# COSC 2779 | Deep Learning Assignment 1 [Submitted by: Jyoti, s3880522]

### Assignment 1: Introduction to Deep Convolutional Neural Networks

### (Action Unit and Emotion Recognition)

## Introduction

This report delves into the exploration of the domain of facial expression analysis using deep learning techniques. The primary objective is to present the results, methodologies, and insights obtained from the development of a Convolutional Neural Network (CNN) – Rest net model capable of simultaneously predicting high-level emotions [Positive, Negative, Surprise] and detecting Facial Action Coding System (FACS) codes within facial images. The report draws upon a curated subset of the CK+ dataset, consisting of 560 labeled images from 123 subjects, as a core dataset.

## Background

Facial expression analysis through deep learning has gained significant importance due to its applications in computer vision, machine learning, and behavioural sciences. This technology finds use in various fields, including security, human-computer interaction, driver safety, and healthcare. It is also the subject of ethical discussions regarding privacy, bias, and transparency. This project presents an opportunity to explore the capabilities and ethical considerations of emotion recognition technology.

## Project Objective

In this assignment, I aim to develop a deep convolutional neural network (CNN) to identify facial emotions and Facial Action Coding System (FACS) codes in images.

- High Level Emotion: Does the image show “Positive”, “Negative” or “surprised” subject?
- FACS codes: Does the image show a particular facial action (FACS code)? This is a yes(1)/no(0) output.

## Data Analysis & Preprocessing

1. **Exploration of Extended Cohn-Kanade (CK+) Dataset & Data Preprocessing:** A numerical representation - "emotion_num" is created for high level emotions to facilitate machine learning algorithms. Certain columns, such as 'sequence,' 'image_index,' 'file_prefix,' and demographic columns, are considered for removal to simplify the dataset.
2. **Image Visualization and Data Description:** Images are displayed using PIL and Matplotlib to visually inspect a subset of the dataset. Image information includes file path, emotional label, numerical emotion representation, image size, colour mode, and format. Unique image sizes are observed: (720, 480), (640, 490), and (640, 480).
3. **Action Units Analysis & Correlation:** I observed that three unique emotions are present: 'negative,' 'surprise,' and 'positive.' The dataset includes rows with specific emotions and associated Action Units (AU) values. During Heatmap visualization, it is observed that there are Strong positive correlations are found between AUs 1 and 2, AUs 1 and 27, AUs 2 and 27, AUs 6 and 12, and AUs 23 and 24. Strong negative correlations are observed between AUs 11 and 26, and AUs 14 and 15.
4. **Data Splitting:** I split CK dataset into three subsets: training, validation, test sets. The training set consists of 60% of original dataset. The validation set and test each make up 20% of original dataset. Why Validation data : The validation set aids in hyperparameter tuning, model selection, preventing overfitting, and assessing how well my model generalizes to new data, all of which are essential for building robust and effective models.
5. **Data Loader:** I have defined a DataGenerator class to prepare batches of image data with associated labels for training deep learning models. It incorporates various data preprocessing, augmentation, and normalization steps to improve the model's ability to learn from the CK dataset, specifically for tasks involving FACS labels and emotion detection.

- Output layers: I have added 2 output layers, one to predict emotions, 2nd to predict FACS.
- Batch size : Instead of a 32 batch size, I chosen a batch size of 16 for the efficient model.
- The on_epoch_end method shuffles data indices after each epoch to ensure diverse data
- Normalizing the data using the provided mean and standard deviation.

## Data Augmentation

Augmentations is employed to enhance model robustness through a parameter as part of data generation. It is only applied to train data but not to validation and test sets. The validation and test sets are meant to assess how well the model generalizes to new, unseen data. If augmentations were applied to these sets, it could introduce data leakage.

- **Augmentation for Emotion Labels:** Augmentations enhance the model's ability to recognize facial expressions at various orientations and positions. The padding ensures consistent image dimensions, accommodating variations in image sizes. Brightness adjustment makes the model robust to varying lighting conditions, crucial for accurate emotion recognition.
- **Augmentation For FACS Labels:** FACS labels represent specific facial muscle activations. Augmentations simulate changes in head angles, positioning, and lighting conditions. This helps the model understand how different action units are activated under real-world variations.

## Baseline model [ResNet Model]

I chose ResNet model for facial emotion recognition and FACS label due to its ability to handle deep architectures, extract meaningful features, leverage pre-trained models, and achieve state-of-the-art performance**.** It came as a compelling choice for building a robust and accurate facial analysis system.

**A Custom CNN layer as a residual block:** While I can use the sequential/functional API with the existing blocks in TensorFlow/keras to build a large model, the code will become unreadable when the network size increases. As a solution, I created a custom layer for a local structure to repeat as follows.

- conv1: First convolutional layer with specified filters, kernel size (3x3), and stride.
- bn1: Batch normalization layer applied after the first convolution.
- conv2: Second convolutional layer with the same filter count and kernel size.
- bn2: Batch normalization layer applied after the second convolution.

**ResNet Model Architecture**: My model consists of an input layer, convolutional layers, 12 residual blocks with varying numbers of convolutional layers and filter sizes, a global average pooling layer, a flattened layer.
**Input Layer:** Input shape: (490, 490, 3)
**Convolutional Layers:** Conv2D Layer with 64 filters, kernel size (3x3), and ReLU activation. There is Batch Normalization Layer after Conv2D.
**Residual Blocks (Total of 12 Blocks**): Each block includes multiple Conv2D layers with 64 filters, kernel size (3x3), and ReLU activation.
Batch Normalization is applied within each residual block.
The number of filters and spatial dimensions may change as the blocks progress.
Blocks 1-6 have 64 filters, and spatial dimensions are reduced to half (e.g., from 490x490 to 245x245).
Blocks 7-12 have 256 filters.
**Global Average Pooling Layer:** Reduces the spatial dimensions to 1x1, resulting in a shape of (256,).
**Flatten Layer:** Converts the 1x1 spatial dimensions into a flat vector of length 256.

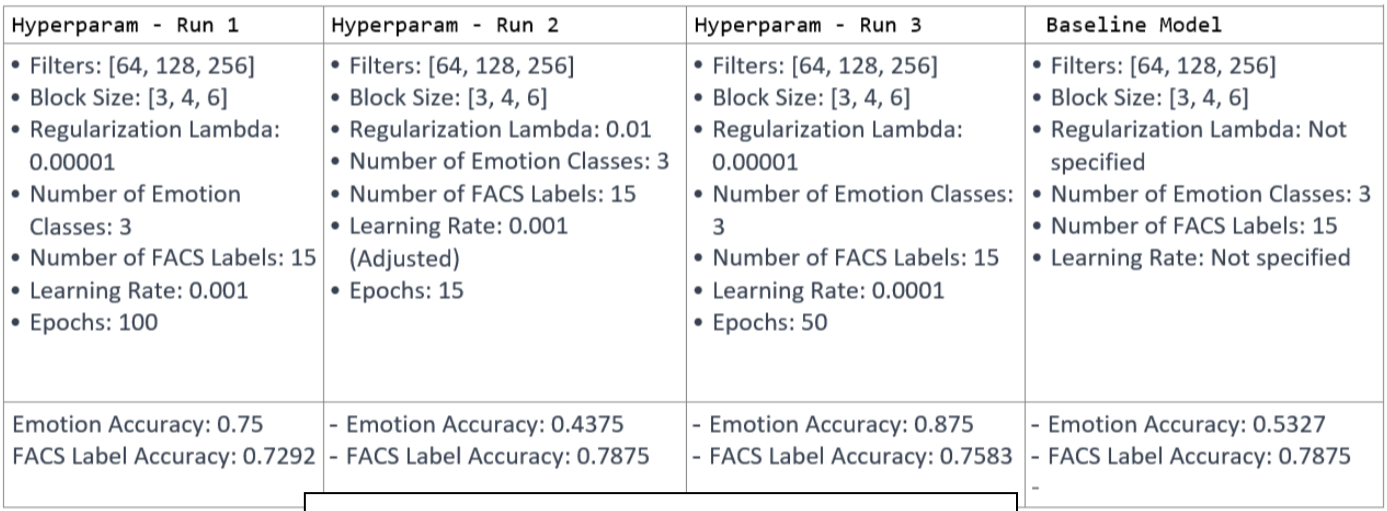
**Hyper parameters selection strategy:** I have used Regularization lamb and changed with no of epcohs. I observed that my model is not efficient with 100 epcohs. Epoch 75 & 50 are giving me good accuracy and less loss. Also, with Regularization Lambda less the 0.00001 is not giving a good model and score outcome.

- Filters: I used filter sizes of 64, 128, and 256 to influence model depth and complexity.

- Block Size: I have residual blocks have sizes of 3, 4, and 6, affecting model depth and expressiveness.
- Regularization Lambda: A small value of 0.00001 to 0.1 is applied for regularization.
- Learning Rate: I set the learning rate to 0.001 for optimization and a scheduler to adjust the learning rate during training.

**Model Compilation:** I have used Adam optimizer, Two different loss functions for emotion output (categorical_crossentropy) and the FACS output (binary_crossentropy). In addition to it, I used F1 score metrics for both tasks along with accuracy.

**Experiments Evaluation:** Below data shows ResNet model outcome at an initial run.



| Hyperparam - Run 1 | Hyperparam - Run 2 | Hyperparam - Run 3 | Baseline Model |
|---|---|---|---|
| • Filters: [64, 128, 256]<br>• Block Size: [3, 4, 6]<br>• Regularization Lambda: 0.00001<br>• Number of Emotion Classes: 3<br>• Number of FACS Labels: 15<br>• Learning Rate: 0.001<br>• Epochs: 100 | • Filters: [64, 128, 256]<br>• Block Size: [3, 4, 6]<br>• Regularization Lambda: 0.01<br>• Number of Emotion Classes: 3<br>• Number of FACS Labels: 15<br>• Learning Rate: 0.001 (Adjusted)<br>• Epochs: 15 | • Filters: [64, 128, 256]<br>• Block Size: [3, 4, 6]<br>• Regularization Lambda: 0.00001<br>• Number of Emotion Classes: 3<br>• Number of FACS Labels: 15<br>• Learning Rate: 0.0001<br>• Epochs: 50 | • Filters: [64, 128, 256]<br>• Block Size: [3, 4, 6]<br>• Regularization Lambda: Not specified<br>• Number of Emotion Classes: 3<br>• Number of FACS Labels: 15<br>• Learning Rate: Not specified |
| Emotion Accuracy: 0.75<br>FACS Label Accuracy: 0.7292 | - Emotion Accuracy: 0.4375<br>- FACS Label Accuracy: 0.7875 | - Emotion Accuracy: 0.875<br>- FACS Label Accuracy: 0.7583 | - Emotion Accuracy: 0.5327<br>- FACS Label Accuracy: 0.7875<br>- |

Table1 : Analysis of output with diff hyper param

## Result Analysis & Ultimate Judgement

**Overfitting (Run 3):** Run 3 exhibits signs of overfitting, where the model performs exceptionally well on the training data but struggles to generalize to the validation set. The high emotion accuracy (0.875) on the training data is significantly better than the validation accuracy (0.7946). This suggests that the model has learned to fit the training data too closely, capturing noise or specific patterns that do not generalize well.

**Underfitting (Run 2):** Run 2 seems to suffer from underfitting. The low emotion accuracy (0.4375) on the training data indicates that the model did not learn the underlying patterns in the data effectively. This may be due to the relatively high regularization lambda (0.01) and insufficient training epochs (15). The model's performance on both training and validation data is suboptimal.

**Balanced (Run 1 and Run 4):** Runs 1 and 4 demonstrate more balanced performance. While they achieve a reasonable emotion accuracy on the training data (0.75), their generalization to the validation data is consistent (0.8839 and 0.8839, respectively). So, model has learned patterns without overfitting or underfitting.

**Ultimate judgment:** Run 1 and Run 4 strike a balance between overfitting and underfitting, showcasing the importance of careful hyperparameter selection for achieving good model performance. Run 3 overfits the data, while Run 2 underfits.

**Accuracy & F 1 score Metrics For emotion prediction:** The precision score of 0.7656 indicates that when the model predicts an emotion, it is correct approximately 76.56% of the time. The recall score of 0.625 suggests that the model captures approximately 62.5% of all actual positive emotions.The F1 score of 0.4808 is the harmonic mean of precision and recall, providing a balance between the two.

**Accuracy & F 1 score Metrics For FACS label detection:** The weighted FACS label precision score of 0.9778 indicates that when the model predicts FACS labels, it is correct approximately 97.78% of the time. Weighted precision considers class imbalance. The weighted FACS label recall score of 0.3111 suggests that the model captures approximately 31.11% of all actual positive FACS labels, again considering class imbalance. The weighted FACS label F1 score of 0.3111 is the harmonic mean of precision and recall for FACS label detection.

**Limitations of the model:** Very high loss, Data Imbalance, Limited Generalization, Model Complexity, Interpretable Features as black box, Resource Requirements, Robustness to Noisy Data, Ethnic and Gender Bias, Limited Data for FACS Labels. To mitigate some of ResNet model limitations, I can explore techniques such as transfer learning, fine-tuning on larger and more diverse datasets, and experimenting with different model architectures like MobileVnet. Additionally, addressing data quality and balance issues will be critical for improving the robustness and accuracy of your emotion and FACS label detection model.

**Identified Ethical issues and biases**: In the provided dataset for facial emotion recognition for this project, there is a noticeable bias in the distribution of emotion labels. Specifically, the 'negative' emotion label occurs more frequently than the other emotion labels, with 328 occurrences, while 'positive' and 'surprise' appear less frequently with 121 and 111 occurrences, respectively. As it be seen by histogram fig 1 clearly depicts the same. Negative emotions are double the other two.

**Ethnic Bias:** The dataset is predominantly composed of Euro-American individuals (81%).

**Gender Bias:** The dataset is biased toward females, with 69% of the participants being female. This gender imbalance may affect the generalizability of model.

**Age Bias**: The dataset includes participants aged between 18 and 50 years. The exclusion of younger or older age groups may limit the applicability of the dataset to broader age ranges.

**Label Bias/ Expression Bias:** Participants were instructed to perform a series of 23 facial displays.

**Image View Bias:** The dataset includes image sequences for both frontal views and 30-degree views. it could impact the generalization of models trained on the dataset.

**Sampling Bias:** The participants in the dataset are those who agreed to participate in the study and followed the experimenter's instructions. This sampling bias may not fully represent the broader population's diversity in terms of facial expressions and emotions.

**Privacy concerns :**The provided abstract for the Extended Cohn-Kanade (CK+) dataset does not explicitly detail how all privacy and ethical concerns are addressed. Handling privacy and ethical concerns typically involves various measures, including obtaining informed consent, anonymizing data, ensuring data security, and adhering to relevant legal and ethical guidelines. It is the responsibility of the dataset creators and users to implement these measures to protect participant privacy and comply with ethical standards.

**Informed Consent:** Participants understand and consent to study's nature and consequences.

**Data Privacy:** For security of facial data to prevent misuse, I signed a data agreement to access it.

## Conclusions and Future Work

1. Adjusting hyperparameters can significantly impact a model's performance and reduced losses
2. No. of epcohs also palyed a mojor role. Any epoch less than 10 was not training model well at all.
3. Regularization in Hyper param :A lower regularization lambda of 0.0001 resulted in the best overall performance in terms of emotion accuracy and competitive FACS label accuracy. A higher regularization lambda of 0.1 led to lower emotion accuracy, indicating that strong regularization may hinder the model's ability to capture essential patterns in the data. A very low regularization lambda of 0.00001 performed reasonably well but required significantly more training epochs to achieve similar results to the optimal lambda of 0.0001

In conclusion, the emotion prediction model demonstrates reasonably good performance, with precision and recall scores indicating a balance between precision and recall. However, there is room for improvement in predicting emotions, as the F1 score could be higher. On the other hand, the FACS label detection model shows high precision but relatively low recall, which suggests that it tends to make accurate predictions for FACS labels but may miss some of them. Further optimization may be required to improve recall without sacrificing precision in FACS label detection. Overall, the model's performance can be considered promising but with potential for further refinement and optimization, especially in the context of both emotion prediction and FACS label detection.

## References:

1. *S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," in IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195-1215, 1 July-Sept. 2022, doi: 10.1109/TAFFC.2020.2981446.*
2. *P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression,"* 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, San Francisco, CA, USA, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.*

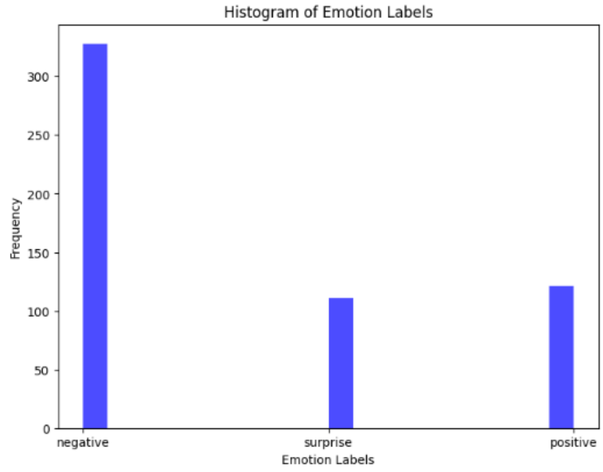


Fig 1 Imbalance of data

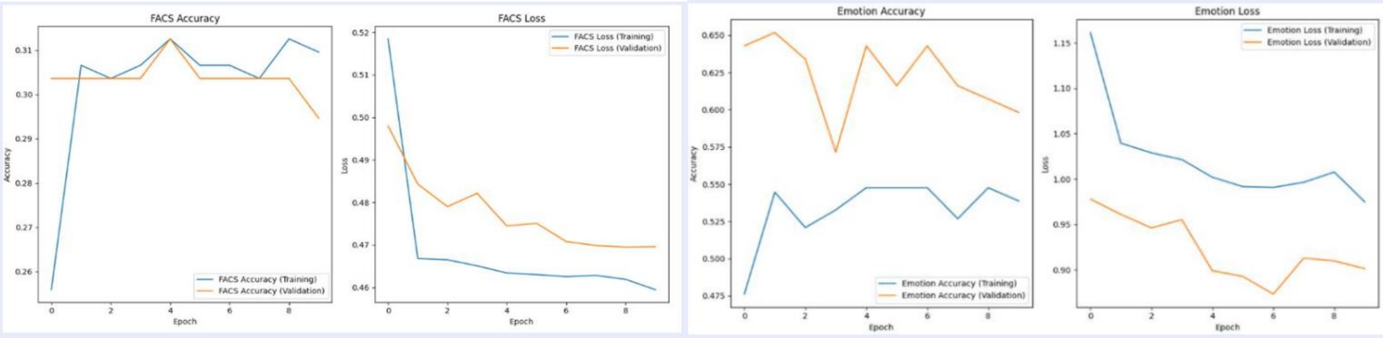
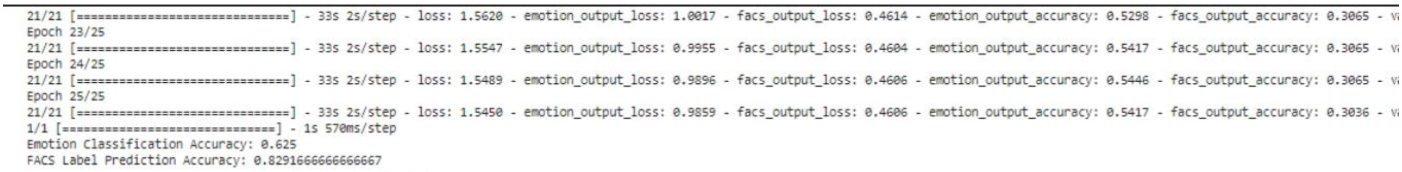
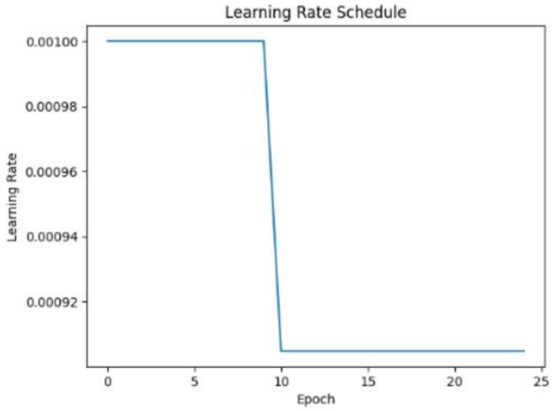


Fig 2: Baseline Model results



```
21/21 [==============================] - 33s 2s/step - loss: 1.5620 - emotion_output_loss: 1.0017 - facs_output_loss: 0.4614 - emotion_output_accuracy: 0.5298 - facs_output_accuracy: 0.3065 - va
Epoch 23/25
21/21 [==============================] - 33s 2s/step - loss: 1.5547 - emotion_output_loss: 0.9955 - facs_output_loss: 0.4604 - emotion_output_accuracy: 0.5417 - facs_output_accuracy: 0.3065 - va
Epoch 24/25
21/21 [==============================] - 33s 2s/step - loss: 1.5489 - emotion_output_loss: 0.9896 - facs_output_loss: 0.4606 - emotion_output_accuracy: 0.5446 - facs_output_accuracy: 0.3065 - va
Epoch 25/25
21/21 [==============================] - 33s 2s/step - loss: 1.5450 - emotion_output_loss: 0.9859 - facs_output_loss: 0.4606 - emotion_output_accuracy: 0.5417 - facs_output_accuracy: 0.3036 - va
1/1 [==============================] - 1s 570ms/step
Emotion Classification Accuracy: 0.625
FACS Label Prediction Accuracy: 0.8291666666666667
```



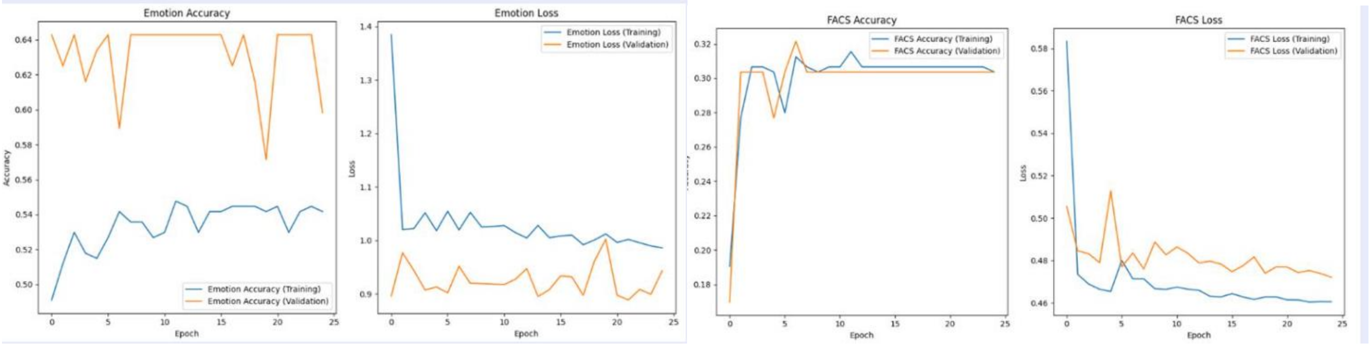


Fig 3: Run 3 Model results Imbalance

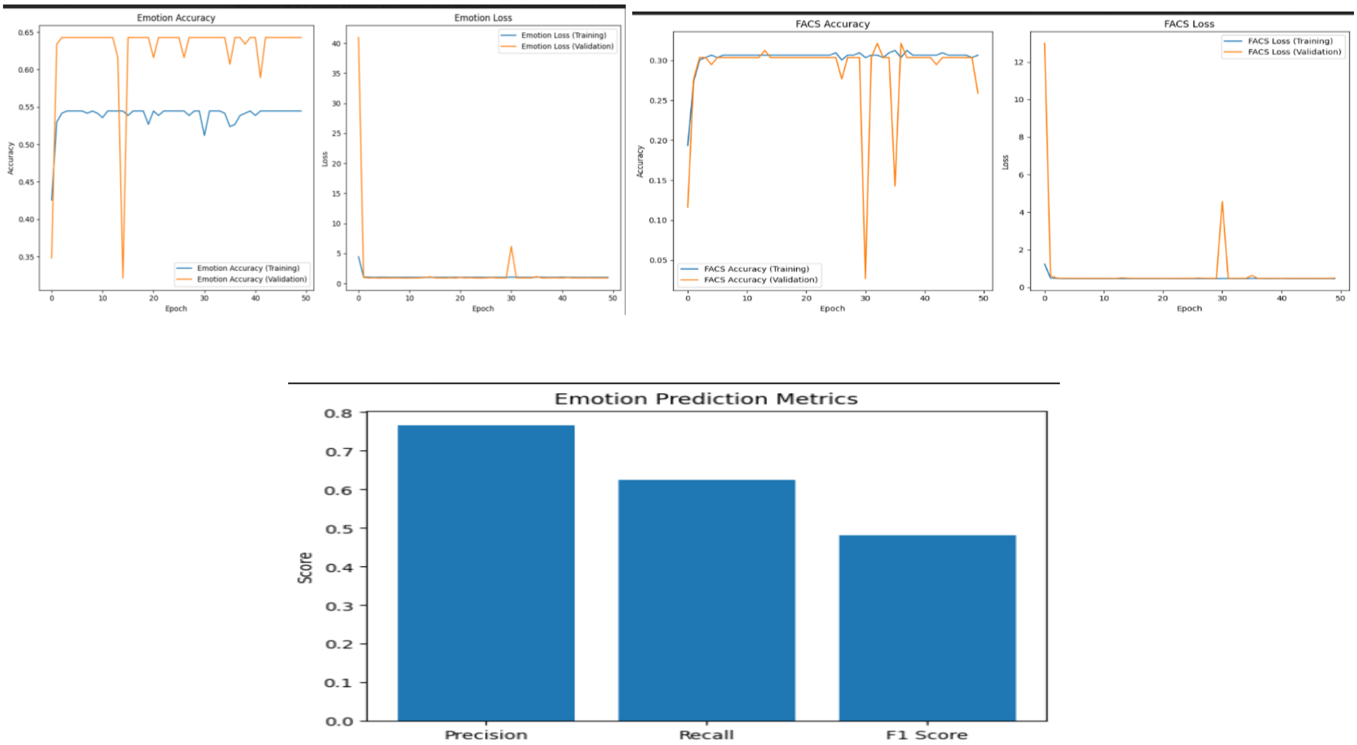


Fig 4 : Final Model results