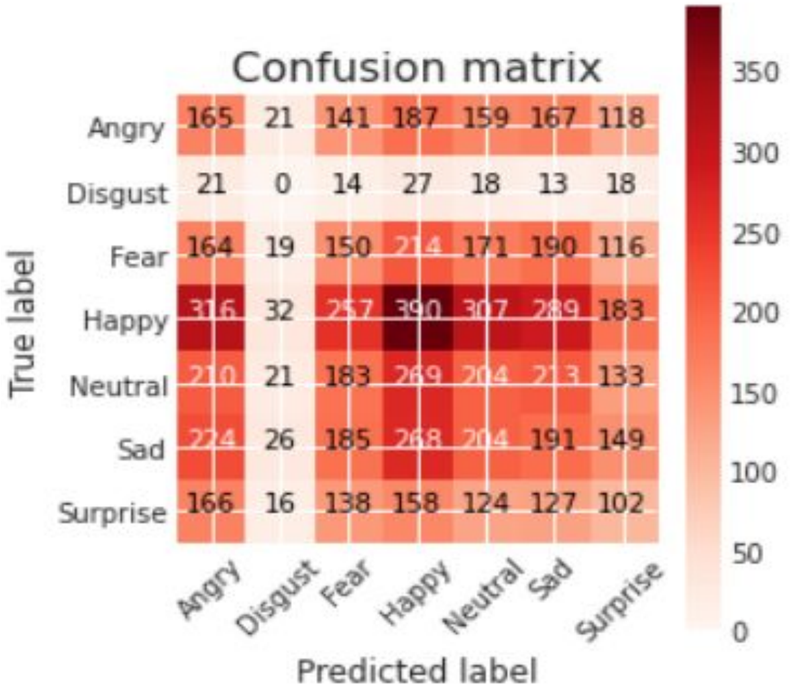
VGG16:

FER 2013-
Training Accuracy: 95%
Validation Accuracy; 82.8%
F1 Score: 0.64

FER 2018-
Training Accuracy: 95%
Validation Accuracy: 83%
F1 Score: 0.68



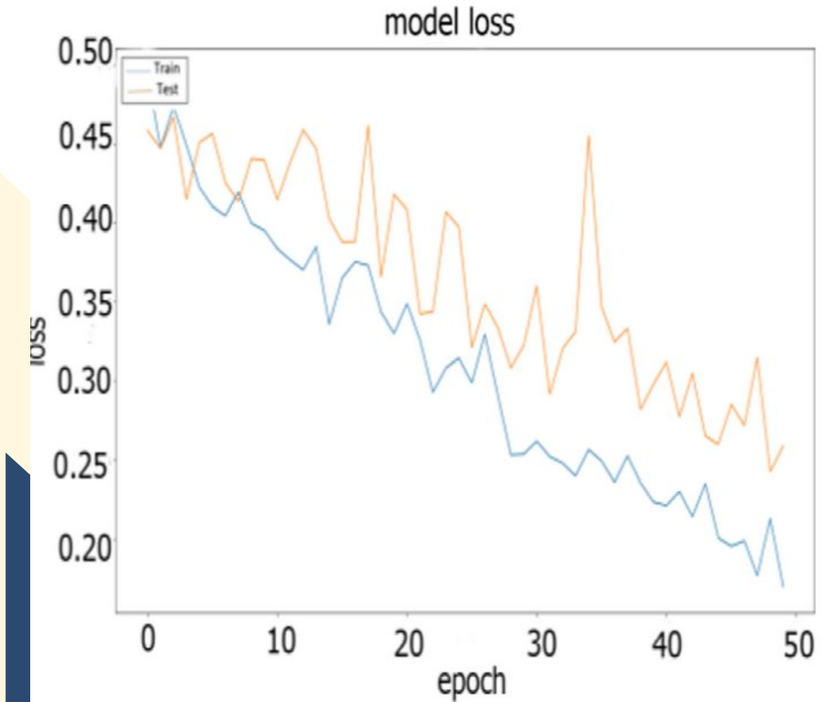


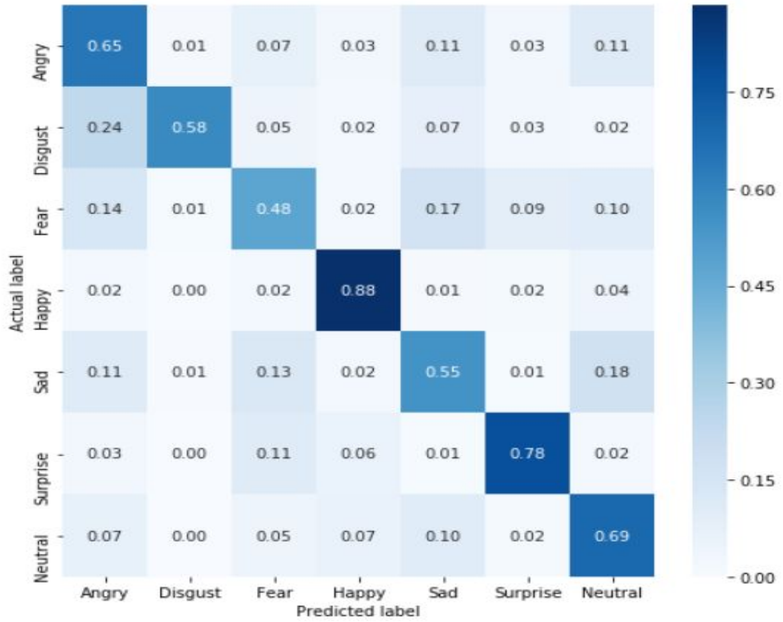


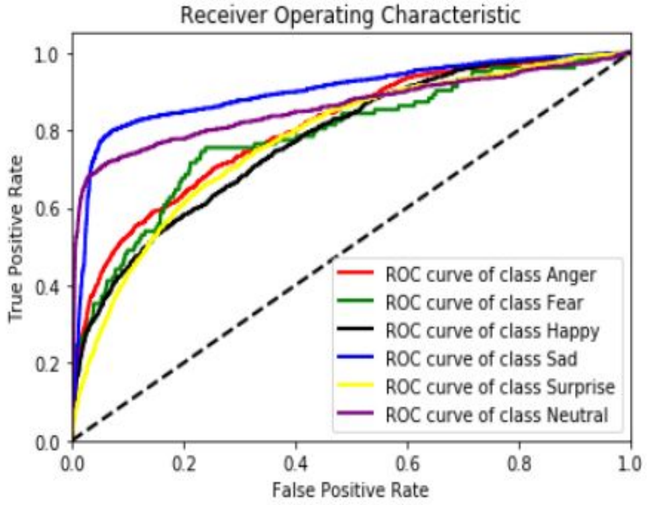
VGG19:

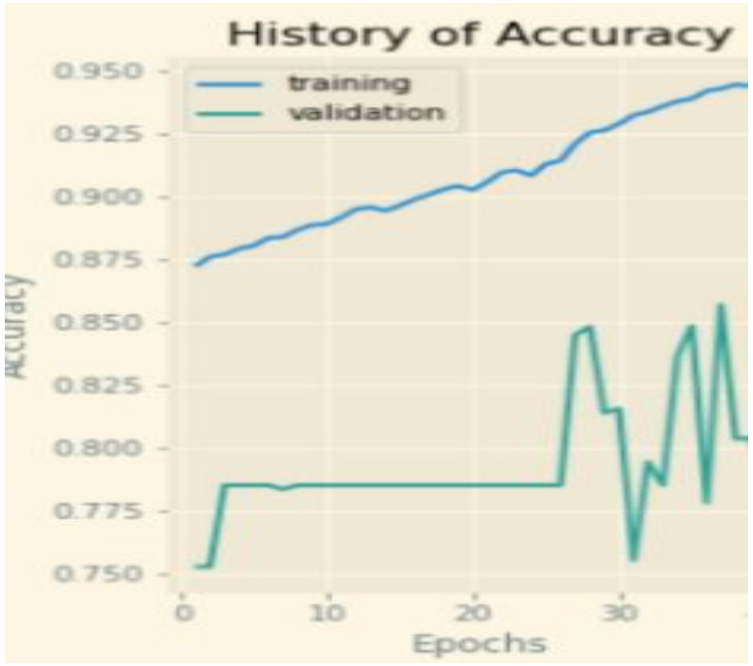
FER 2013-
Training Accuracy: 95%
Validation Accuracy; 84%
F1 Score: 0.61

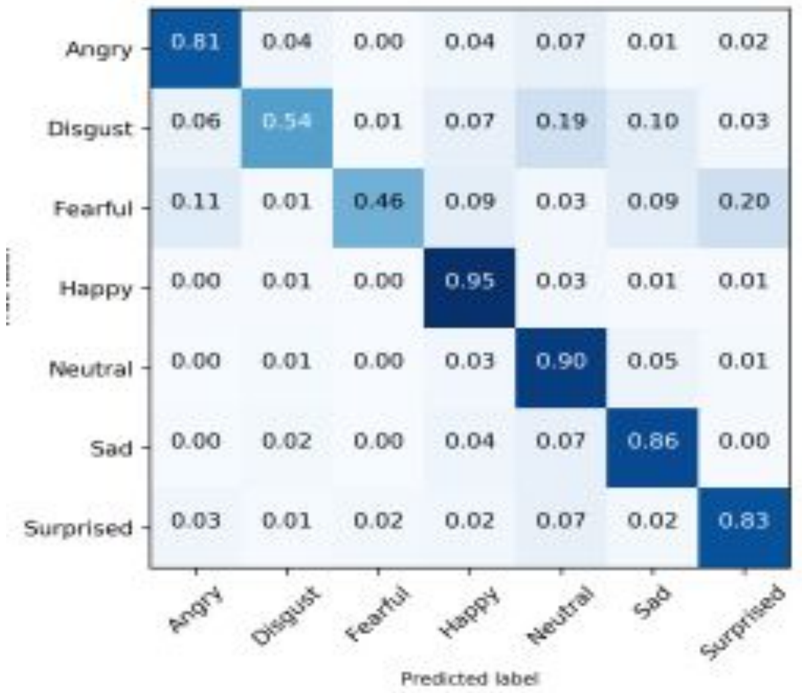
FER 2018-
Training Accuracy: 96%
Validation Accuracy: 84%
F1 Score: 0.66





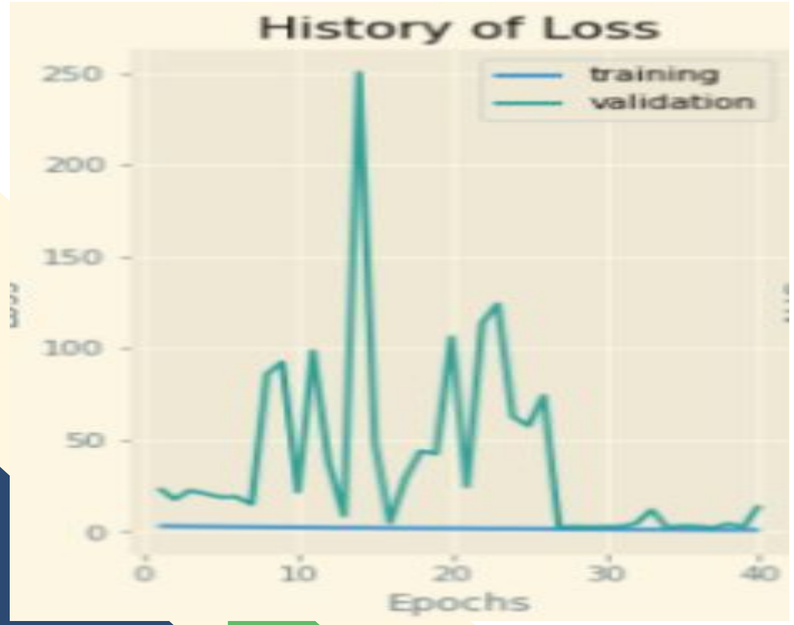


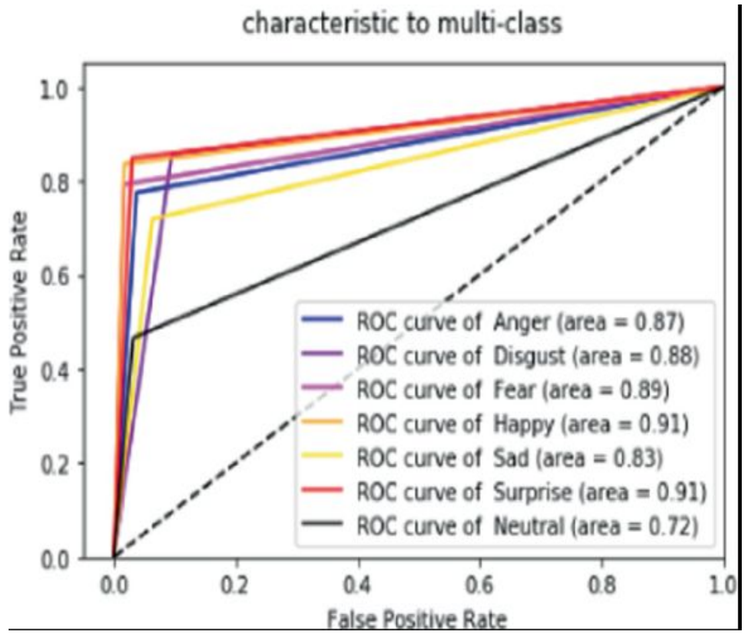




ResNet-50:

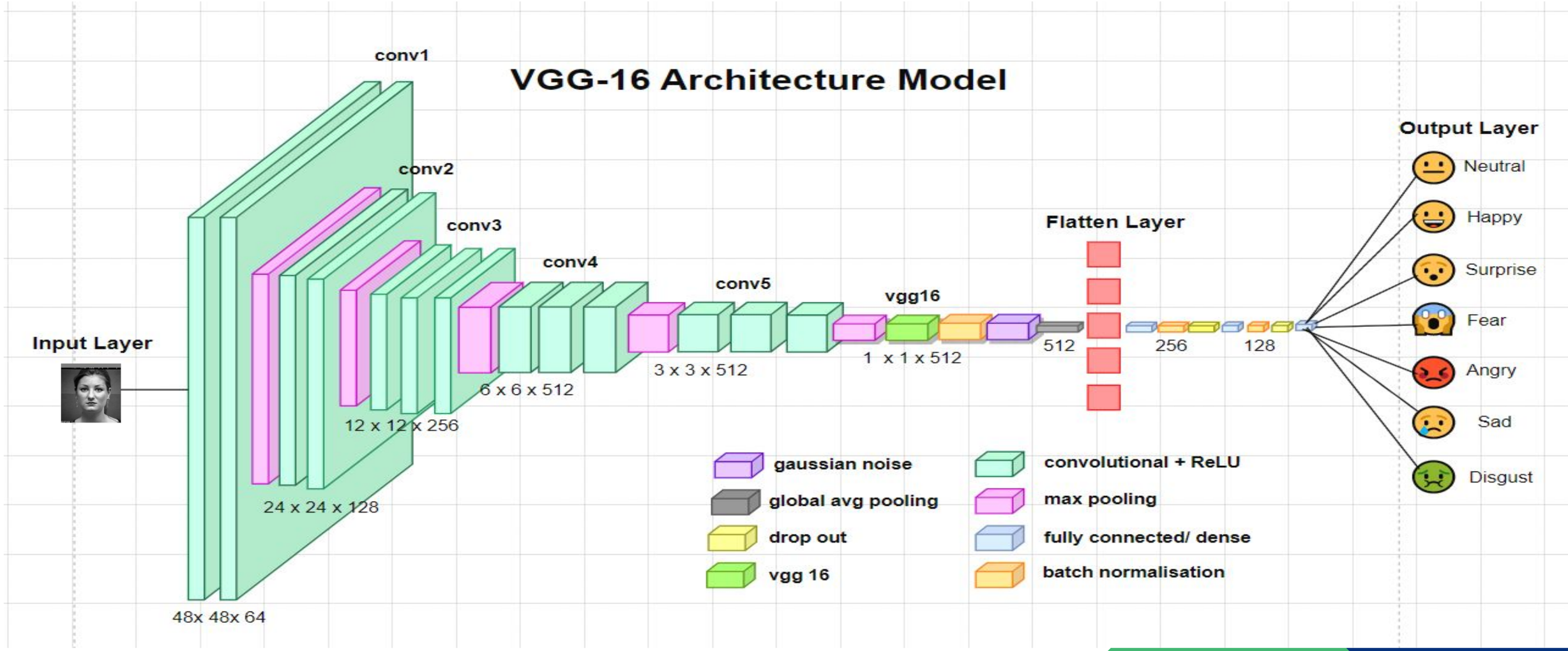
Ck+ 48-
Training Accuracy: 96%
Validation Accuracy: 85%
F1 Score: 0.61





Proposed Method

The basic VGG16 model is overfit. So we tried to remove it by adding further layers with VGG16 and achieved an accuracy of 90% with validation accuracy of 97% under FER2013 dataset.



CONCLUSION

- We tuned our parameters, increased layers, and applied max-pooling, dropout layers to obtain good accuracy.
- We observed a combination of CNN and LSTM provides a slighter enhanced result than simple CNN model.
- The results of Bidirectional LSTM with CNN had similar results as CNN+LSTM.
- Then, we applied the VGG16, VGG19, and ResNet50 transfer learning methods where we observed comparatively enhanced results after adding extra layers over the basic models.
- **Applications**: Video games, robots, medical pain analysis,

THANK YOU