



Facial Expression Recognition

Milestone 2

Malak Gaballa - 900201683

Masa Tantawy - 900201312

CSCE 4604 - Advanced Machine Learning
Dr. Moustafa Youssef

Table of contents



01

**Proposal
Summary**



02

**Progress
Report**



03

**Timeline &
Member
Contribution**



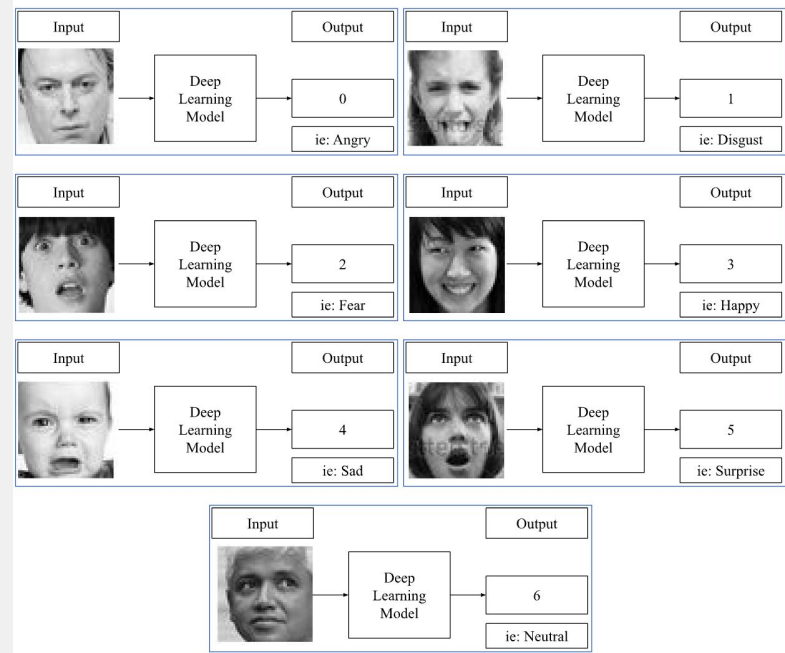
Problem Statement: Facial expressions recognition (FER)

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

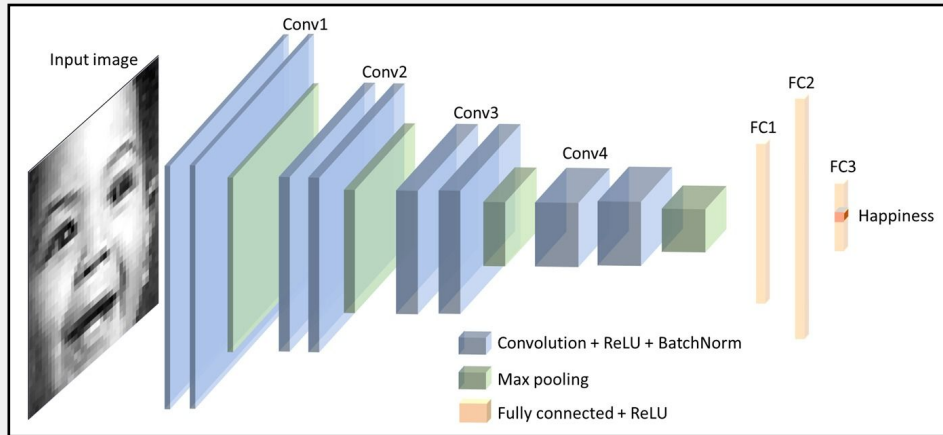
The VGGNet model, short for Visual Geometry Group Network

- *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.

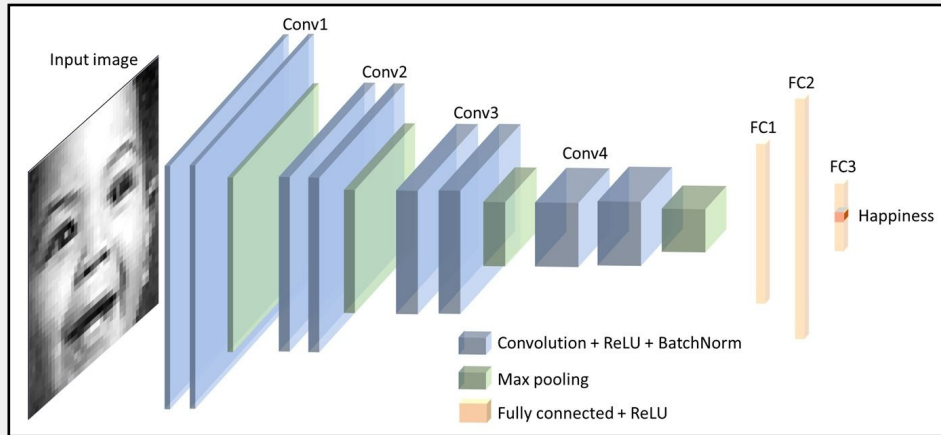


- **Each convolutional stage:** 2 convolutional blocks & a max-pooling layer.
- **Convolution block:** consists of a convolutional layer, a ReLU activation, and a batch normalization layer. Batch normalization is used to speed up the learning process, reduce the internal covariance shift, and prevent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

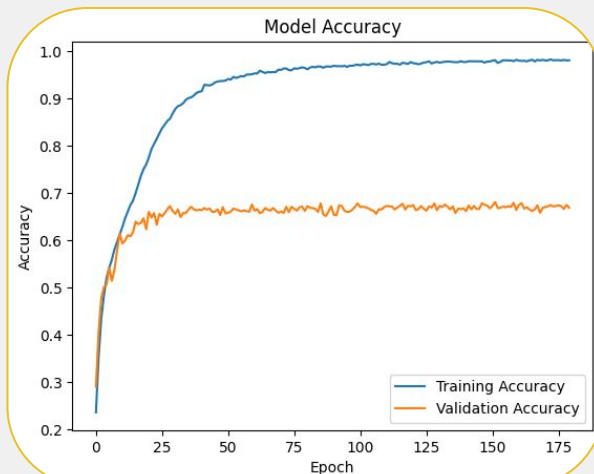
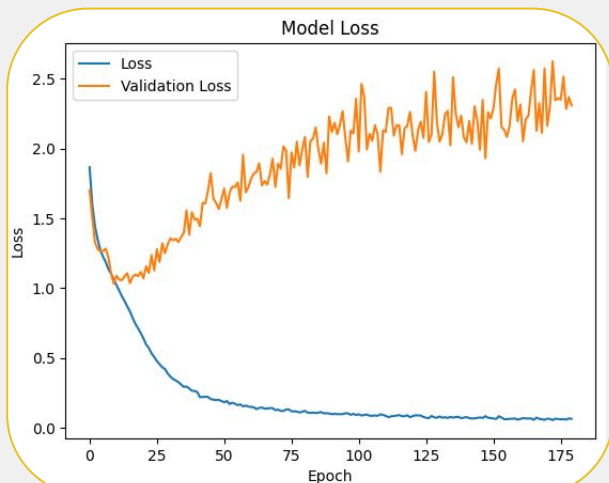
To afford high generalisability on the Egyptian/Arab race



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%



Confusion Matrix

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	276	6	53	23	58	5	46
Disgust	14	35	2	0	4	0	1
Fear	45	0	248	10	119	20	54
Happy	19	0	29	736	33	15	63
Sad	66	4	73	26	380	8	96
Surprise	13	0	44	22	10	314	12
Neutral	36	0	42	41	114	3	371

Y-axis: true label
X-axis: predicted label



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

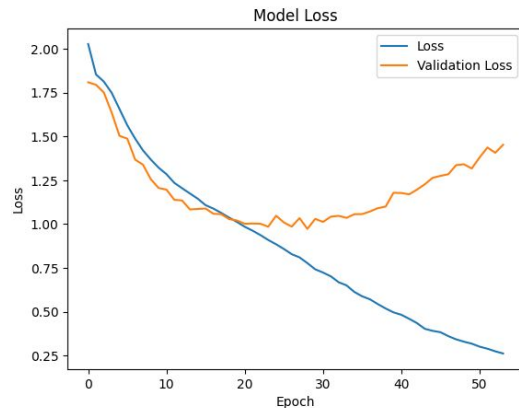
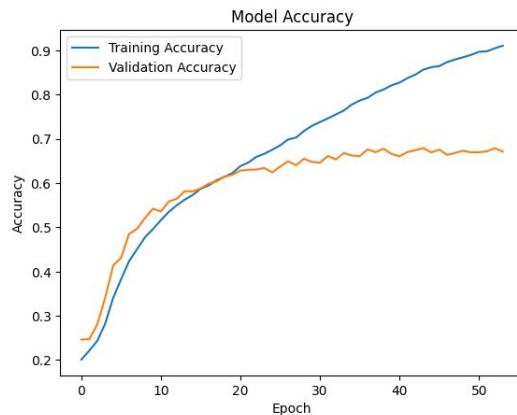
- Training process stopped after 54 epochs
 - Top-1 Accuracy: 66.15%
 - Top-2 Accuracy: 82.22%
 - Top-3 Accuracy: 90.89%

Confusion Matrix

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	241	13	58	31	70	8	46
Disgust	11	36	4	0	3	0	2
Fear	44	6	234	21	112	33	46
Happy	11	2	17	787	17	19	42
Sad	48	4	74	40	376	9	102
Surprise	9	1	37	18	13	327	10
Neutral	31	1	42	68	85	7	373

true label

predicted label



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

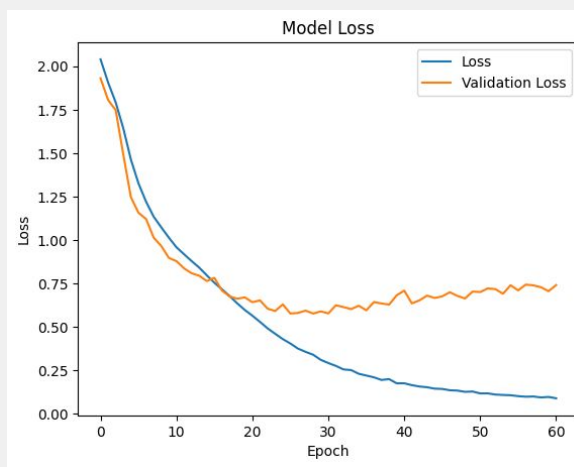
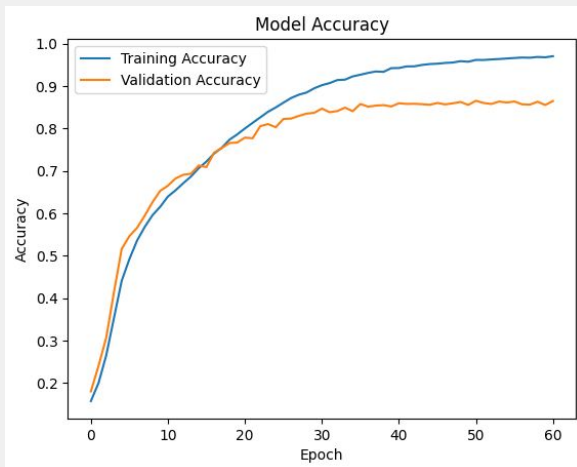
Highly Imbalanced
A total of 32,809 images

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%



Confusion Matrix

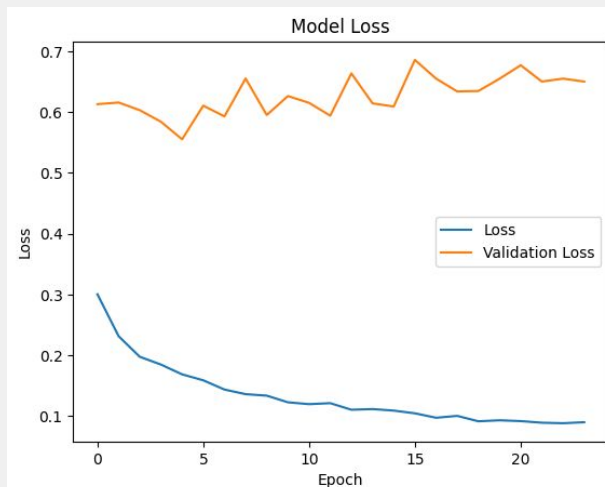
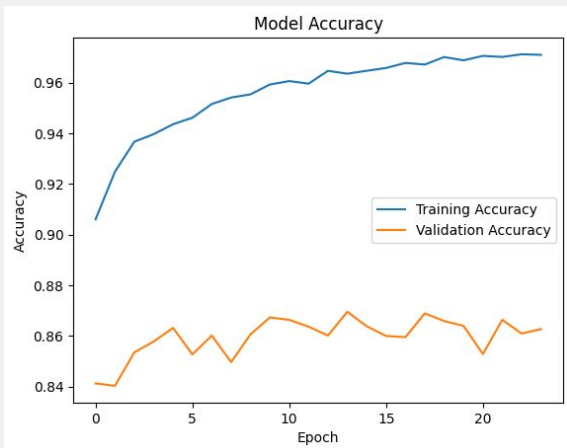
	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	734	5	33	16	55	4	34
Disgust	0	862	0	0	0	0	0
Fear	42	2	708	13	65	29	31
Happy	18	2	10	758	28	17	52
Sad	43	1	47	28	739	5	94
Surprise	3	3	24	15	2	872	6
Neutral	30	0	18	44	73	5	723

true label

predicted label

SMOTE Model

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



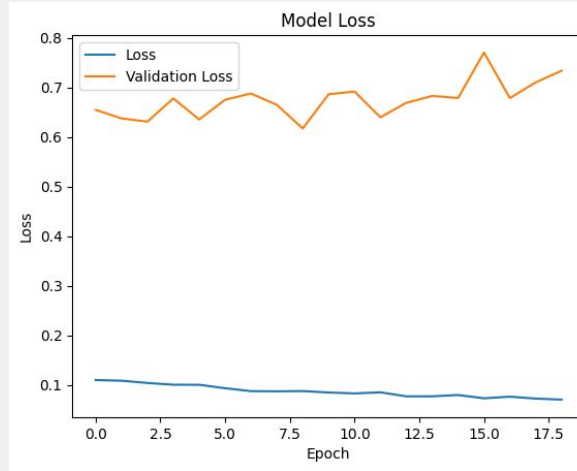
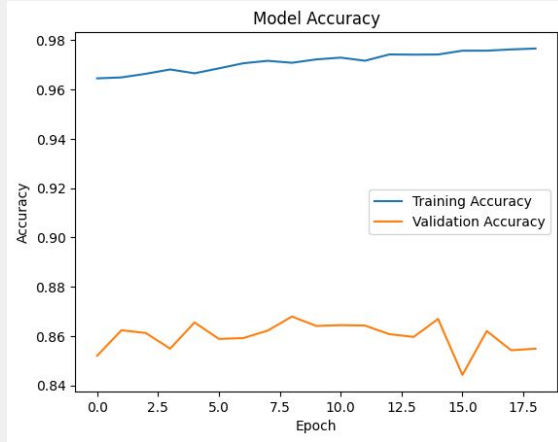
Confusion Matrix

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	747	6	36	14	49	1	28
Disgust	0	862	0	0	0	0	0
Fear	39	3	765	3	39	19	22
Happy	22	3	24	729	34	17	56
Sad	48	1	72	16	743	8	69
Surprise	6	1	25	10	4	872	7
Neutral	23	2	27	27	73	4	737

Y-axis: true label
X-axis: predicted label

SmoteTomek Model

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



Confusion Matrix

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	756	5	22	17	46	4	31
Disgust	0	862	0	0	0	0	0
Fear	53	3	718	6	59	22	29
Happy	18	0	17	763	26	12	49
Sad	46	3	53	25	758	7	65
Surprise	4	2	21	10	1	880	7
Neutral	23	0	21	41	65	5	738

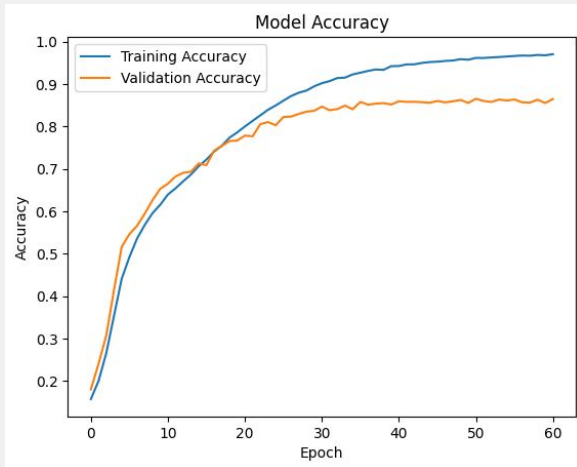
Comparative Analysis

Random

Oversampling (ROS)

61 Trained Epochs

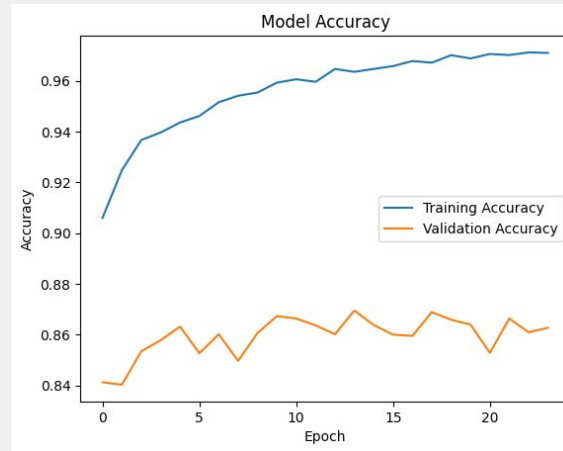
- Top-1 Accuracy: 85.75%
- Top-2 Accuracy: 92.74%
- Top-3 Accuracy: 96.55%



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

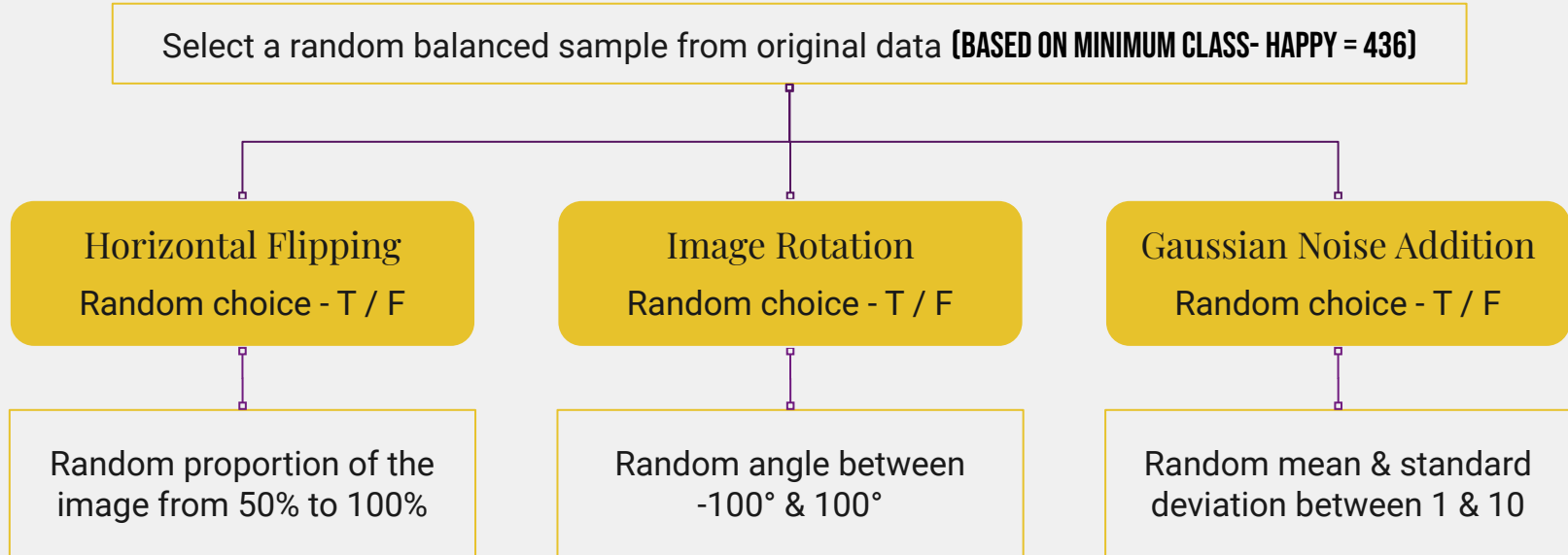


Gaussian Noise Addition



3. Image Manipulation

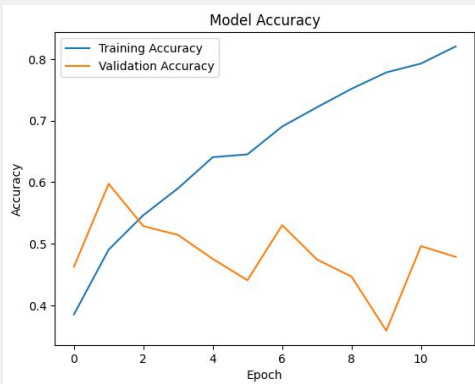
- Adopted suitable approaches to FER problem: only 3 techniques
- Randomization used to select technique(s) and amount of change in each image



3. Image Manipulation

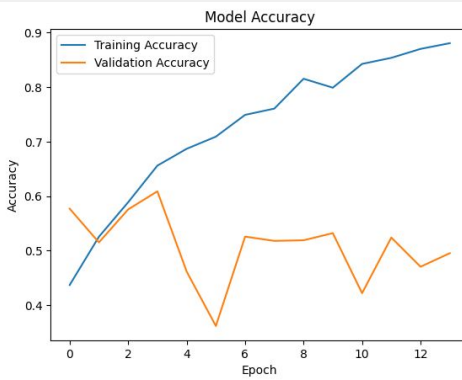
Random Oversampling

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



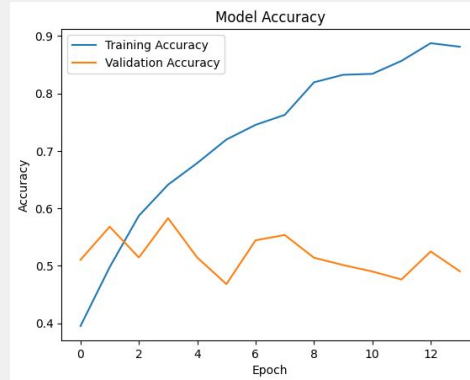
SMOTE

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%

		Confusion Matrix						
true label	Angry	290	6	38	13	93	9	42
	Disgust	20	29	1	0	1	1	3
	Fear	116	1	150	12	143	72	34
	Happy	63	0	23	645	66	36	46
	Sad	81	5	38	15	369	5	81
	Surprise	18	3	45	13	9	322	6
	Neutral	58	6	32	22	160	9	339
		Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
		predicted label						

SMOTE

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

		Confusion Matrix						
true label	Angry	182	12	18	35	169	14	61
	Disgust	4	35	1	2	11	1	1
	Fear	35	5	118	30	184	66	90
	Happy	7	3	6	725	66	22	50
	Sad	15	3	12	34	410	8	112
	Surprise	6	2	7	26	24	324	27
	Neutral	15	4	6	67	135	8	391
		Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
		predicted label						

SmoteTomek

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

		Confusion Matrix						
true label	Angry	210	65	27	26	77	28	58
	Disgust	0	51	0	2	1	1	0
	Fear	61	39	120	36	117	104	51
	Happy	33	36	13	669	48	32	48
	Sad	45	38	34	47	332	20	78
	Surprise	6	6	10	21	10	355	8
	Neutral	23	34	13	36	140	25	355
		Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
		predicted label						

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

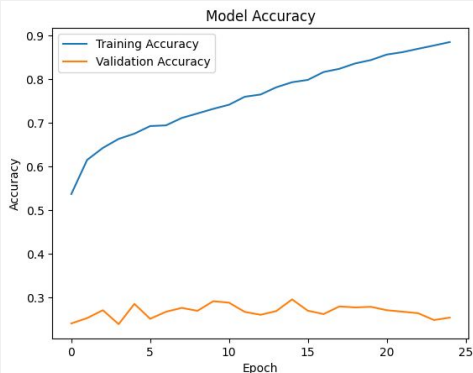
- Over 91,000 instances → large size (given GPU constraints)



4. Auxiliary Data

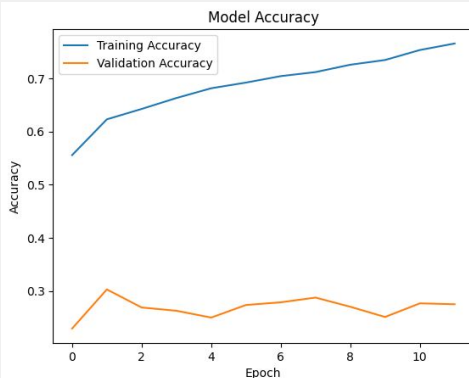
Random Oversampling

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



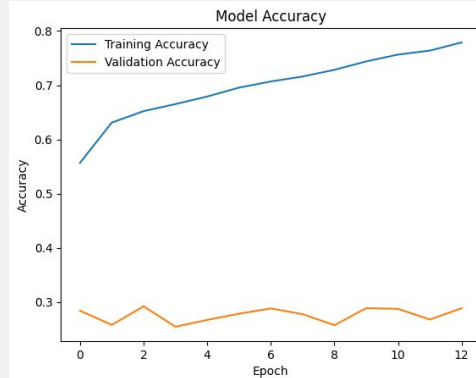
SMOTE

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%

Confusion Matrix

Angry	170	53	54	14	112	58	30
Disgust	10	32	5	3	3	2	0
Fear	73	31	95	13	166	127	23
Happy	54	153	25	105	175	350	17
Sad	83	19	25	29	318	60	60
Surprise	10	3	77	3	25	295	3
Neutral	82	32	26	5	300	137	44

true label

predicted label

SMOTE

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

Confusion Matrix

Angry	224	11	18	0	172	46	20
Disgust	23	10	3	0	17	2	0
Fear	97	2	40	7	226	126	30
Happy	75	18	41	44	411	259	31
Sad	68	0	9	3	411	40	63
Surprise	30	1	36	1	30	315	3
Neutral	75	3	12	1	405	86	44

true label

predicted label

SmoteTomek

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

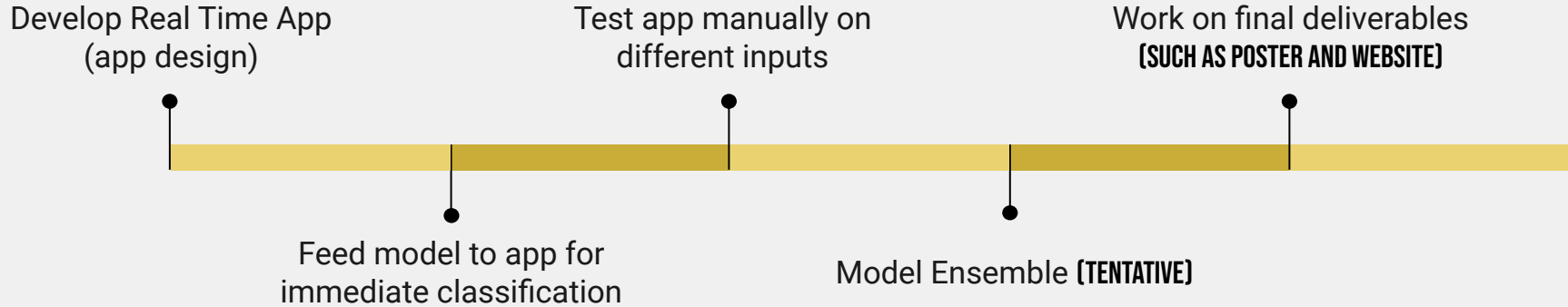
Confusion Matrix

Angry	121	49	48	8	186	61	18
Disgust	9	16	2	2	22	4	0
Fear	40	27	99	8	213	126	15
Happy	11	78	61	91	327	304	7
Sad	28	9	32	19	404	73	29
Surprise	2	5	100	2	21	285	1
Neutral	28	20	26	15	399	107	31

true label

predicted label

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 Gabal - 900201683
Masa Tantawy - 900201312



CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**