Facial Expression Recognition

Final Milestone

Malak Gaballa - 900201683
Masa Tantawy - 900201312

CSCE 4604 - Advanced Machine Learning
Dr. Moustafa Youssef

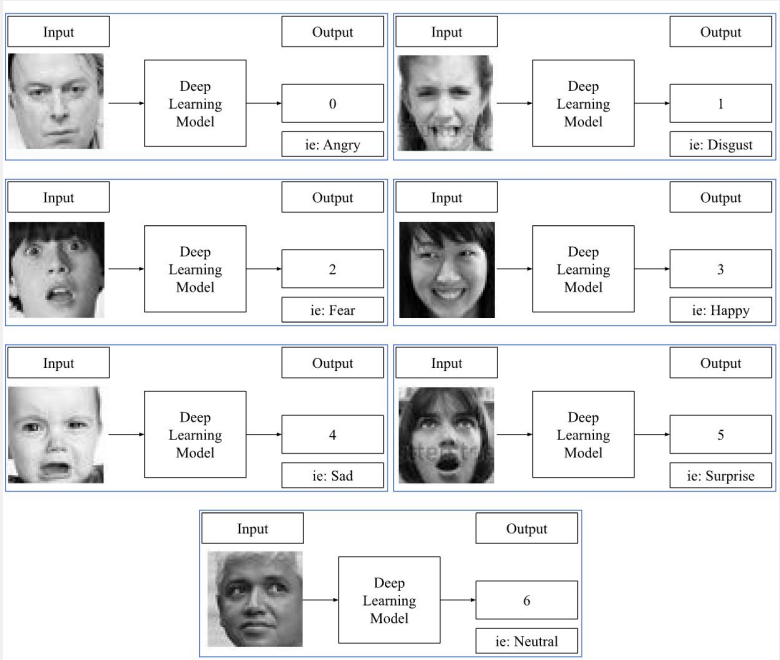
Table of contents

Problem Statement: Facial Expressions Recognition (FER)

Given images of human faces showing different expressions, the model should be able to **categorise each image into one of 7 categories, each representing a facial expression**. These are: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

- *Model input:* image - vector of pixels for a 48x48 pixel grayscale image
- *Model output:* Number from 0 to 6 which indicates the facial expression illustrated in the image

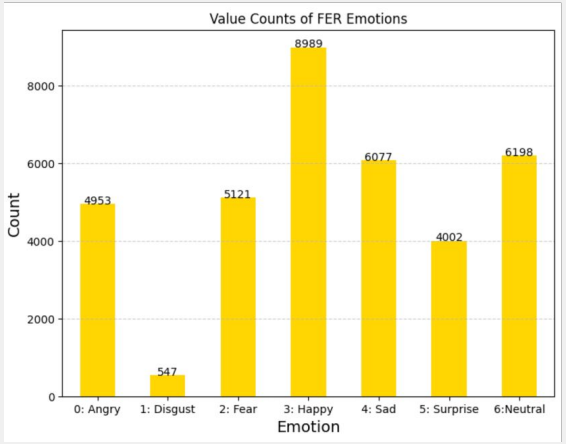
To evaluate the model effectiveness, we opt for the **weighted accuracy metric**, which accounts for class imbalance in the data.



Datasets

Selected Dataset
FER2013

- 35,887 facial grayscale images (48x48 pixels), 63 MBs
- 7 categories (highly imbalanced)
- Has a test-train split.



Other Datasets

Overview of selected most relevant/ suitable datasets

AffectNet

- 12,809 images, 5 GBs
- 8 categories
- Slightly balanced

ExpW

Expression in-the-Wild Dataset

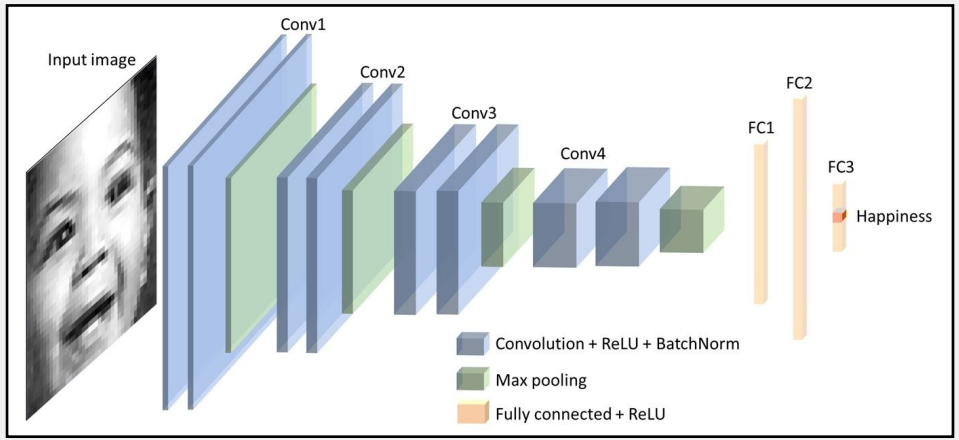
- 91,793 images, 8 GBs
- 7 categories

Baseline Model - VGGNet

The VGGNet model, short for Visual Geometry Group Network

- *Training:* FER2013 dataset achieving an accuracy of 73.28%
- *Research Paper:* Facial Emotion Recognition: State of the Art Performance on FER2013
- *Repository:* Github link
- *Frameworks:* PyTorch

A classical CNN consisting of 4 convolutional stages and 3 fully connected layers.



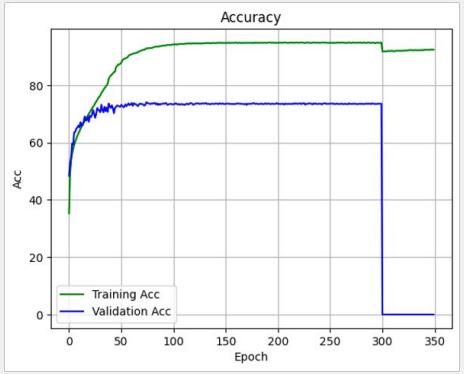
- ***Each convolutional stage:*** 2 convolutional blocks & a max-pooling layer.
- ***Convolution block:*** consists of a convolutional layer, a ReLU activation, and a batch normalization layer.
- The first 3 fully connected layers are followed by a ReLU activation. The 3rd fully connected layer is for classification.

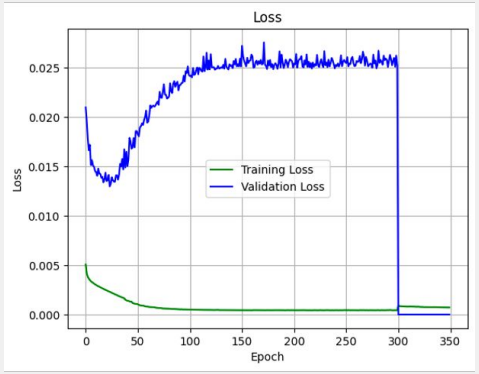
Baseline Model - Performance

PyTorch

Original model; hyperparameters cannot be modified.

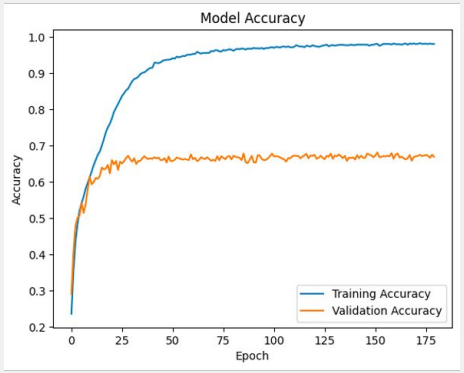
- Epochs = 350
  - Top-1 Accuracy : 73.27%
  - Top-2 Accuracy : 86.45%

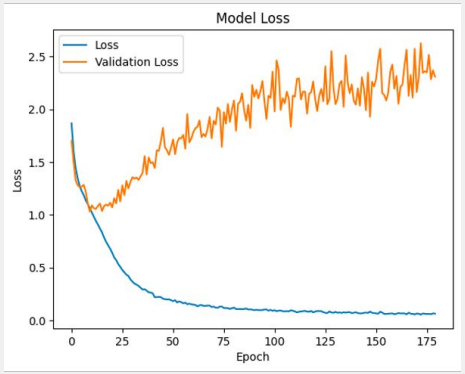




Keras TensorFlow

- Epochs = 180 instead of 350 due to GPU limit
  - Top-1 Accuracy : 65.76%
  - Top-2 Accuracy : 79.91%
  - Top-3 Accuracy : 88.49%





02

Methodology & Results

Methodology

Hyperparameters Tuning

Different regularizers & optimizers with varying learning rates were experimented with in addition to early stopping.

Final modifications:

- **ADAM LEARNING RATE = 0.0001** instead of 0.001
- **EARLY STOPPING** with patience = 10

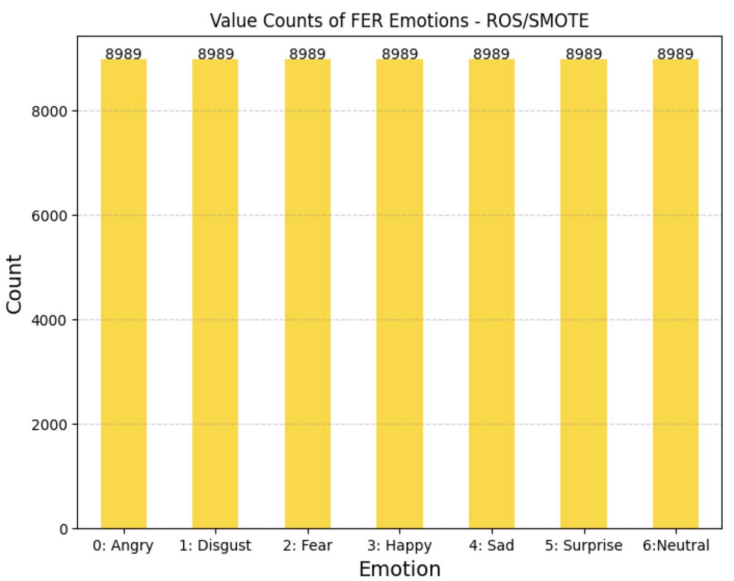
Data Imbalance Handling

- Oversampling: **ROS** & **SMOTE**
- ~~Undersampling:~~ RUS & Tomeklinks (deep models require large datasets)
- **SMOTE + TOMEK** & ~~Smote + ENN~~ (reversed the imbalance)

→ Before being added to the model, each dataset was split using sklearn in the same ratio as the original data (80% train ,10% test ,10% validation)

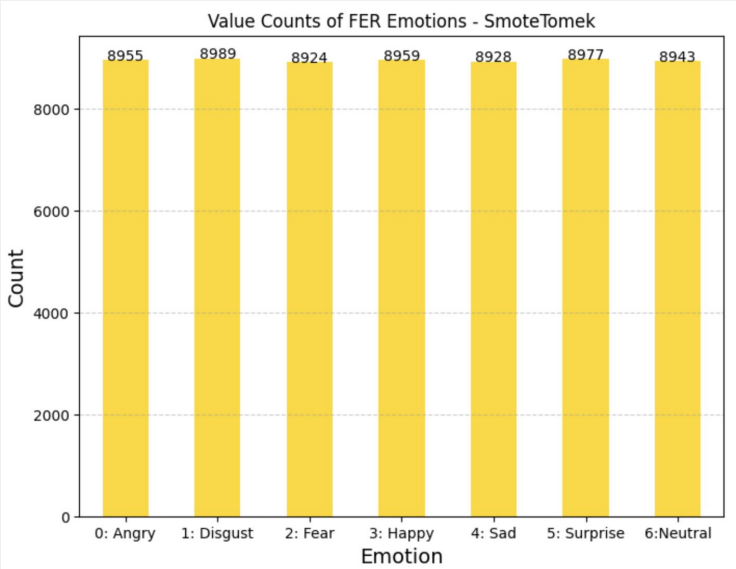
Data imbalance

Balanced Data



ROS and SMOTE
A total of 62,923 images

Balanced Data



SmoteTomek
A total of 62,675 images

Methodology

Data Augmentation

A random balanced subset of the dataset has undergone different combinations (none, one, or multiple) of **HORIZONTAL FLIPPING, ROTATION, GAUSSIAN NOISE ADDITION** with different ratios from given set ranges.

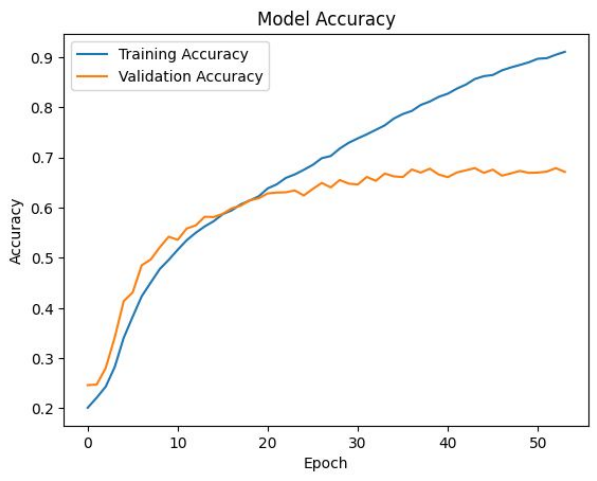
Auxiliary Data

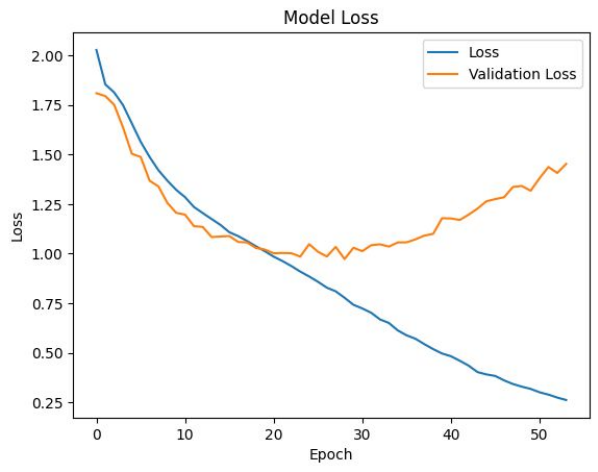
The **AFFECTNET DATASET** was used. It originally contained 8 categories of 96x96 coloured images.

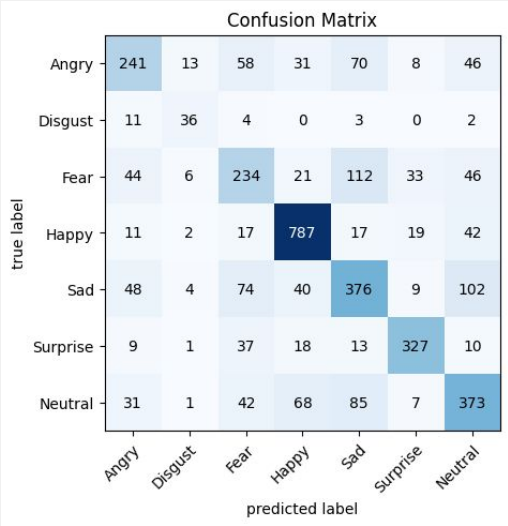
→ Before being added to the model, only the common 7 categories were selected, images were converted to grayscale and resized to 48x48.

Results - Hyperparameters Tuning

- Training process stopped after 54 epochs
  - Top-1 Accuracy: 66.15%
  - Top-2 Accuracy: 82.22%
  - Top-3 Accuracy: 90.89%







Note: All extra training to follow was done on the hyper tuned model.

Results - Extra training on balanced data

Random Oversampling (ROS)

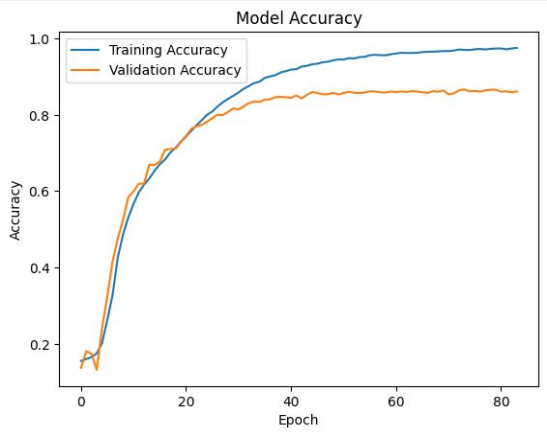
- Top-1 Accuracy: 85.97%
- Top-2 Accuracy: 92.71%
- Top-3 Accuracy: 96.49%

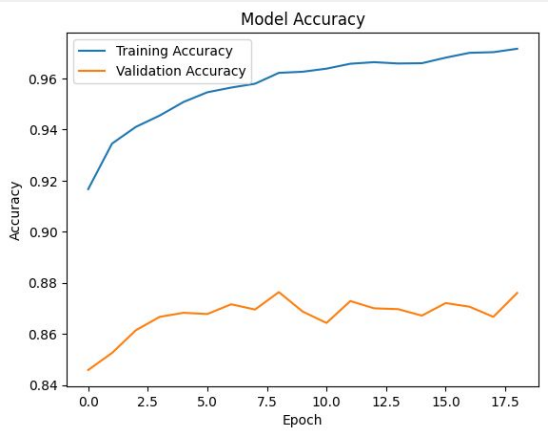
SMOTE

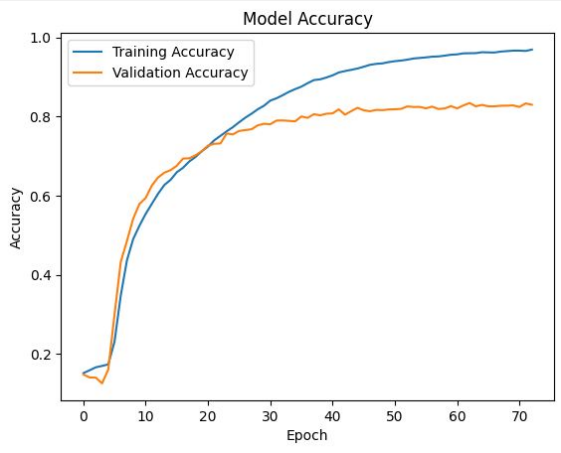
- Top-1 Accuracy: 87.18%
- Top-2 Accuracy: 93.88%
- Top-3 Accuracy: 96.84%

SmoteTomek

- Top-1 Accuracy: 82.88%
- Top-2 Accuracy: 91.82%
- Top-3 Accuracy: 95.64%







Results - Extra training on augmented data

Random Oversampling (ROS)

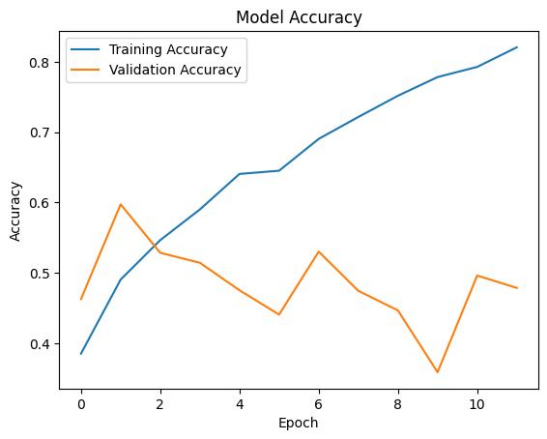
- Top-1 Accuracy: 59.74%
- Top-2 Accuracy: 79.91%
- Top-3 Accuracy: 88.63%

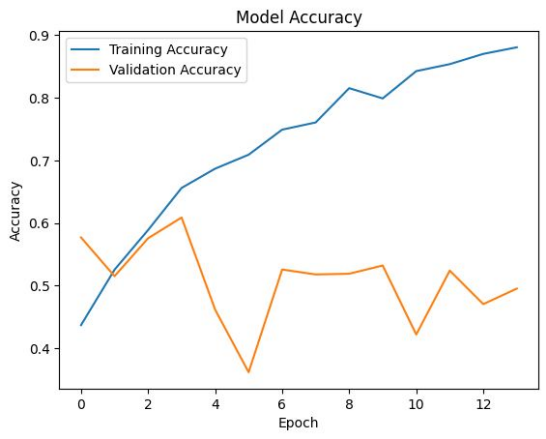
SMOTE

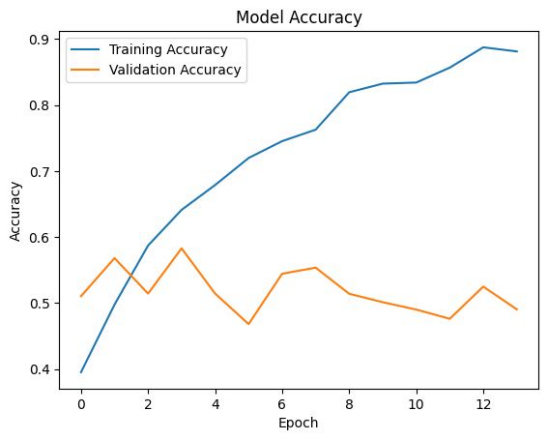
- Top-1 Accuracy: 60.88%
- Top-2 Accuracy: 78.99%
- Top-3 Accuracy: 89.22%

SmoteTomek

- Top-1 Accuracy: 58.29%
- Top-2 Accuracy: 77.46%
- Top-3 Accuracy: 86.77%







Results - Extra training on auxiliary data

Random Oversampling (ROS)

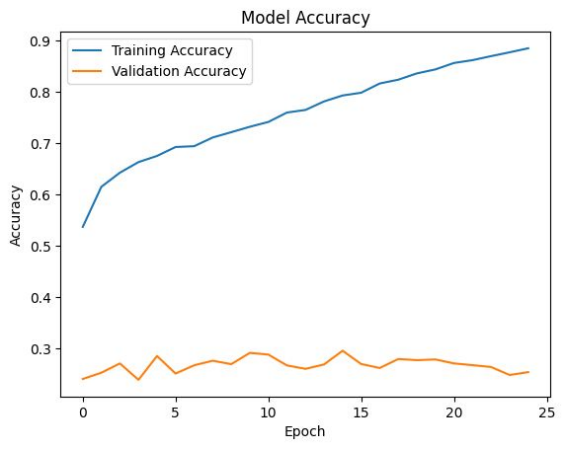
- Top-1 Accuracy: 29.51%
- Top-2 Accuracy: 45.33%
- Top-3 Accuracy: 53.41%

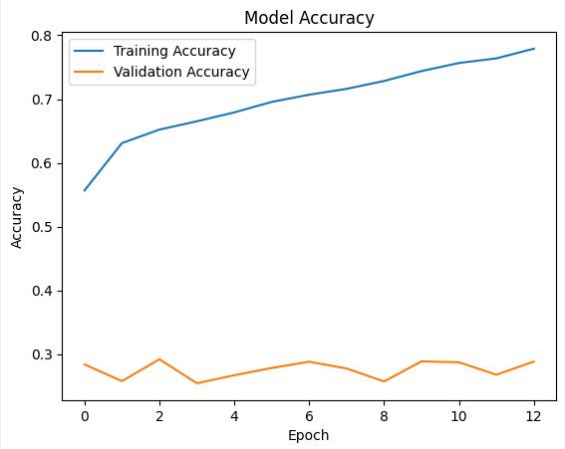
SMOTE

- Top-1 Accuracy: 30.31%
- Top-2 Accuracy: 44.58%
- Top-3 Accuracy: 54.25%

SmoteTomek

- Top-1 Accuracy: 29.17%
- Top-2 Accuracy: 42.71%
- Top-3 Accuracy: 51.96%





Note: The model zoo for each model has been saved for future use.

03

Discussion & Conclusion

Discussion

Ensemble

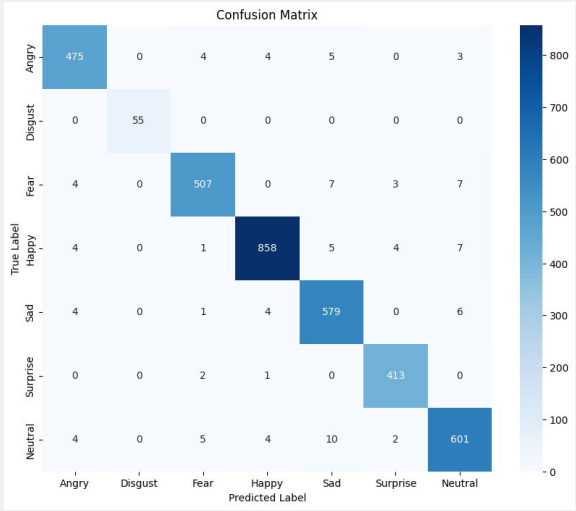
Final Output = average output of 3 distinct VGGNet models → ROS, SMOTE, SmoteTomek

(augmented and auxiliary data excluded as they deteriorated the model)

- Top-1 Accuracy: 97.18%
- Top-2 Accuracy : 99.58%
- Top-3 Accuracy : 99.72%

Baseline Model

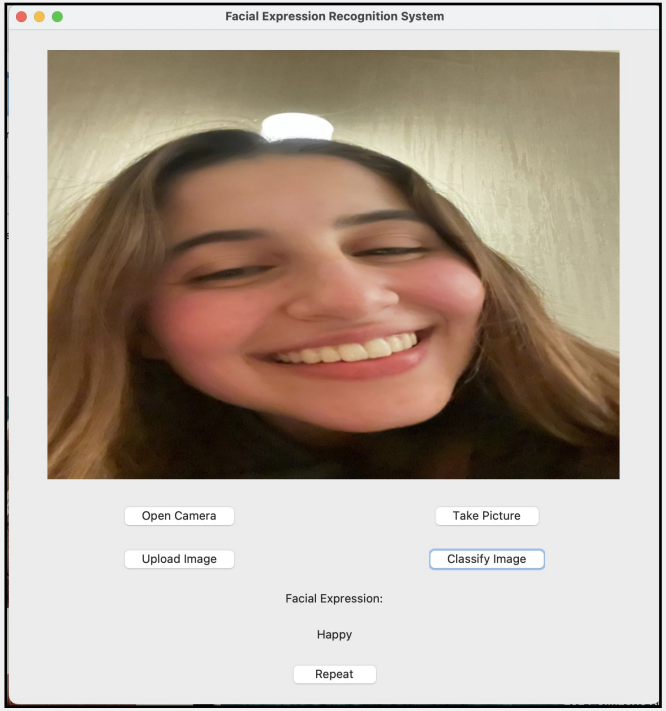
- Top-1 Accuracy : 65.76%
- Top-2 Accuracy : 79.91%
- Top-3 Accuracy : 88.49%

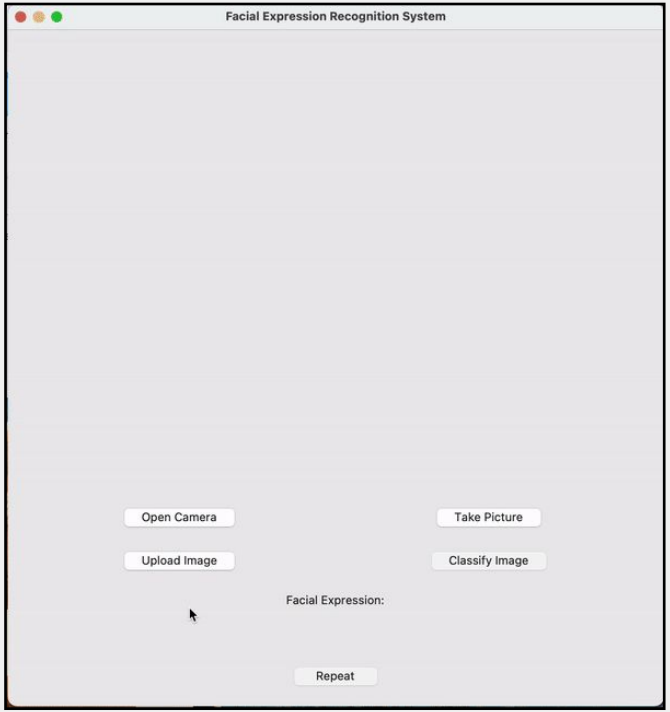


Discussion

Real Time App

Model demo





Conclusion

Lessons Learnt

- Data imbalance handling significantly enhances the performance of the model.
- Constructing an ensemble model using multiple VGGNet models trained on balanced datasets further optimized performance.
- It is also concluded that extra training on auxiliary or augmented data may lead to worse performance of the model instead of enhancing it.
- The confusion matrix highlights *Angry, Fear,* and *Neutral* as the most challenging expressions to classify. This difficulty may arise from subtle facial differences or dataset imbalances.

Future recommendations

- It is suggested to train the model on a more diverse and generalized database of facial expressions, such as Exp-W which posed a challenge due to GPU limitations and the dataset size.
- Considering the inclusion of an 8th category of facial expression, such as *contempt* as seen in the AffectNet Dataset, could enhance model comprehensiveness.
- Lastly, transitioning from grayscale to RGB images for input might yield better results.

Thanks!

Facial Expression Recognition
Final Milestone

Malak Gaballa - 900201683
Masa Tantawy - 900201312

Malak

- Model Ensemble
- Website

Masa

- App
- Poster