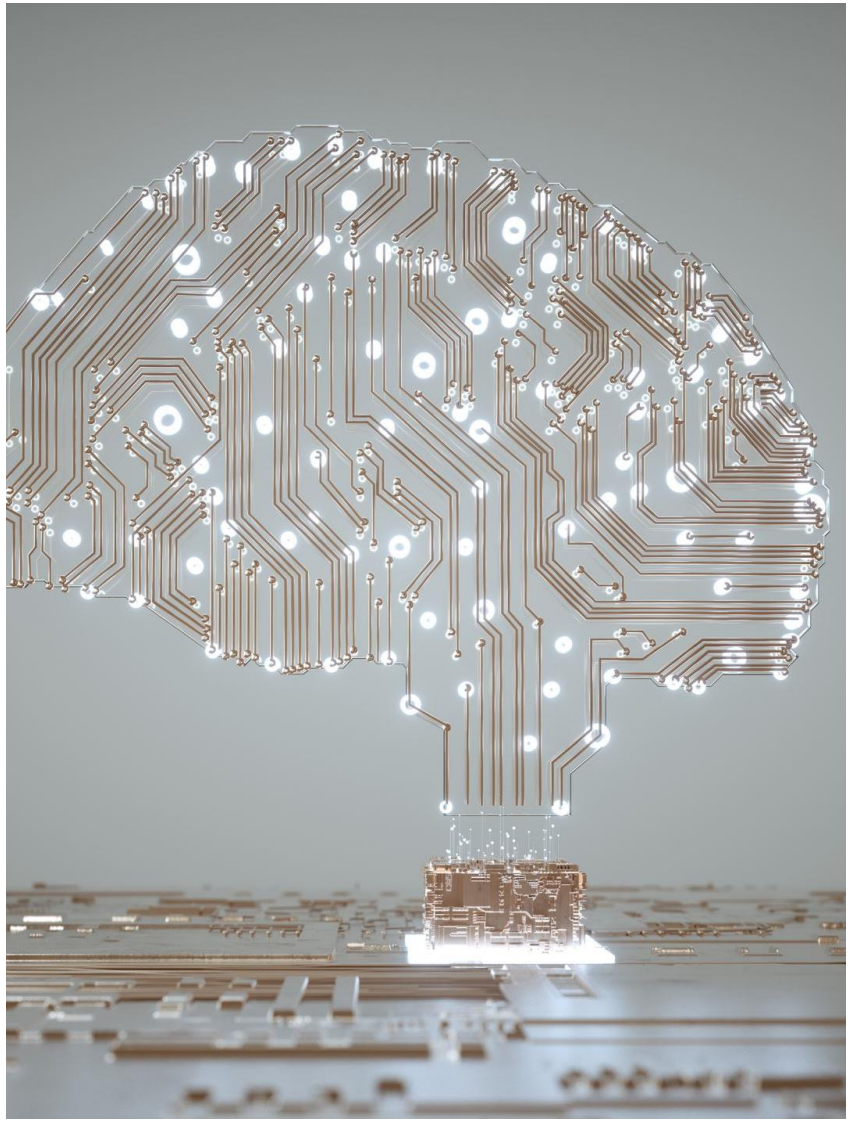


A.I.ducation Analytics

GROUP NAME : NS – 01

GROUP MEMBERS:

DARA RAHMAT SAMII (40281972)

NUMAN SALIM SHAIKH (40266934)

SHAHAB AMROLLAHIBIOKI (40292670)

Introduction

The goal of this project is to use PyTorch to create a Deep Learning Convolutional Neural Network (CNN) that can classify students into different states or activities based on the analysis of photos of them in a classroom or online meeting environment.

Stages of this Project :

- Data Collection, Cleaning, Labelling & Preliminary Analysis
- CNN Model Development and Basic Evaluation

Part 1

PREPROCESSING DATA

Data Collection, Cleaning, Labelling & Preliminary Analysis

The primary dataset employed in this project is the FER2013 dataset and FER+ dataset.

Comprising grayscale images, each measuring 48 × 48 pixels, the FER2013 dataset boasts an impressive total of 32,298 individual samples.

The FER2013 dataset categorizes emotions into seven distinct classes, each representing a unique spectrum of emotional expressions. These emotions encompass:

1. Angry 2. Disgust 3. Fear 4. Happy 5. Sad 6. Surprise 7. Neutral

FER+ is a crowed sourced scores for the given emotions.

Table 1.2: Sample of FER+ merged with FER2013

emotion	pixels	neutral	happiness	surprise	sadness	anger	disgust	fear	contempt	unknown	NF
Angry	70 80 ...	4	0	0	1	3	2	0	0	0	0
Angry	151 150 ...	6	0	1	1	0	0	0	0	2	0
Fear	231 212 8 ...	5	0	0	3	1	0	0	0	1	0
Sad	24 32 ...	4	0	0	4	1	0	0	0	1	0

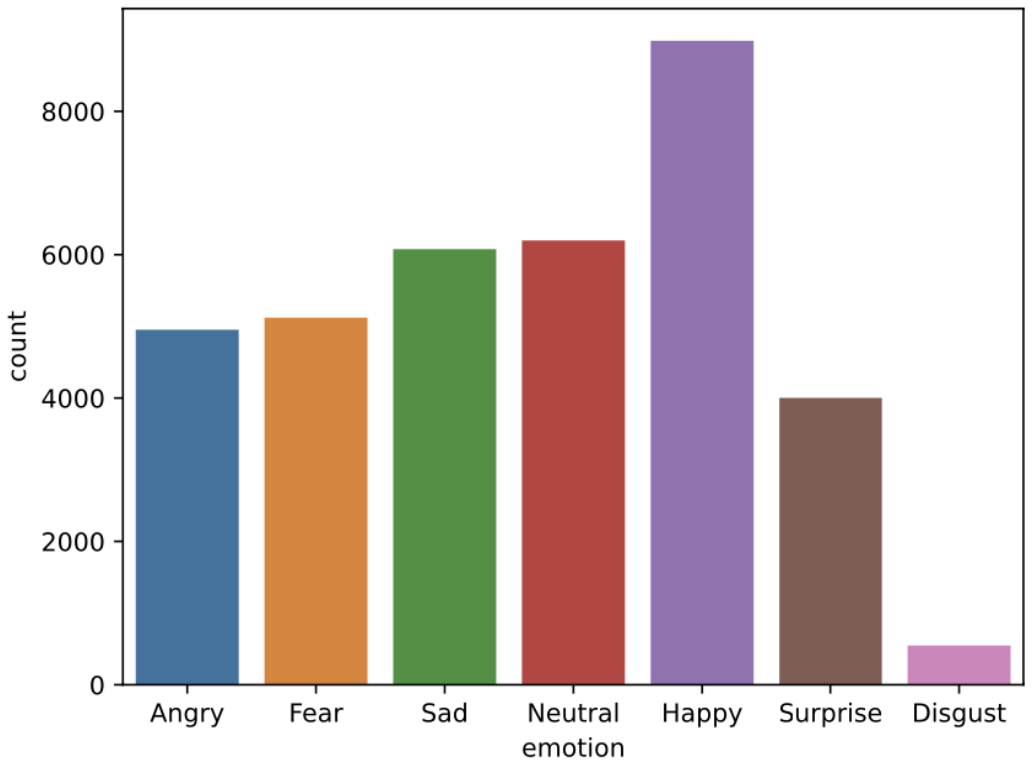
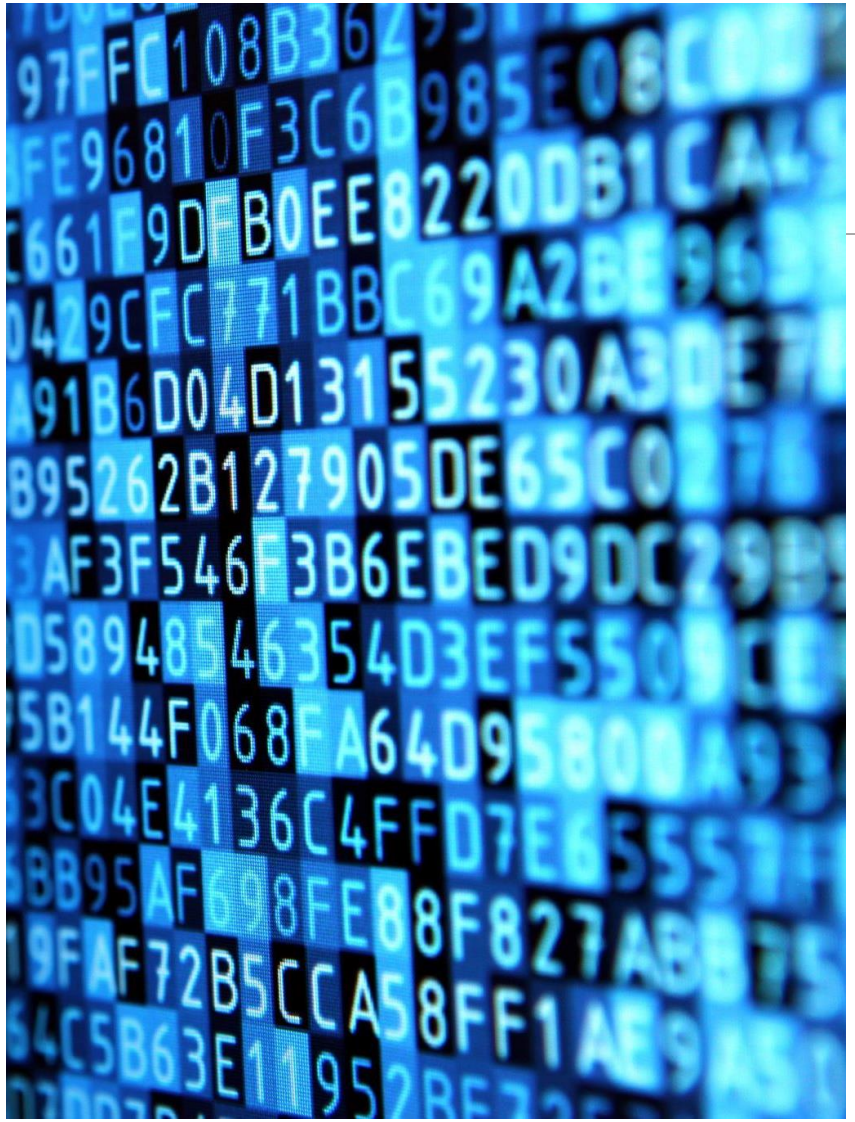


Figure 1.1: Number of images per each emotion in FER2013

Facial Expression Initial Dataset

Raw dataset containing image samples of various facial expressions. We will apply filters to transform the images as per our requirement



Cleaning the Data

The number of images with a score N F = 10 is 176.

The number of images with a score N F = 4 is 2.

The number of images with a score N F = 2 is 4.

The number of images with a score N F = 1 is 167.

number of images with score unknown=8 is 3.

number of images with score unknown=7 is 3.

number of images with score unknown=6 is 18.

number of images with score unknown=5 is 55.

Anger	• Anger score > other emotion's score • Anger score > 2
Neutral	• Neutral score > other emotion's score • Neutral score > 6
Focused	• sadness score == 0 • anger score == 0 • neutral score > other emotion's score
Bored	• happiness score == 0 • fear score == 0 • sadness score ≠ 0 • neutral score ≠ 0 • anger score + 2 < sadness score

Data classification criterion

- Number of samples labeled as 'Neutral': 3789
- Number of samples labeled as 'Angry': 3954
- Number of samples labeled as 'Focused': 4553
- Number of samples labeled as 'Bored': 3960

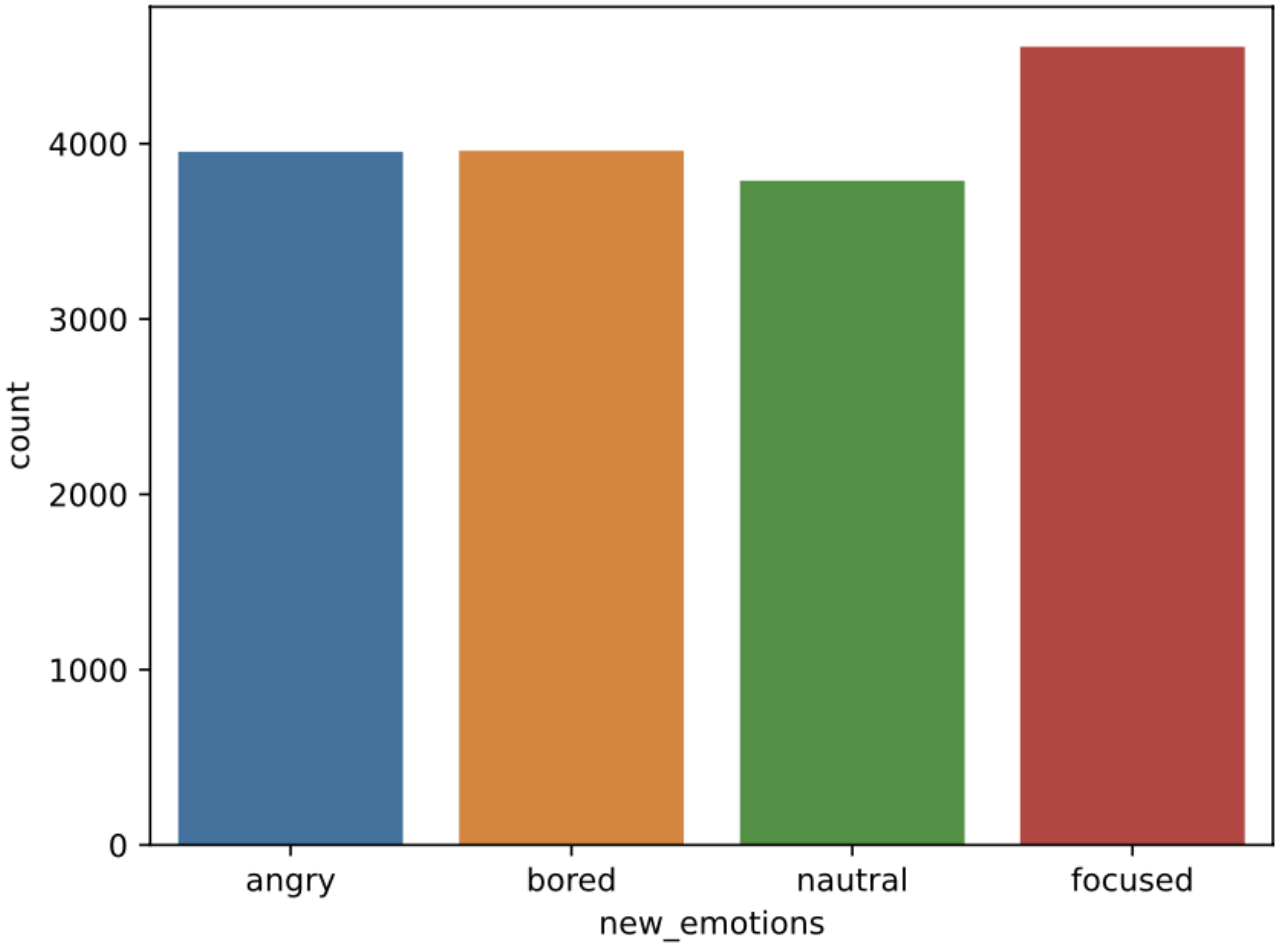


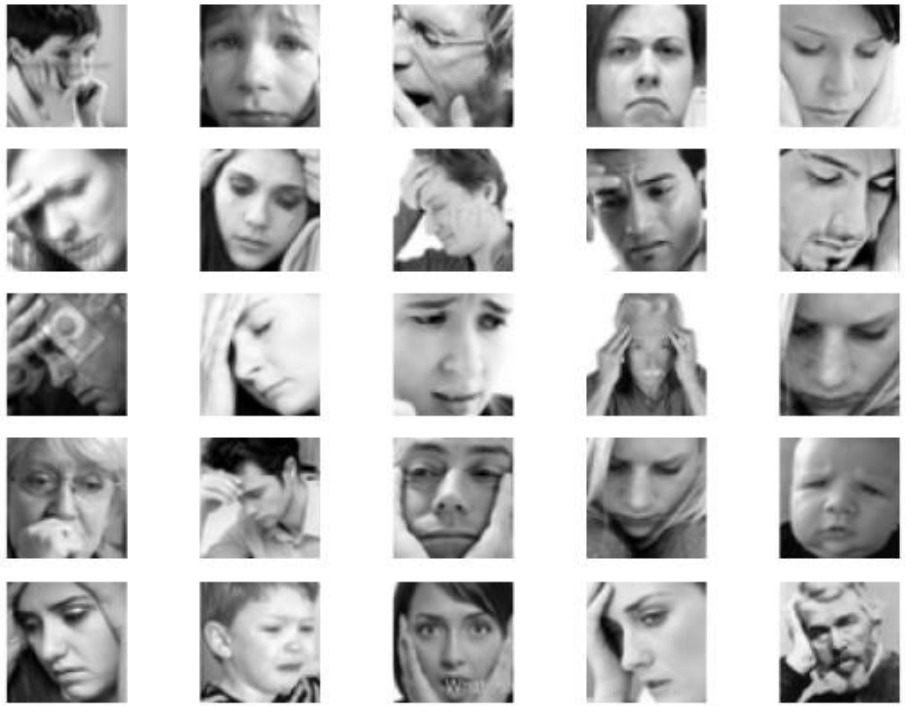
Figure 1.4: Bar-plot of images per desired emotion

Final facial expression dataset after cleaning

It is imperative to classify images into four distinct labels. While the original dataset inherently includes the emotions Anger and Neutral, it lacked explicit representations for Bored and Focused emotions.

Sample of Dataset

(a) Samples of images labeled as 'Angry'



(b) Samples of images labeled as 'Bored'

Sample of Dataset(Continue)

(c) Samples of images labeled as 'Neutral'

(d) Samples of images labeled as 'Focused'

Part 2

TRAINING THE MODELS

Model Architecture

TOTAL NUMBER OF 12 MODELS WERE TRAINED AND EVALUATED

Kernel Sizes and Padding

- (3 × 3): Padding was set to 1.
- (5 × 5): Padding was set to 2.
- (7 × 7): Padding was set to 3.

Number of Channels

- 16 channels
- 32 channels

Number of Resblocks

- 4 ResBlocks and 1 FC Layer
- 8 ResBlocks and 2 FC Layers

Table 2.1: Hyper Parameters used in Training

Parameter	Values
Learning Rate	0.001
Epochs	100
Batch Size	64
Optimizer	Adam
Loss Function	Cross-Entropy
Weight Decay	0.0001

Model Architecture

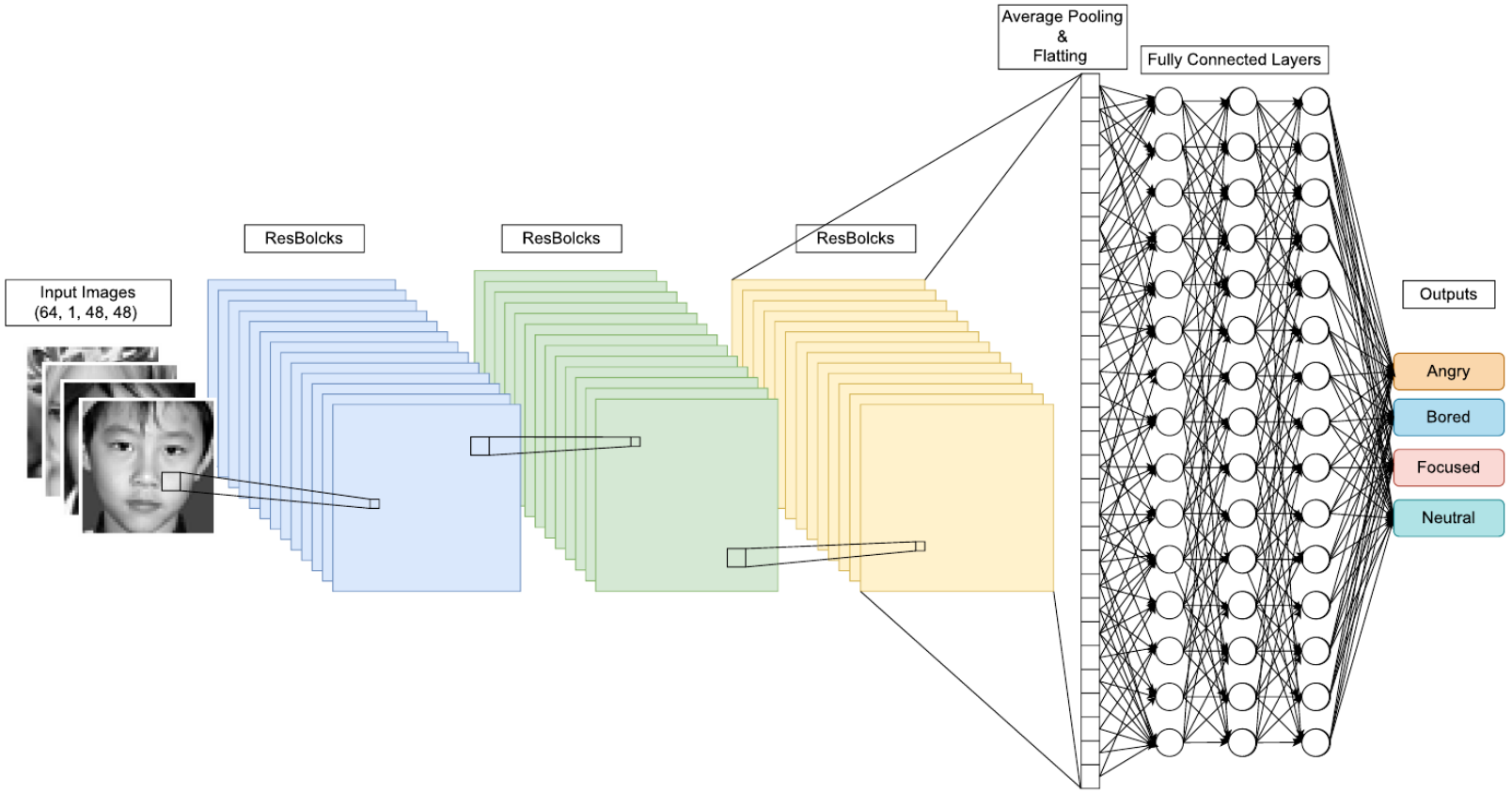


Figure 2.1: Schematic of Model Architecture

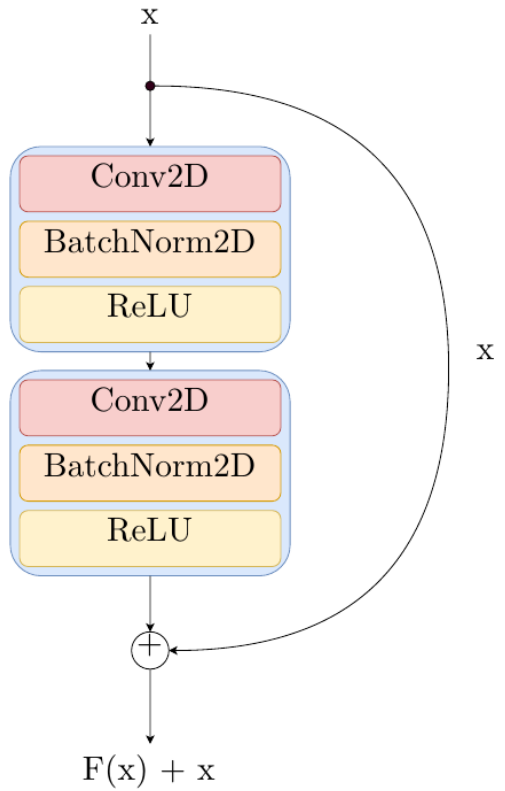


Figure 2.2: Resblock

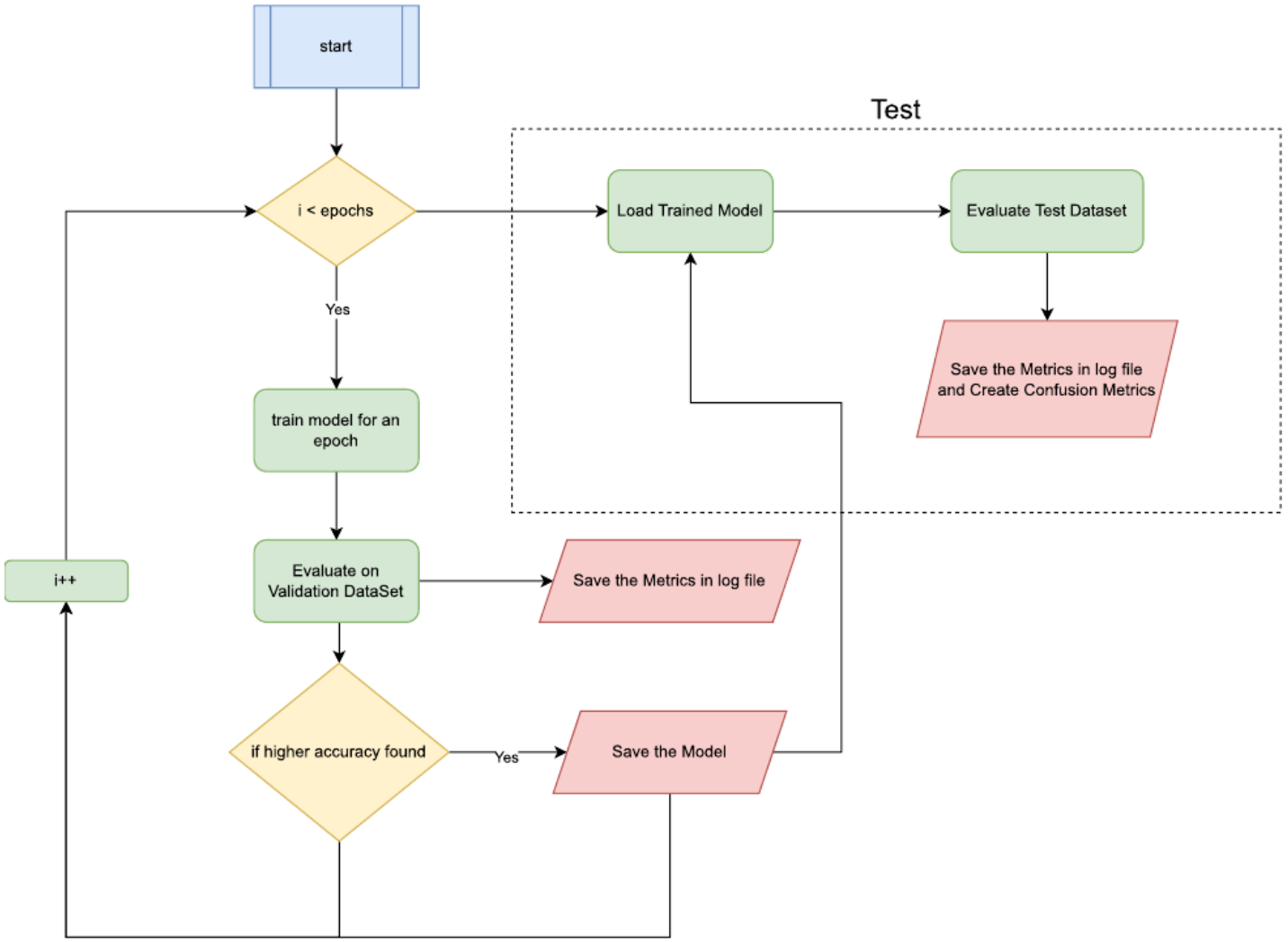


Figure 2.3: Flow-chart of training and evaluation of the models

Coding Method

Class Learner

- Train_one_epoch
- Eval
- Train_eval
- Load
- Save
- Save_log
- Load_logs
- Print metrics
- Plot metrics
- Plot confusion matrixes

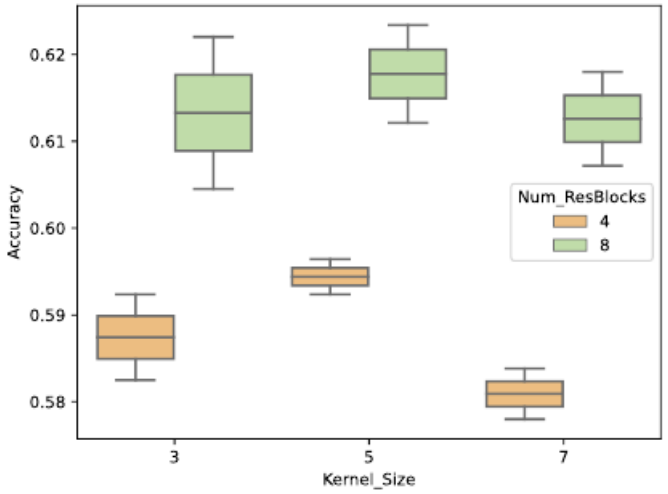
Results

Table 2.2: Summary of Model Architecture and the represented Accuracy

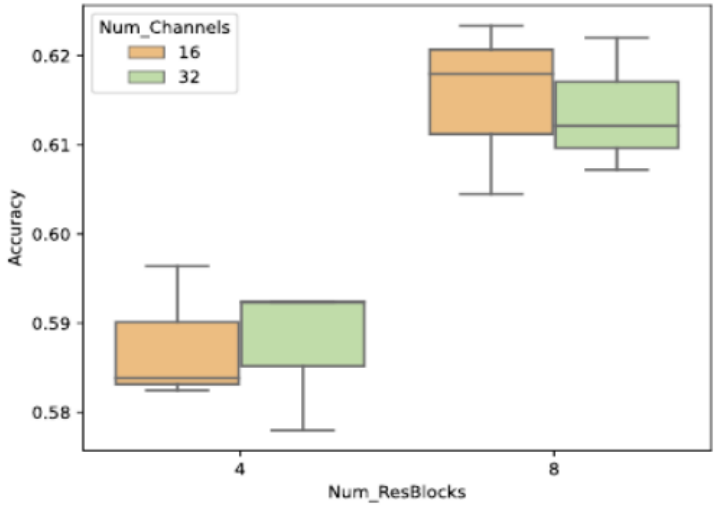
Model Name	K Size	Block	Channels	FCs	Accuracy	Recall	Precision	F1 Score
B16_N4_FC1_K3_AP20	3	4	16	1	0.582472	0.581015	0.599142	0.585643
B32_N4_FC1_K3_AP20	3	4	32	1	0.59236	0.595532	0.622576	0.596638
B16_N8_FC2_K3_AP20	3	8	16	2	0.604494	0.600417	0.620104	0.59736
B32_N8_FC2_K3_AP20	3	8	32	2	0.622022	0.618432	0.650313	0.622679
B16_N4_FC1_K5_AP20	5	4	16	1	0.596404	0.598403	0.612813	0.595999
B32_N4_FC1_K5_AP20	5	4	32	1	0.59236	0.595048	0.629317	0.590056
B16_N8_FC2_K5_AP20	5	8	16	2	0.623371	0.621407	0.654631	0.611707
B32_N8_FC2_K5_AP20	5	8	32	2	0.612135	0.615304	0.652023	0.613317
B16_N4_FC1_K7_AP20	7	4	16	1	0.58382	0.580199	0.619654	0.579667
B32_N4_FC1_K7_AP20	7	4	32	1	0.577978	0.582606	0.592769	0.584087
B16_N8_FC2_K7_AP20	7	8	16	2	0.617978	0.611103	0.627087	0.609395
B32_N8_FC2_K7_AP20	7	8	32	2	0.607191	0.610399	0.651253	0.608782

- As the categories are balanced in all the logs the micro metrics were exactly the same as the accuracy score.

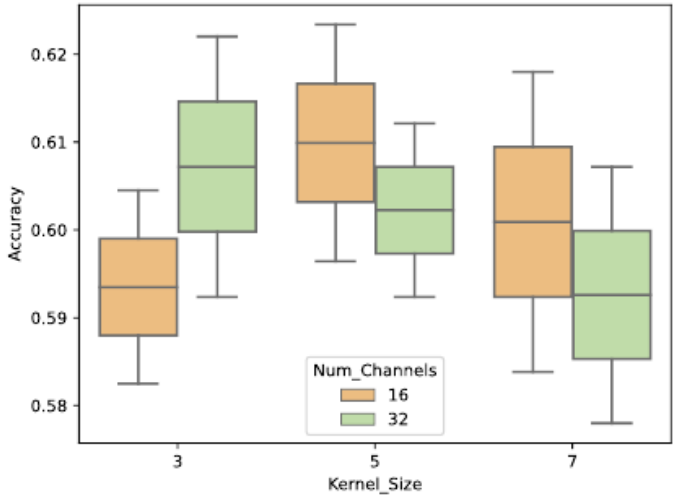
Model Comparison



(d) Accuracy scores compared to Kernel Sizes and number of ResBlocks



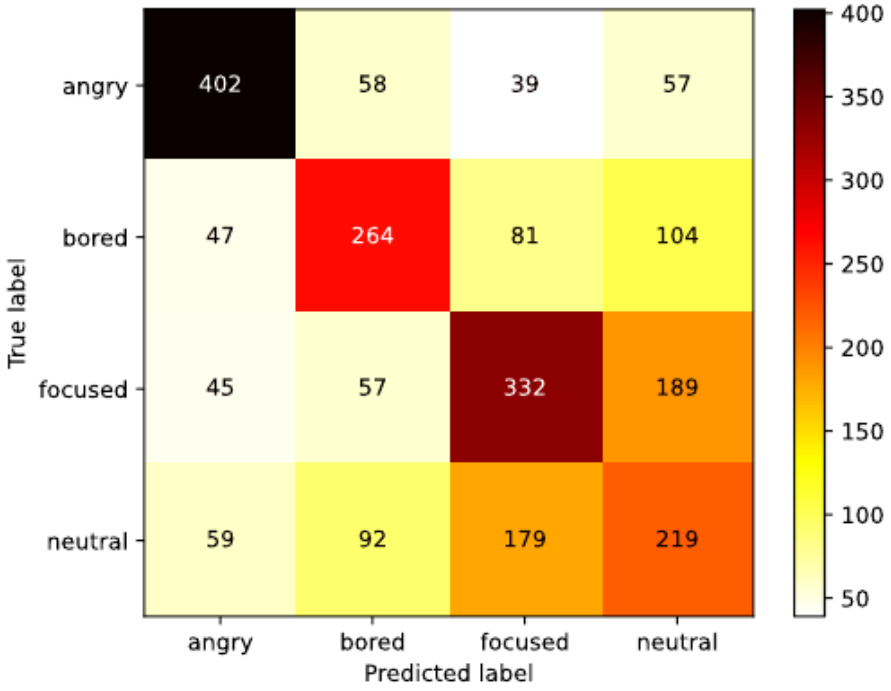
(f) Accuracy scores compared to ResBlock's number of channels and number of ResBlocks



(c) Accuracy scores compared to Kernel Sizes and ResBlock number of Channels

- Models with (5 × 5) kernel sizes outperform (3 × 3) and (7 × 7).
- Positive correlation observed between ResBlocks and accuracy; 8 ResBlocks outperform 4.
- No significant correlation found between channel count in ResBlocks and accuracy.

Confusion Matrix



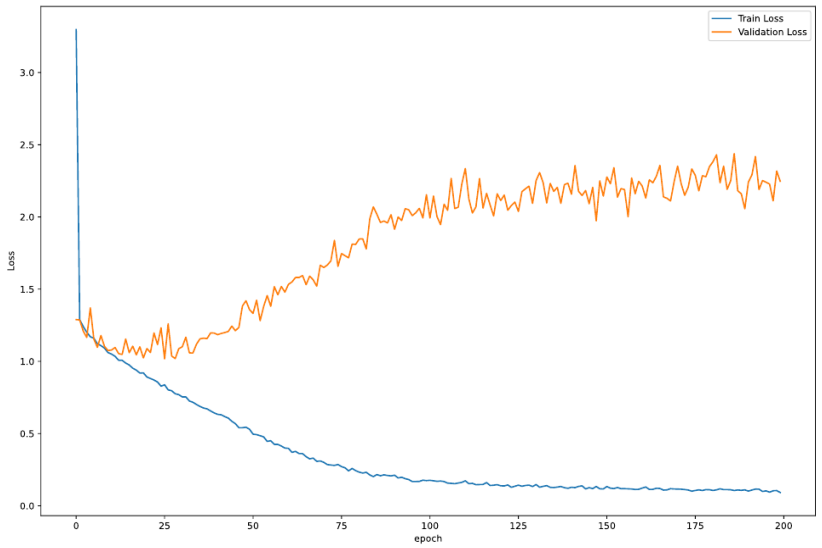
(a) Confusion Matrix on Test

Confusion Matrix shows strong performance in predicting ”Angry” and ”Bored” expressions.

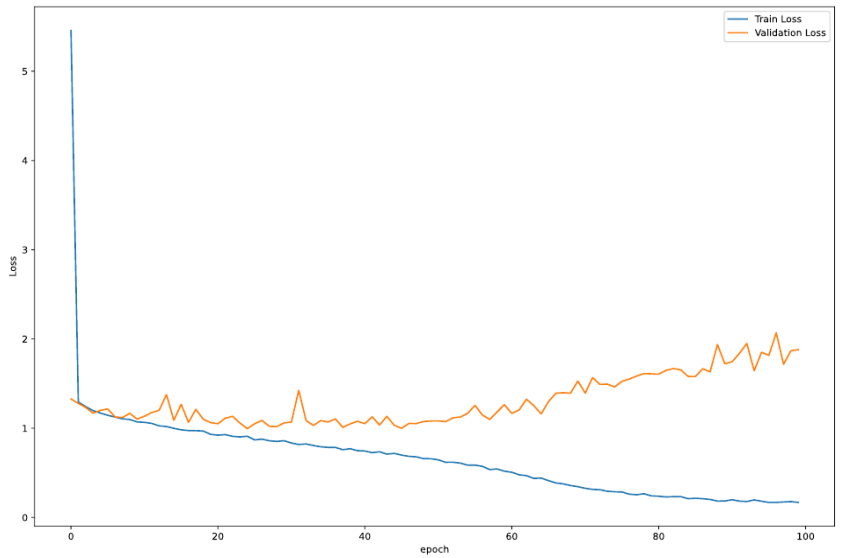
Difficulty distinguishing between ”Focused” and ”Neutral” expressions.

Train and Validation Loss

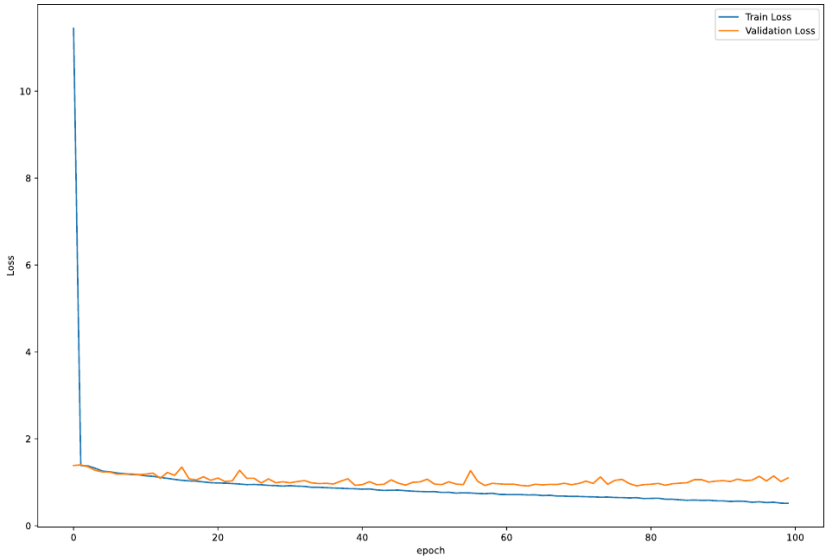
B16 N4 FC1 K3 AP20



B32 N4 FC1 K3 AP20



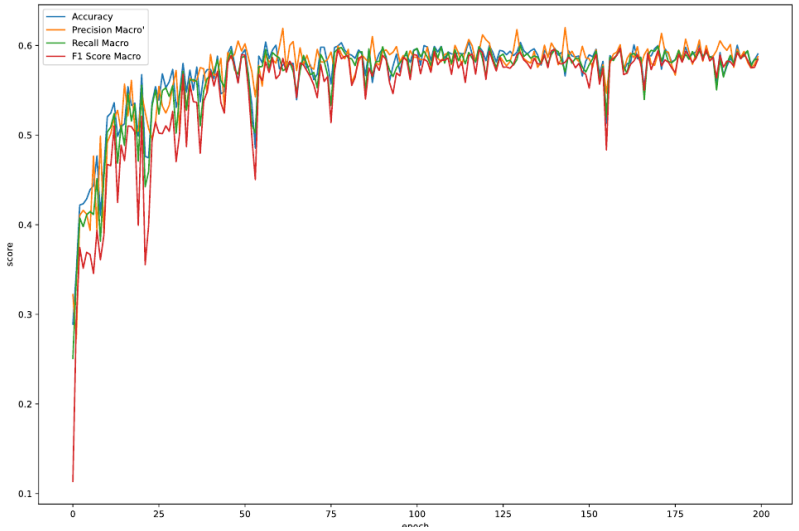
B32 N8 FC2 K3 AP20



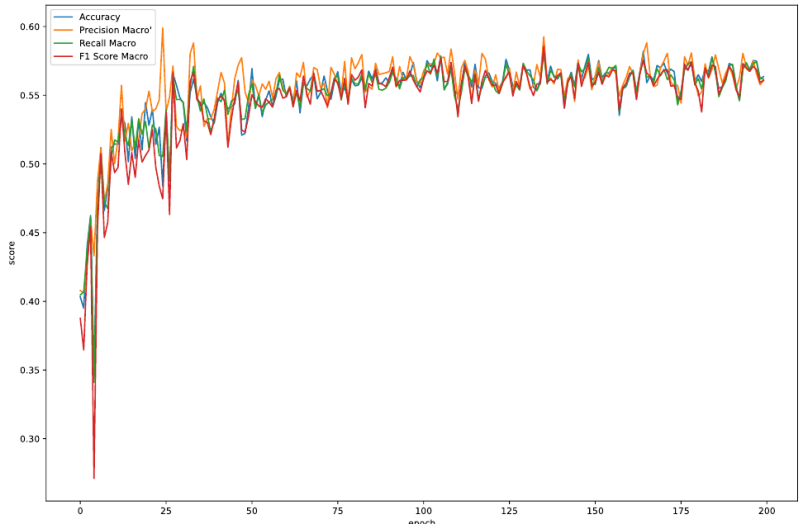
- As Models become Larger, with training progress then tend to fluctuate less

Metrics Plot

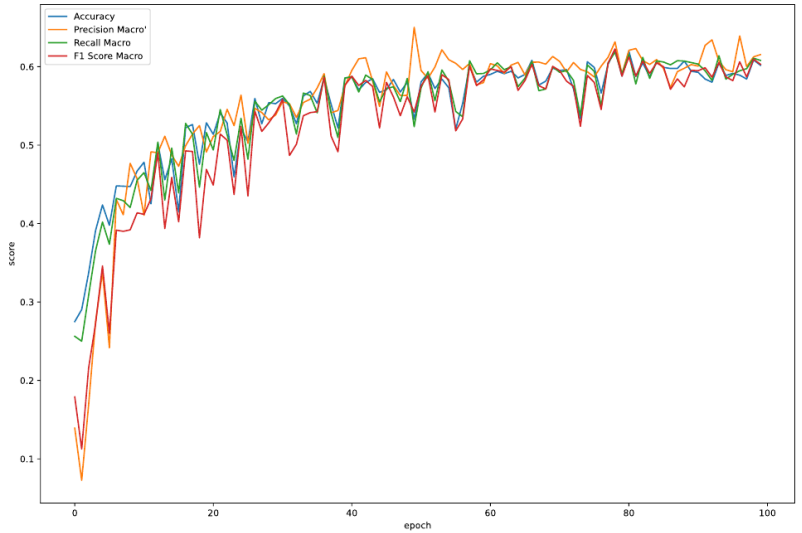
B16 N4 FC1 K3 AP20



B32 N4 FC1 K3 AP20



B32 N8 FC2 K3 AP20



- Smaller Models' metrics rise and reach the threshold faster than bigger models

Summary

Models with (5 × 5) kernel sizes outperform (3 × 3) and (7 × 7).

Positive correlation observed between number of ResBlocks and accuracy; 8 ResBlocks outperform 4.

No significant correlation found between channel count in ResBlocks and accuracy.

Model B16_N8_FC2_K5_AP20 excels with an accuracy of 0.623371.

Confusion Matrix shows strong performance in predicting ”Angry” and ”Bored” expressions.

Difficulty distinguishing between ”Focused” and ”Neutral” expressions.

Loss plots indicate increased overfitting risk with fewer layers; larger models show more stability.

K-Fold

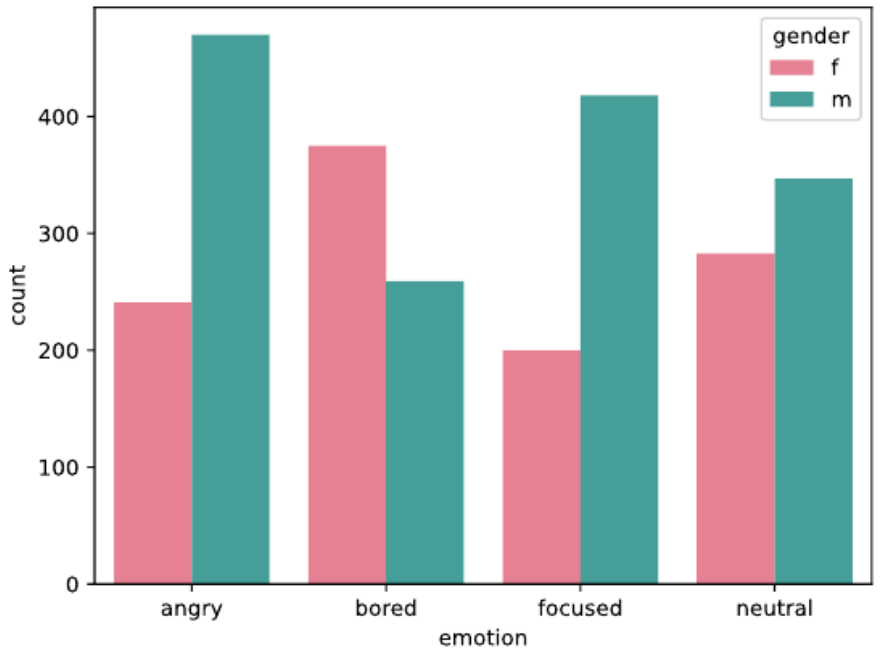
Table 3.2: Table of K-Fold Micro Metrics

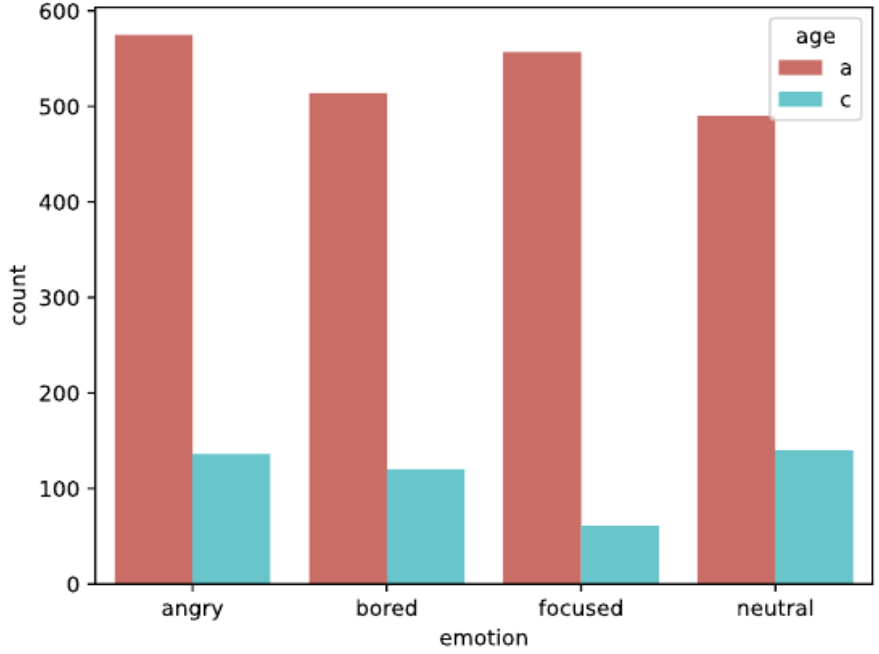
fold	Max Accuracy	Recall	Precision	F1 Score
0	0.648218	0.648218	0.648218	0.648218
1	0.652767	0.652767	0.652767	0.652767
2	0.636088	0.636088	0.636088	0.636088
3	0.645944	0.645944	0.645944	0.645944
4	0.633055	0.633055	0.633055	0.633055
5	0.648218	0.648218	0.648218	0.648218
6	0.625474	0.625474	0.625474	0.625474
7	0.633813	0.633813	0.633813	0.633813
8	**0.661865**	0.661865	0.661865	0.661865
9	0.636846	0.636846	0.636846	0.636846
Average	0.6422288	0.6422288	0.6422288	0.6422288

Table 3.2: Table of K-Fold Micro Metrics

fold	Max Accuracy	Recall	Precision	F1 Score
0	0.648218	0.648218	0.648218	0.648218
1	0.652767	0.652767	0.652767	0.652767
2	0.636088	0.636088	0.636088	0.636088
3	0.645944	0.645944	0.645944	0.645944
4	0.633055	0.633055	0.633055	0.633055
5	0.648218	0.648218	0.648218	0.648218
6	0.625474	0.625474	0.625474	0.625474
7	0.633813	0.633813	0.633813	0.633813
8	**0.661865**	0.661865	0.661865	0.661865
9	0.636846	0.636846	0.636846	0.636846
Average	0.6422288	0.6422288	0.6422288	0.6422288

Bias Analysis





(a) Distribution of images under the classification of gender (b) Distribution of images under the classification of gender

Figure 3.1: Distribution of images among the different classification

Retraining and Mitigating Bias

Table 3.3: Model's metric on different genders and ages

Attribute	Group	Accuracy	precision		recall		F1 score	
			macro	micro	macro	micro	macro	micro
gender	male	0.7651	0.7568	0.7651	0.7449	0.7651	0.7484	0.7651
	female	0.7489	0.7495	0.7489	0.7609	0.7489	0.7456	0.7489
	average	0.7570	0.7531	0.7570	0.7529	0.7570	0.7470	0.7570
age	Adult	0.7687	0.7651	0.7687	0.7607	0.7687	0.7602	0.7687
	Child	0.7090	0.6998	0.7090	0.7313	0.7090	0.6969	0.7090
	average	0.7388	0.7324	0.7388	0.7460	0.7388	0.7286	0.7388

Table 3.4: Retrained Model performance on different genders and ages

Attribute	Group	Accuracy	precision		recall		F1 score	
			macro	micro	macro	micro	macro	micro
gender	male	0.7675	0.7450	0.7675	0.7061	0.7675	0.7307	0.7675
	female	0.7565	0.7481	0.7565	0.7624	0.7565	0.7438	0.7565
	average	0.7620	0.7466	0.7620	0.7342	0.7620	0.7372	0.7620
age	Adult	0.7699	0.7768	0.7699	0.7844	0.7699	0.7476	0.7699
	Child	0.7507	0.7465	0.7507	0.7547	0.7507	0.7401	0.7507
	average	0.7603	0.7617	0.7603	0.7695	0.7603	0.7438	0.7603

Thank You!