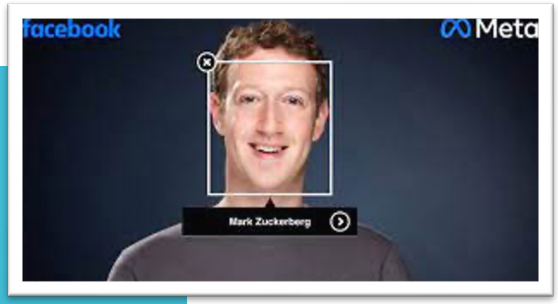
Facial Emotion Detection

Allison Dzubak

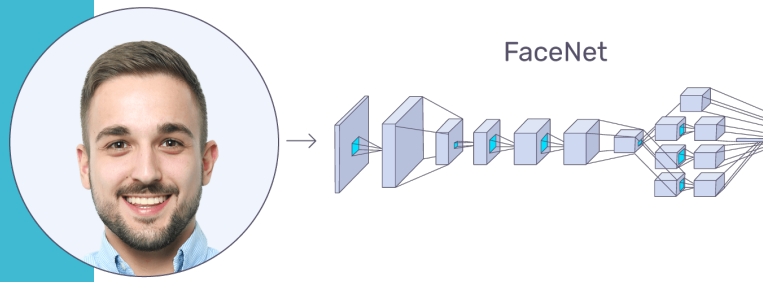
Feb 24, 2024

Current Capabilities



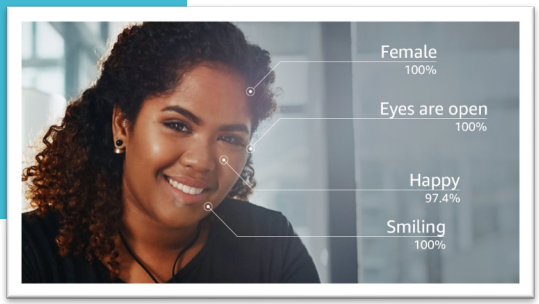
Facebook's New Face-Recognition Software Is Scary Good

SLATE



Google: Our new system for recognizing faces is the best one ever

FORTUNE



Justice Department discloses FBI project with Amazon Rekognition tool

FEDSCOOP

Specific FER Use Case

Stanford MEDICINE

Helping Kids with Autism Read Facial Expressions

Goal: Classify an emotion from a facial image

The Available Data

- Quality
- Diversity
- Class characteristics

Happy

Sad

Surprise

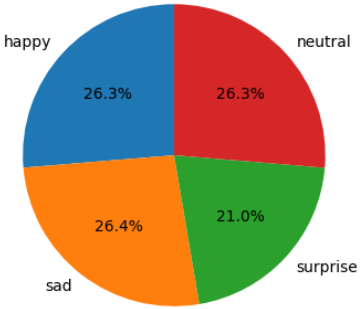
Neutral

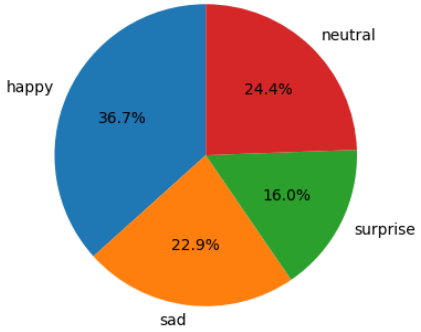
The Available Data

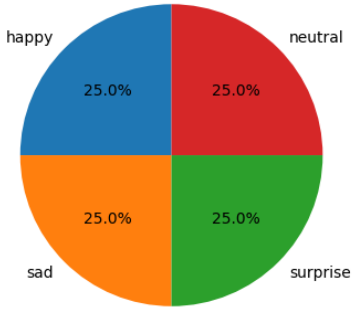
Distribution

20,214 Labeled Images

Train set	Validation set	Test set
15,109	4,977	128
74.8%	24.6%	0.6%







The Available Data

Challenges

Illumination

Camera viewpoint

Watermarks

Non-human faces

Occlusions

Misalignment / Cropping

The Available Data

Observed Issues

Happy	Neutral	Sad	Surprise
			
			
			
			
			

Data Cleaning

Face Detection

Duplicate Detection

- Inter-class duplicates: Exclude
- Intra-class duplicates: Exclude to balance train/validation

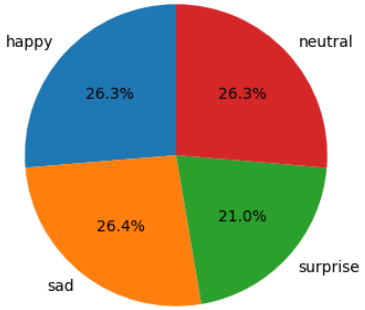
Mislabeled Images

- Known issue: Still to be addressed

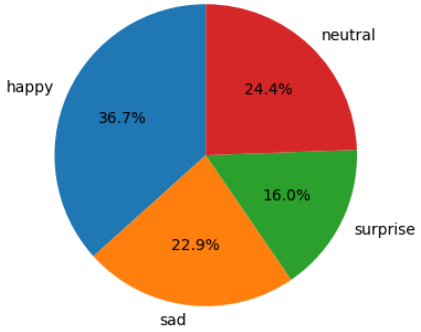
Data Cleaning

Original:
20,214 images

Train set
15,109
74.8%

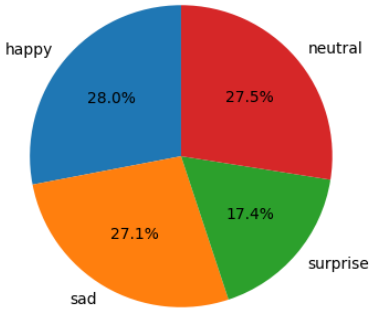


Validation set
4,977
24.6%

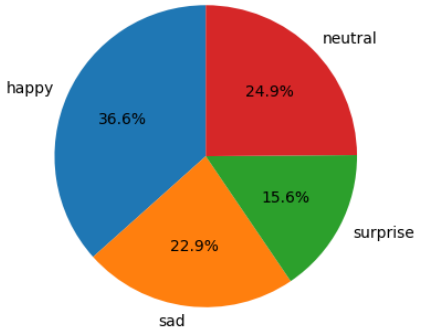


Cleaned:
18,931 images

Train set
13,972
73.8%



Validation set
4,831
25.6%



Model Descriptions

Models

- Pre-trained models: VGG16, ResNet50V2, EfficientNetB0
- Loss/Activation functions: Categorical crossentropy / Softmax
- Optimizer: Adam
- Metric: Accuracy

Versions

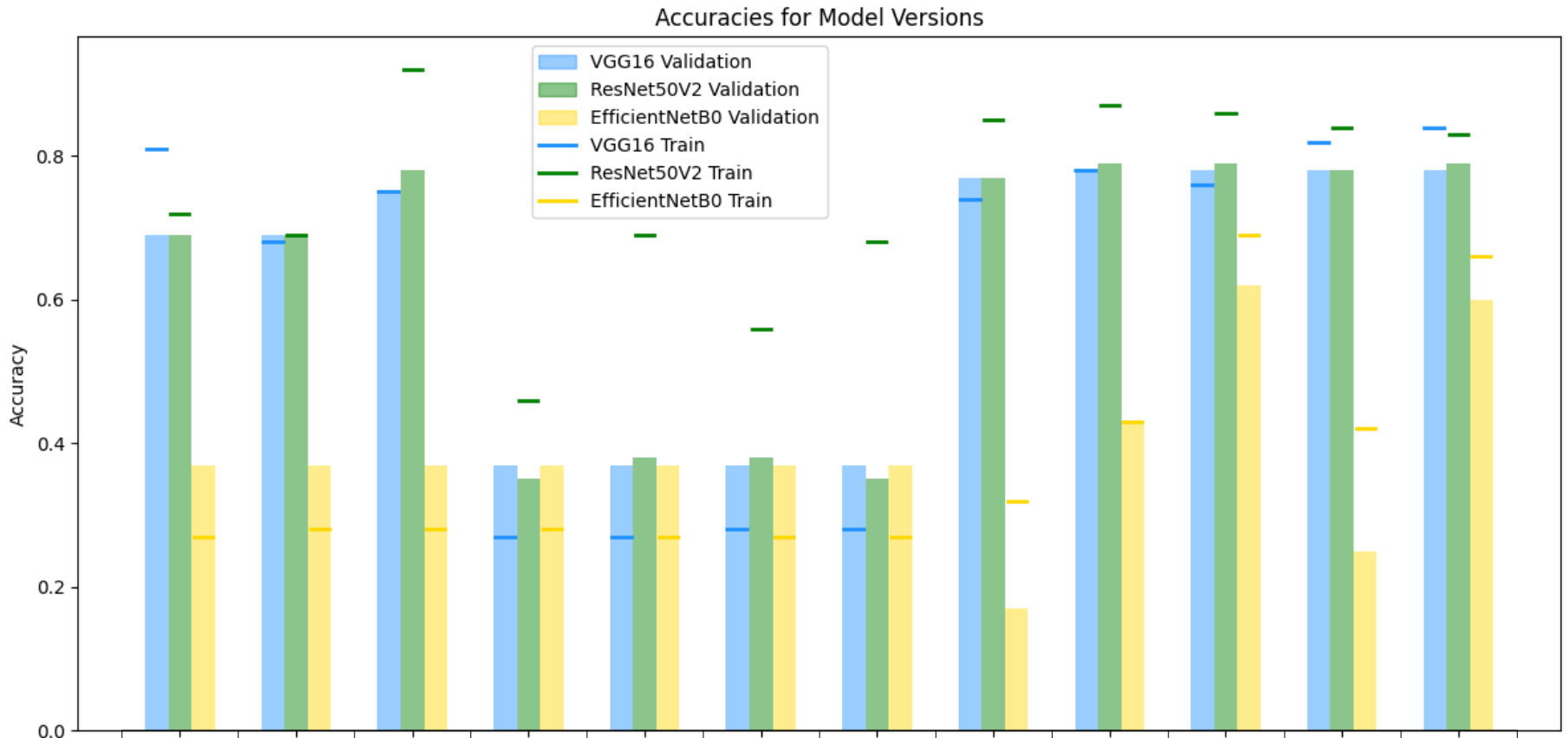
- Base model
- Varying Levels of Data Augmentation
- Varying number of convolution blocks frozen vs. trainable
- Adding an additional dense layer on top
- Varying initial learning rate
- Varying batch size
- Adding class weight balancing

Data Augmentation Levels

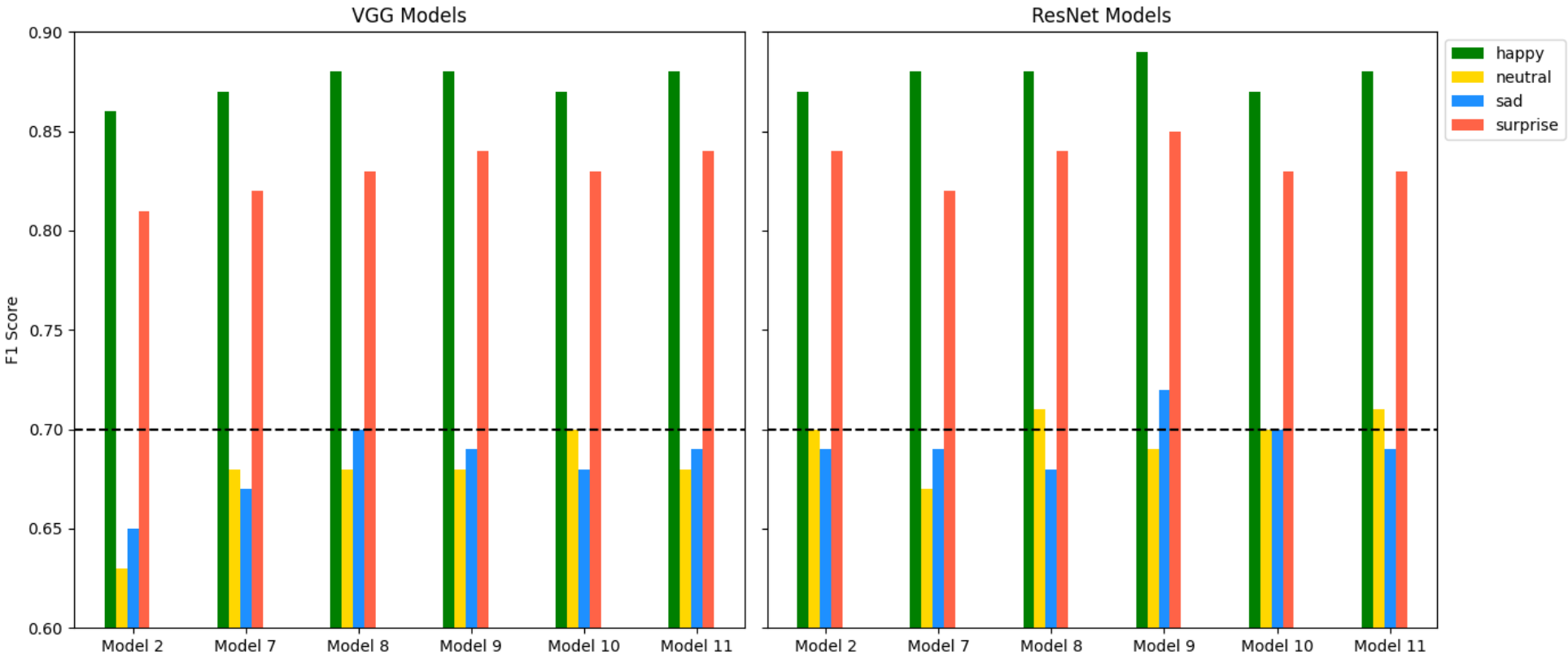
Low data augmentation:		True	5%	5%	5%	
Mid data augmentation:	15 degrees	True	5%	5%	5%	
High data augmentation:	15 degrees	True	5%	5%	5%	Exposure & gamma

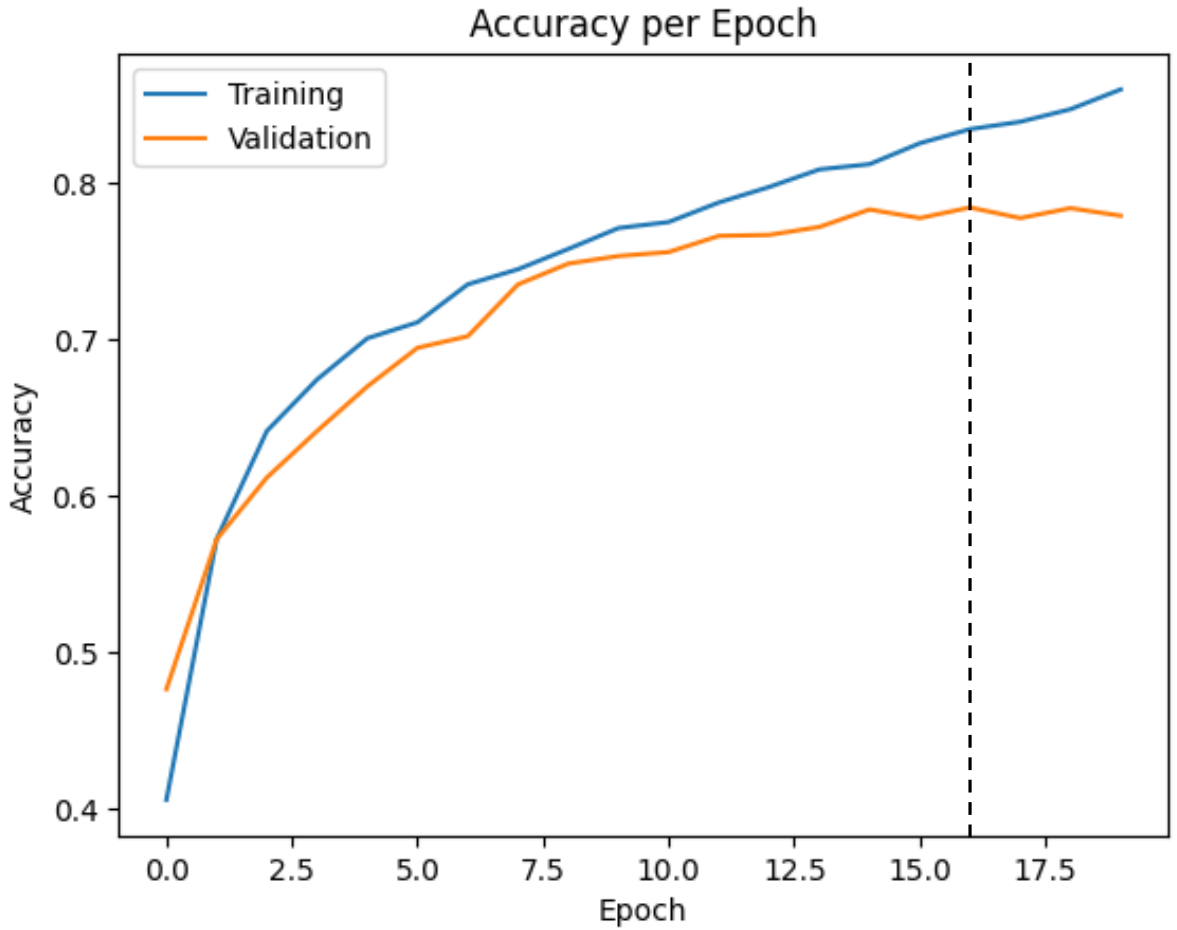
Model Version Descriptions

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Data Augmentation:	none	low		high				mid				
Learning Rate:	0.001		0.0001				0.00001					
Unfreeze top # conv. blocks:	0		1		2				3	all		
Batch size:	32										128	
						add dense layer						weight

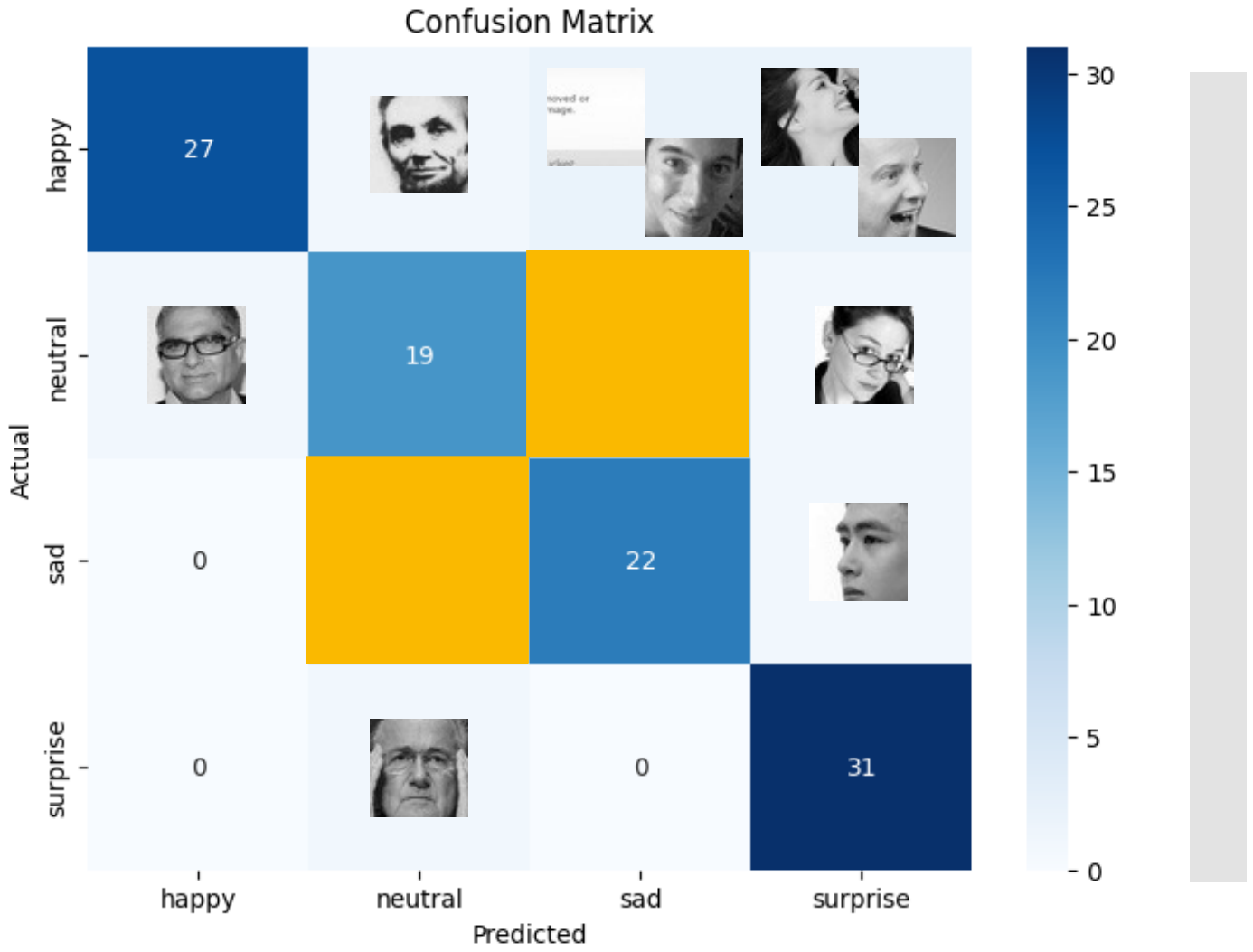


	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Data Augmentation:	none	low		high				mid				
Learning Rate:	0.001			0.0001				0.00001				
Unfreeze top # conv. blocks:	0			1			2			3		all
Batch size:	32											128
						add dense layer						weight

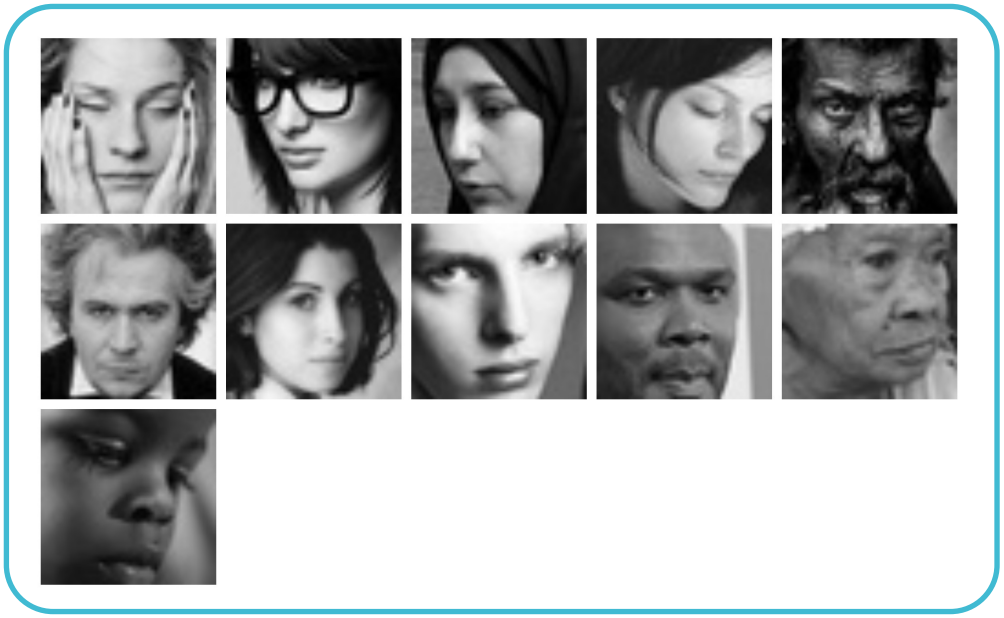

VGG Models
ResNet Models
F1 Score
0.90
0.85
0.80
0.75
0.70
0.65
0.60
Model 2
Model 7
Model 8
Model 9
Model 10
Model 11
happy
neutral
sad
surprise


Accuracy per Epoch
Training
Validation
Accuracy
0.4
0.5
0.6
0.7
0.8
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Epoch

Unseen Test Data Performance



Unseen Test Data Performance



Actual Sad?

Actual Neutral?

Predicted Sad?

Predicted Neutral?

Unseen Test Data Performance

	Validation	Test
Accuracy	0.78	0.77
F1-happy	0.88	0.90
F1-neutral	0.71	0.61
F1-sad	0.69	0.66
F1-surprise	0.83	0.93

Further Analysis Needed

- Human accuracy on emotion labeling within our dataset?
- How much does the model need to generalize?
- How do other FER-specific trained models perform in comparison?
- Does the model have any gender/racial/age biases?
- What features has the model actually learned?
- Must the emotion classes be mutually exclusive?