Facial Expression Recognition

Milestone 2

Malak Gaballa - 900201683
Masa Tantawy - 900201312

CSCE 4604 - Advanced Machine Learning
Dr. Moustafa Youssef

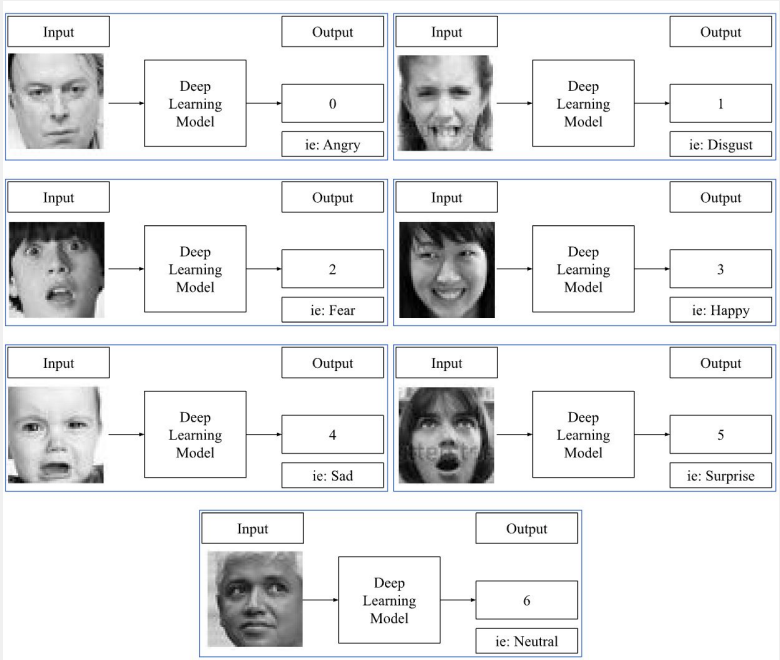
Table of contents

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: A 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.



Related Work (SOTA)

Model	Datasets	Accuracy (%)
VGG,Res-Net, and Inception	FER2013	75.2
LHC-Net	FER2013	74.42
VGGNet	FER2013	73.28
CNN Hyperparameter Optimization	FER2013	72.16
Ad-Corre	FER2013, AffectNet , RAF-DB	72.03
DeepEmotion	FER2013, CK+, FERG, JAFFE	70.02

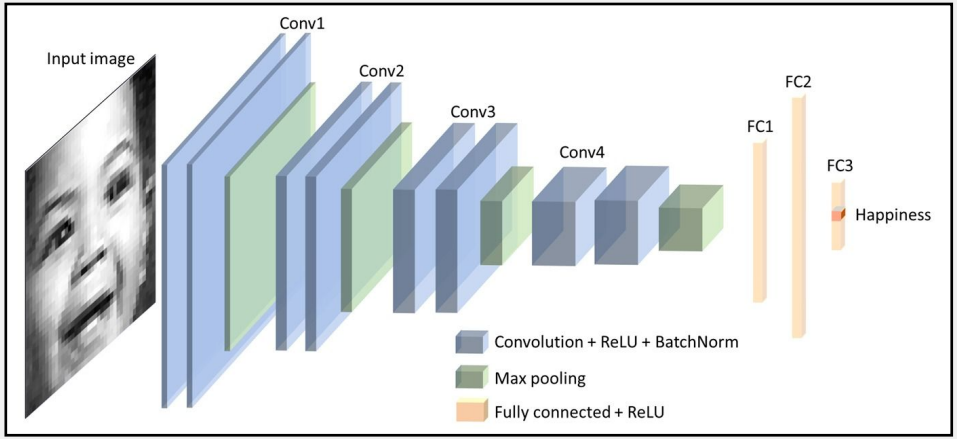
The VGGNet model, short for Visual Geometry Group Network

- *Research Paper:* Facial Emotion Recognition: State of the Art Performance on FER2013
- *Repository:* Github link
- *Frameworks:* PyTorch

Baseline Model – VGGNet

A classical convolutional neural network architecture used in large-scale image processing & pattern recognition.

The network consists of 4 convolutional stages and 3 fully connected layers. The convolutional stages are responsible for feature extraction, dimension reduction, and non-linearity. The fully connected layers are trained to classify the inputs as described by extracted features.

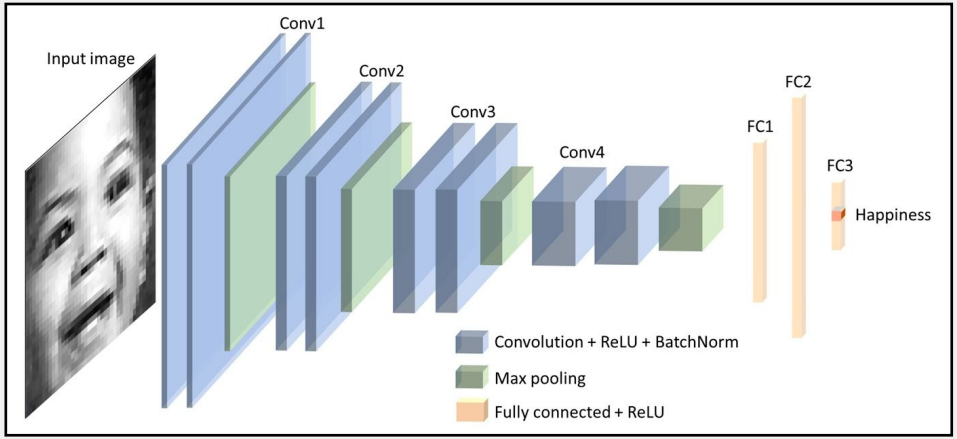


- ***Each convolutional stage:*** 2 convolutional blocks & a max-pooling layer.
- ***Convolution block:*** consists of a convolutional layer, a ReLU activation, and a batch normalization layer. Batch normalization is used to speed up the learning process, reduce the internal covariance shift, and prevent gradient vanishing or explosion.

Baseline Model – VGGNet

A classical convolutional neural network architecture used in large-scale image processing & pattern recognition.

The network consists of 4 convolutional stages and 3 fully connected layers. The convolutional stages are responsible for feature extraction, dimension reduction, and non-linearity. The fully connected layers are trained to classify the inputs as described by extracted features.



- The first two fully connected layers are followed by a ReLU activation. The third fully connected layer is for classification.

Proposed Updates

Hyperparameters Tuning

- Number of epochs,
- Regularisation, ...

Data Imbalance Handling

- Oversampling: ROS & SMOTE
- ~~Undersampling:~~ RUS & Tomeklinks
- Smote + Tomek & Smote + ENN

Model Ensemble

To enhance model performance and interpretability

Data Augmentation

- Adding auxiliary datasets to train the model
- Image manipulation: such as mirroring/reflecting them, adding background noise, or other appropriate approaches.

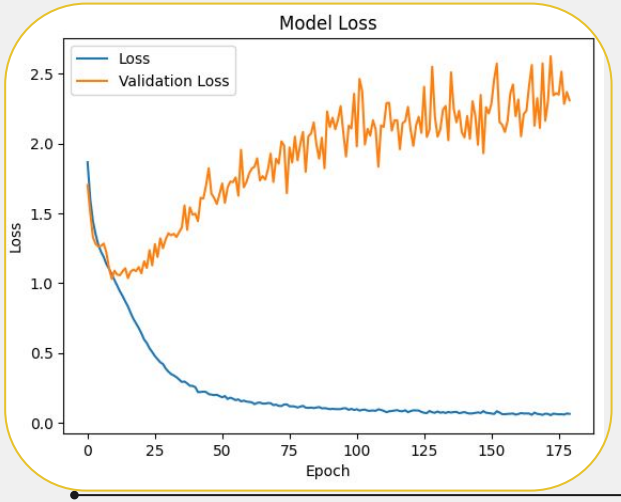
Real-time App (TENTATIVE)

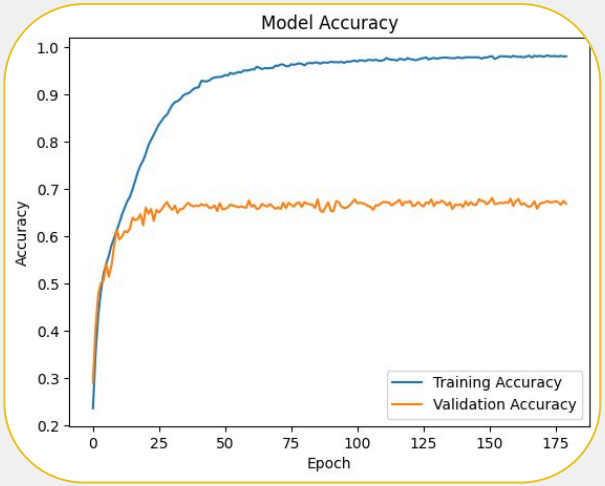
To afford high generalisability on the Egyptian/Arab race

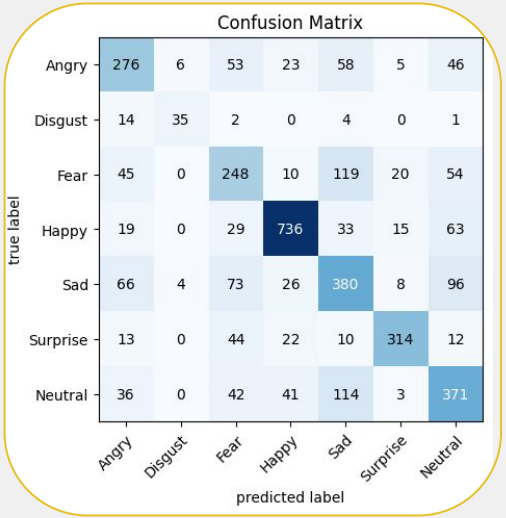
The Model

This is a trained model in **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

Progress Report

Progress on the proposed solutions during this milestone

Milestone 2 Progress

Hyperparameters Tuning

Using different regularizers and optimizers

Data Imbalance Handling

Training the model on the different variations of the data (milestone 1)

Image Manipulation

Horizontal Flipping, Rotation and Noise addition to the images

Auxiliary Data

Extra model training on AffectNet dataset

1. Hyperparameters Tuning

Optimizer

Regularizers and Optimizers with varying learning rates

- *L1 regularization:* neither improved performance nor reduced overfitting.
- *SGD optimizer:* resulted in poor performance.
- Experimented with different learning rates for ADAM optimizer.
  - **lr=0.0001** resulted in highest performance.

Early Stopping

- Utilized due to model's early saturation and lack of improvement in validation performance
- Monitors validation accuracy
  - ***Patience = 10***
- training process will stop if the validation accuracy does not improve for 10 consecutive epochs

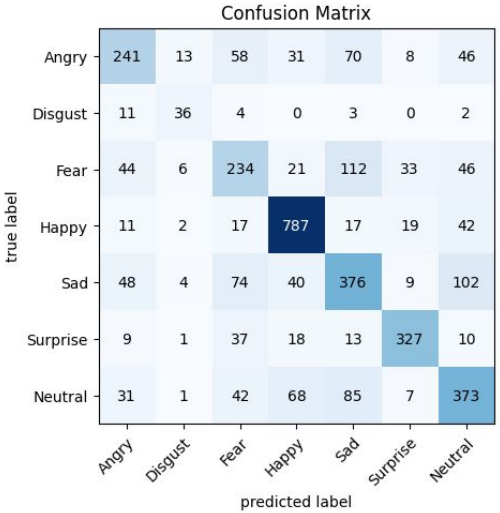
Final Modifications

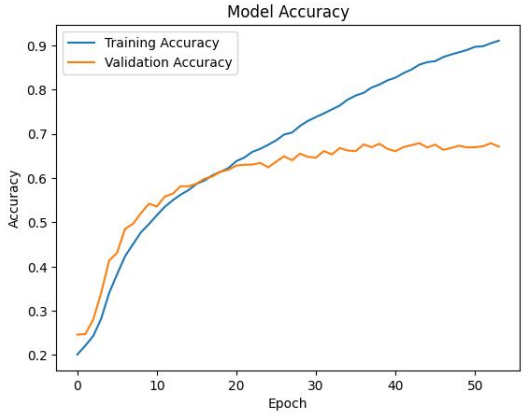
- Adam Learning Rate = 0.0001 instead of 0.001
- Early Stopping with patience = 10

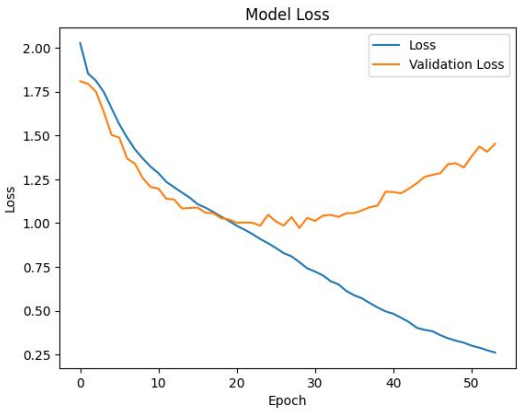
1. Hyperparameters Tuning

Improved Results

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







Final Modifications

- Adam Learning Rate = 0.0001 instead of 0.001
- Early Stopping with patience = 10

2. Data Imbalance Handling

Original Data
Highly imbalanced
A total of 35,887 images

ROS/ SMOTE
Balanced
A total of 62,923 images

SmoteTomek
Almost Balanced
A total of 62,675 images

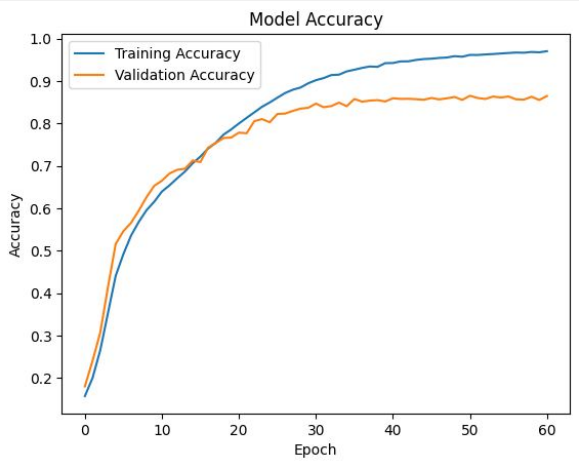
Smoteen
Highly Imbalanced
A total of 32,809 images

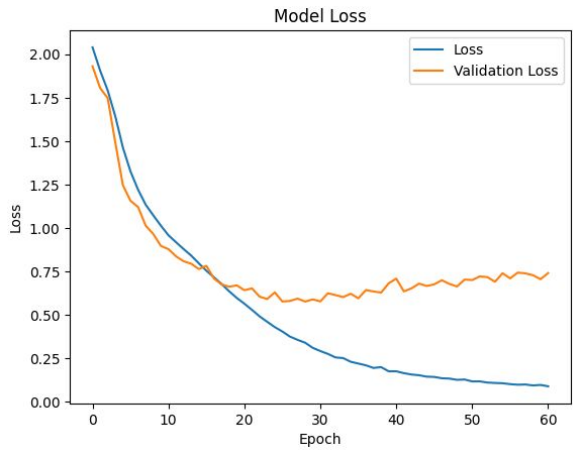
Before Modelling

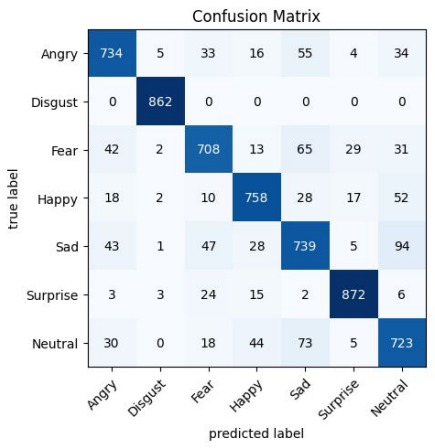
- Split using sklearn (80% train ,10% test ,10% validation)
- Ran model with tuned hyperparameters on each balanced dataset
- 3 Models

ROS Model

- Epochs = 61
  - Top-1 Accuracy: 85.75%
  - Top-2 Accuracy: 92.74%
  - Top-3 Accuracy: 96.55%

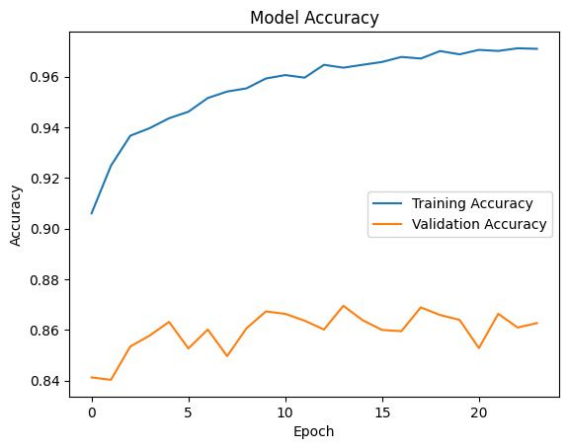


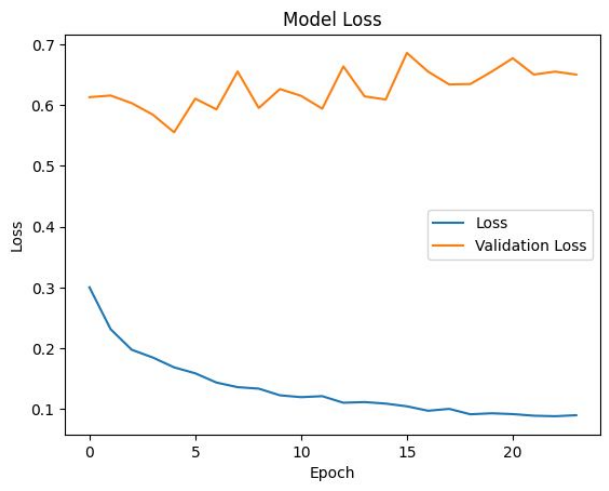


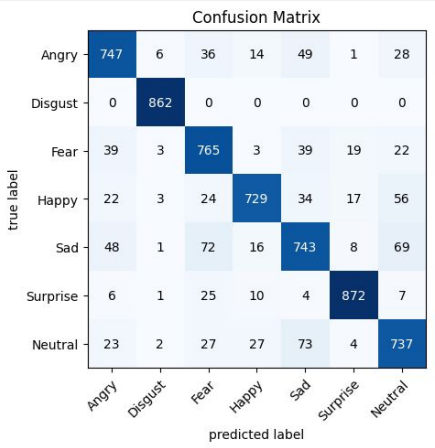


SMOTE Model

- Epochs = 24
  - Top-1 Accuracy: 86.68%
  - Top-2 Accuracy: 93.55%
  - Top-3 Accuracy: 96.68%

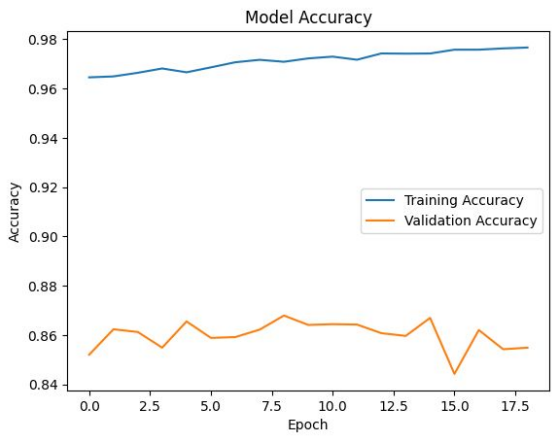


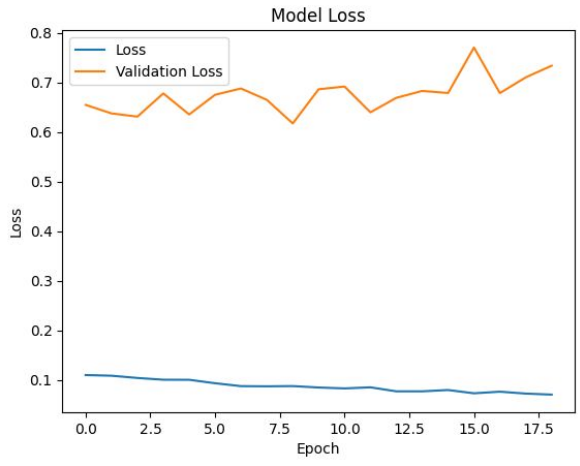


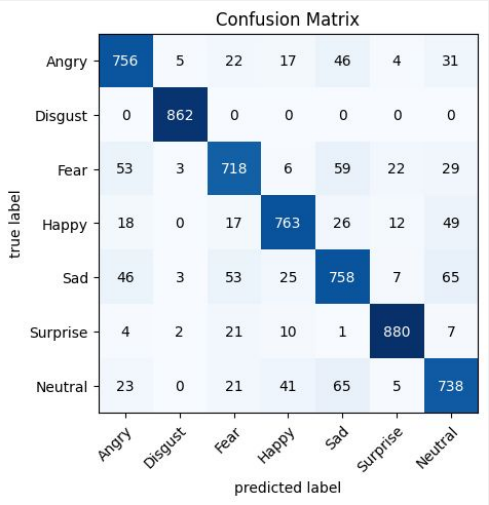


SmoteTomek Model

- Epochs = 19
  - Top-1 Accuracy: 87.00%
  - Top-2 Accuracy: 94.23%
  - Top-3 Accuracy: 97.19%





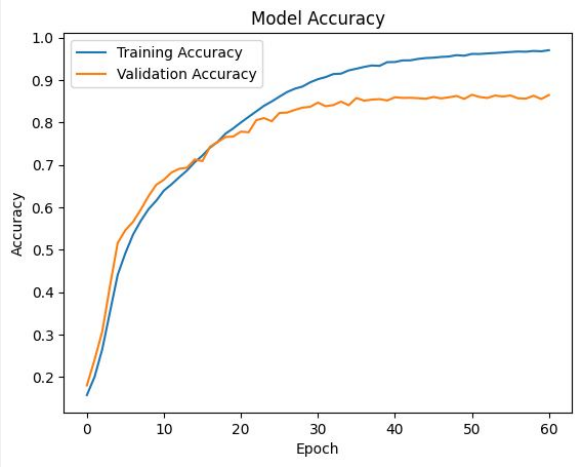


Comparative Analysis

Random Oversampling (ROS)

61 Trained Epochs

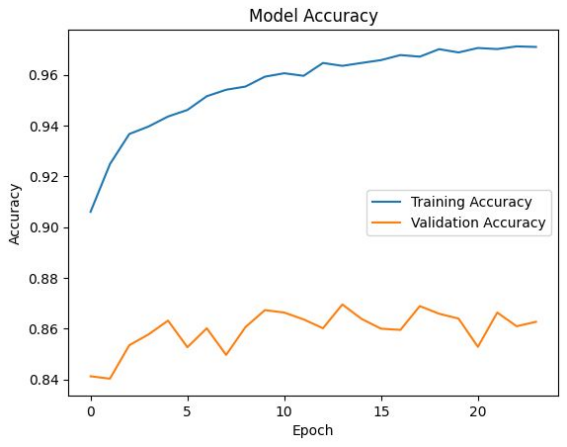
- Top-1 Accuracy: 85.75%
- Top-2 Accuracy: 92.74%
- Top-3 Accuracy: 96.55%



SMOTE

24 Trained Epochs

- Top-1 Accuracy: 86.68%
- Top-2 Accuracy: 93.55%
- Top-3 Accuracy: 96.68%



SmoteTomek

19 Trained Epochs

- Top-1 Accuracy: 87.00%
- Top-2 Accuracy: 94.23%
- Top-3 Accuracy: 97.19%

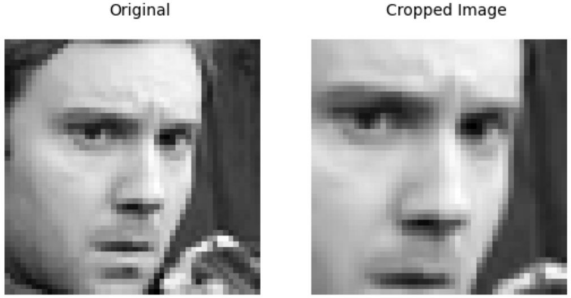
Note: The model zoo for each model has been saved for future use.

3. Image Manipulation

Vertical & Horizontal Flipping

Image Rotation

Image Cropping



Gaussian Noise Addition

3. Image Manipulation

- Adopted suitable approaches to FER problem: only 3 techniques
- Randomization used to select technique(s) and amount of change in each image

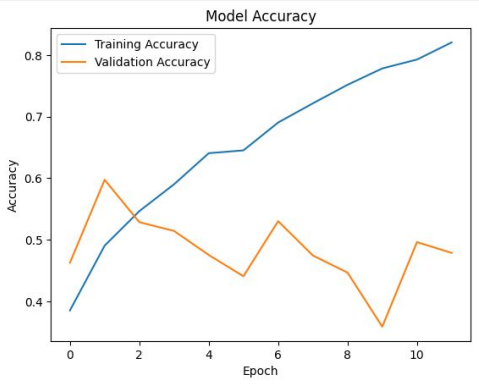
Select a random balanced sample from original data **(BASED ON MINIMUM CLASS- HAPPY = 436)**

Horizontal Flipping	Image Rotation	Gaussian Noise Addition
Random choice - T / F	Random choice - T / F	Random choice - T / F
Random proportion of the image from 50% to 100%	Random angle between -100° & 100°	Random mean & standard deviation between 1 & 10

3. Image Manipulation

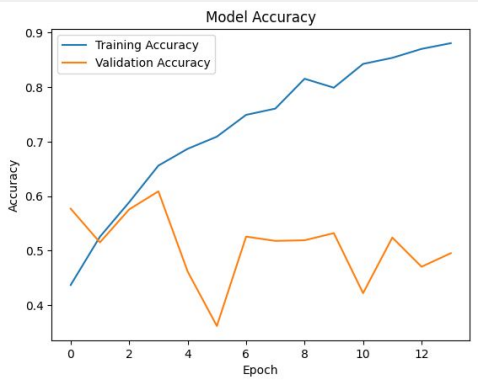
Random Oversampling

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



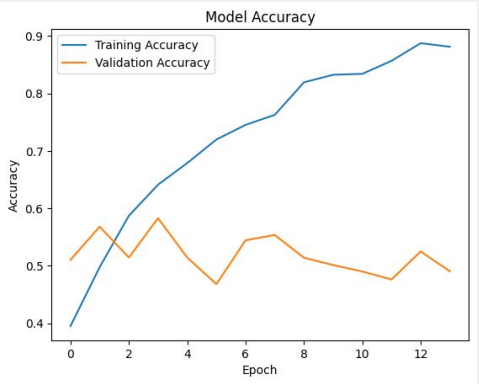
SMOTE

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



SmoteTomek

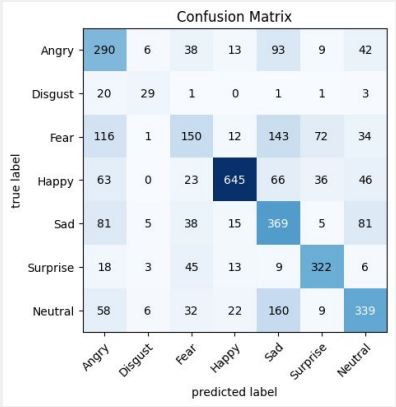
Top-1 Accuracy: 58.29%
Top-2 Accuracy: 77.46%
Top-3 Accuracy: 86.77%



3. Image Manipulation

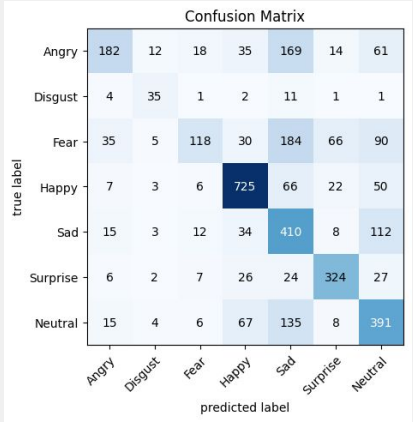
Random Oversampling

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



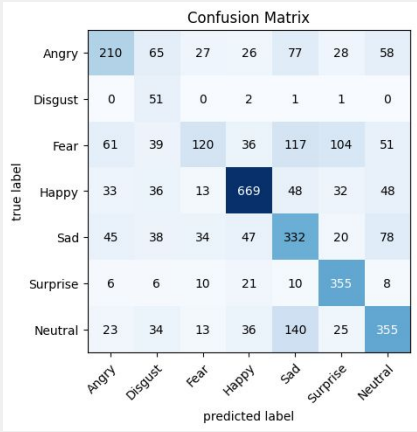
SMOTE

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%



4. Auxiliary Data

AffectNet Data

The largest database of facial expression

Advantage

- Manually Annotated → Higher real-life accuracy

Disadvantage

- 8 categories not 7

1. Select only common 7 categories (read as pixels)
2. Convert from RGB to Grayscale
3. Resize from 96x96 to 48x48

Exp-W

Expression in-the-Wild (ExpW) Dataset

Advantage

- 7 categories
- Accessible
- Manually Annotated → Higher real-life accuracy

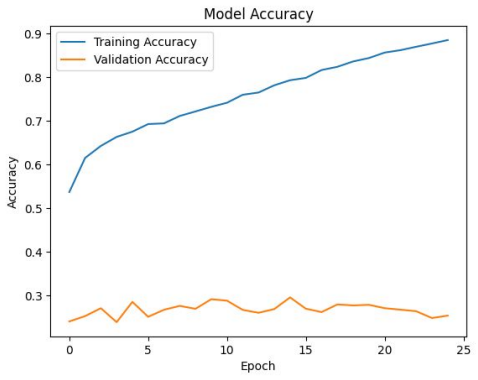
Disadvantage

- Over 91,000 instances → large size (given GPU constraints)

4. Auxiliary Data

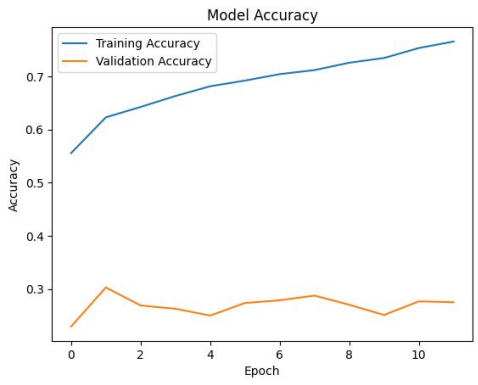
Random Oversampling

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



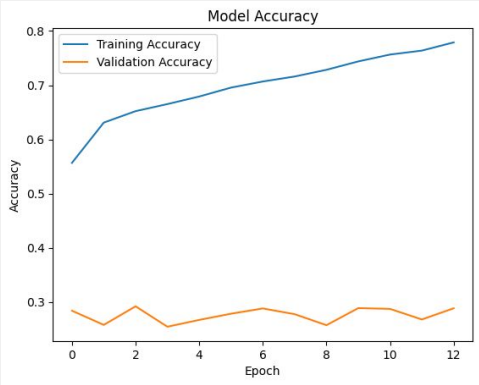
SMOTE

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



SmoteTomek

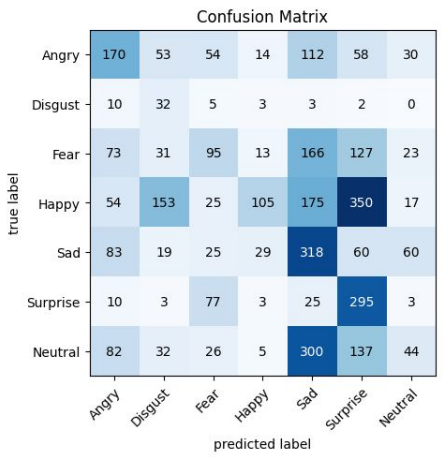
Top-1 Accuracy: 29.17%
Top-2 Accuracy: 42.71%
Top-3 Accuracy: 51.96%



4. Auxiliary Data

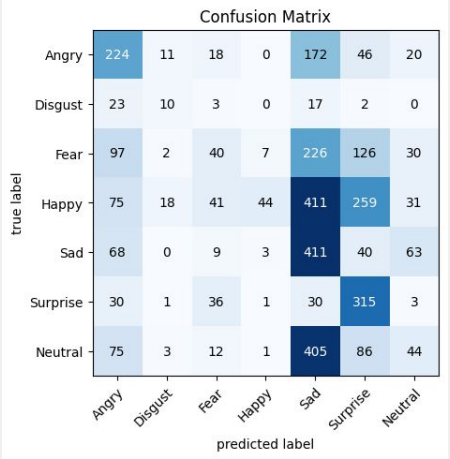
Random Oversampling

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



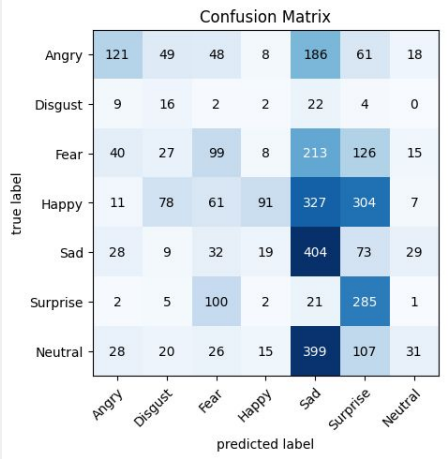
SMOTE

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

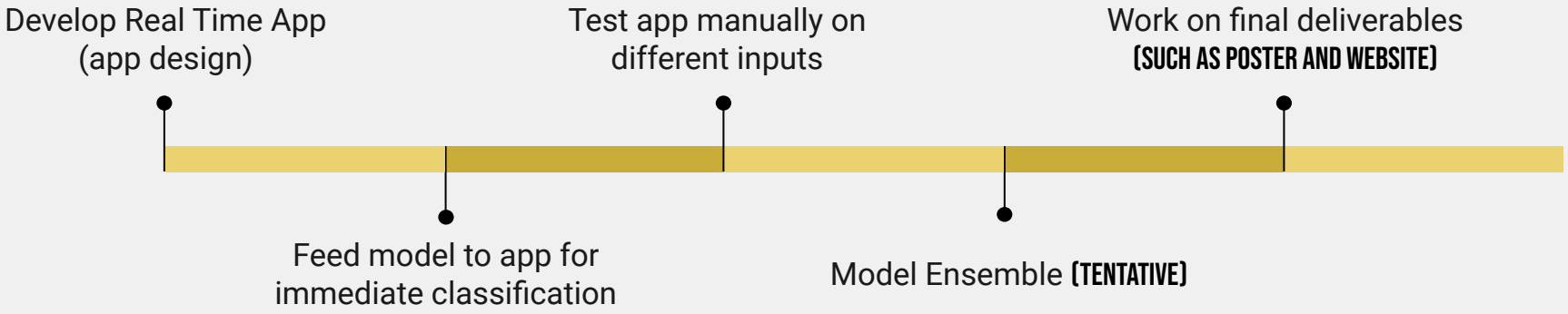


SmoteTomek

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



Timeline & Member Contribution



Note: The timeline is tentative and subject to adjustments depending on progress / time availability.

Malak

- Hyperparameters Tuning
- Model Training (on data resulting from any handling approaches)

Masa

- Image Manipulation Techniques
- Auxiliary Datasets

Thanks!

Facial Expression Recognition
Milestone 2

Malak Gaballa - 900201683
Masa Tantawy - 900201312