FACIAL EMOTION DETECTION

DEEP LEARNING CAPSTONE PROJECT
FE VALVEKENS
AUGUST 2023

PROBLEM DEFINITION

55% of emotional information are visual

Increasing human-machine interaction

Role of machines with emotional intelligence

PROBLEM TO SOLVE

- We are aiming to solve a multi-class classification problem
- What model architectures will be used?
- What data preprocessing steps are required?
- How shall we evaluate each model?

SOLUTION APPROACH

- Perform Exploratory Data Analysis
- Explore two types of architectures
- Compare and evaluate each model

EXPLORATORY DATA ANALYSIS

KEY RESULTS

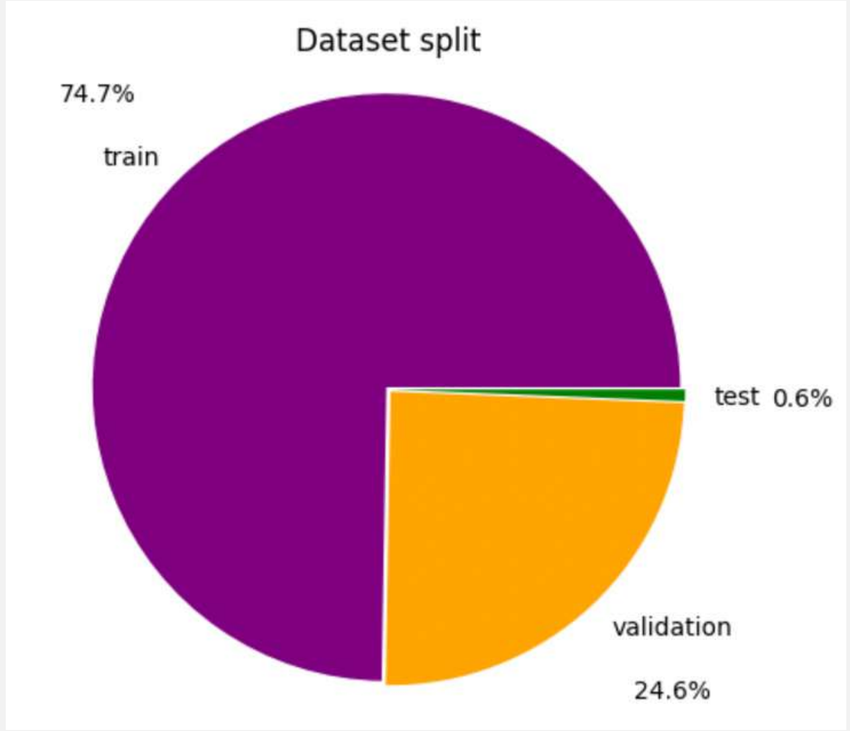


Image color & size

Images are grayscale and of relative small size 48 x 48 pixels.
View a sample of images per class in Appendix 1.

Sample observations

- Different age, gender and ethnicity.
- Various backgrounds
- Faces are of different sizes, and are tilted at different angles.

Class distribution

Four classes: 'happy', 'sad', 'neutral', 'surprise' with a slight imbalance for the class 'surprise'.
See Appendix 2 for the class distribution.

Dataset split

Three datasets:

- train: 15,109 images
- validation: 4,977 images
- test: 128 images

CONVOLUTIONAL NEURAL NETWORKS

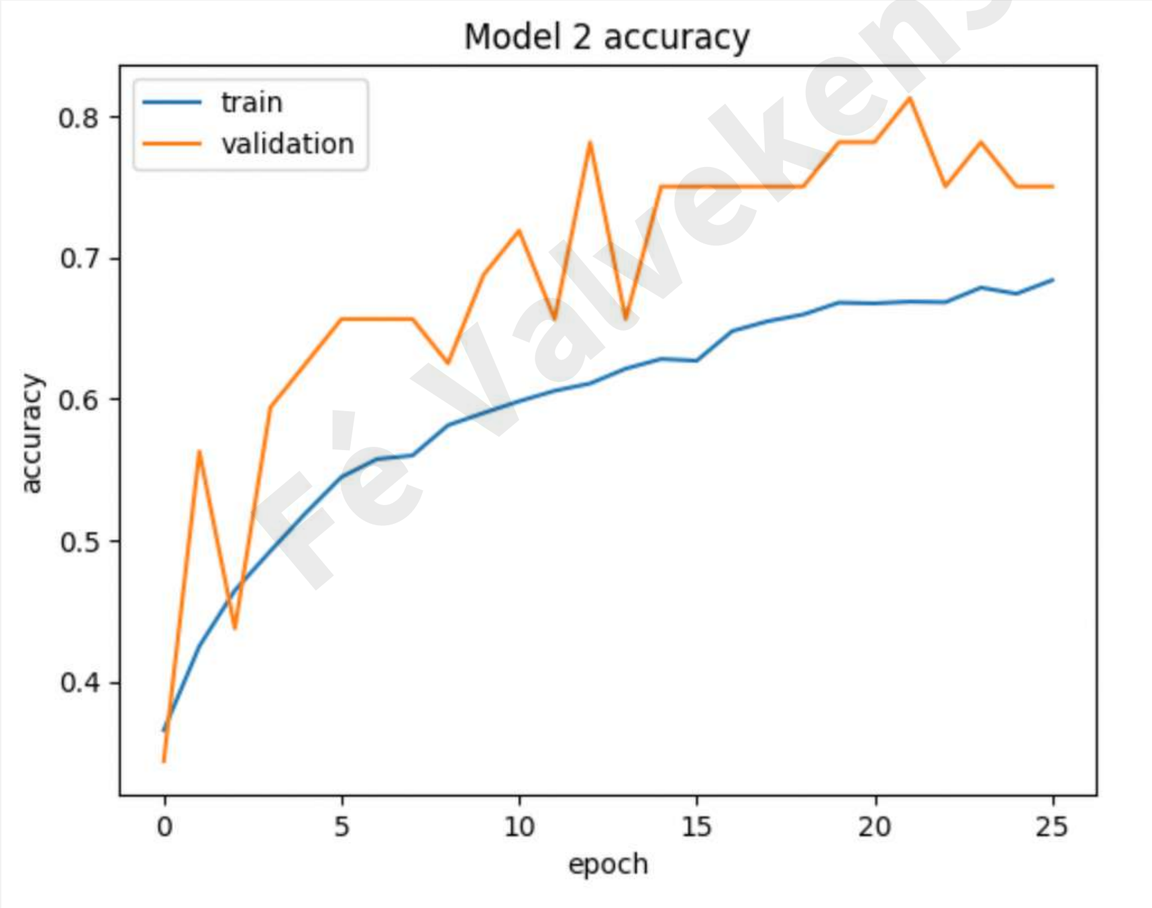
KEY FINDINGS

CNN1

7.4 M trainable parameters

CNN2

1 M trainable parameters



Accuracy

Good performance on test set with slightly more complex model CNN2:

- 75% grayscale
- 65% RGB mode

See Appendix 3 for detailed results.

Color mode

Affects accuracy.

Data Augmentation

Augmented images:

- rotation
- shear
- brightness
- zoom

TRANSFER LEARNING

KEY FINDINGS

VGG16

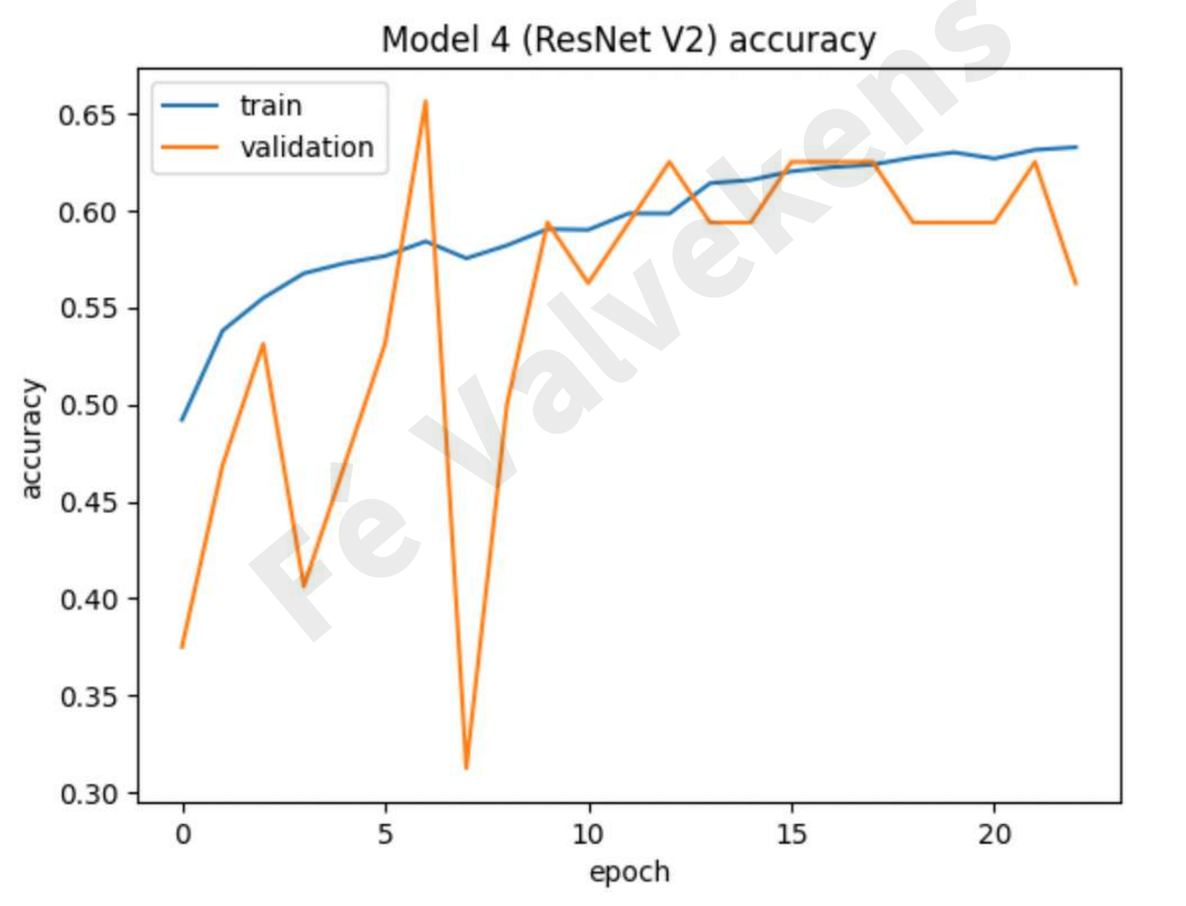
14.7 M trainable parameters

ResNet V2

42 M trainable parameters

EfficientNet B7

63 M trainable parameters



Accuracy

Overall unsatisfactory performance on test set:

- VGG16: 56%
- ResNet V2: 60%
- Efficient B7: 50%

See Appendix 4 for detailed results.

Color mode

Required RGB color mode

PROPOSED MODEL SOLUTION

COMPLEX CONVOLUTIONAL NEURAL NETWORK

Complex CNN

3 M trainable parameters

	Correct prediction*
happy	88%
neutral	78%
sad	75%
surprise	91%

* recall scores from the classification report

Accuracy

Satisfactory performance on test set:

- 82% grayscale
- 74% RGB mode

See Appendices 5 & 6 for detailed results.

Limitations

- Lower accuracy on classes 'neutral' and 'sad'
- Lower accuracy with facial occlusion (Appendix 7)
- training requires significant computational resources

POTENTIAL BENEFITS

Accuracy

The highest level of a accuracy.

Robustness

Good generalisation allowing different genders, age, ethnicity and poses.

Flexibility

Can be adapted to chosen color mode and size of the input images.

Scalability

Can handle large datasets and process high volume of images.

KEY TAKE AWAYS

RECOMMENDATIONS

- Data preprocessing including data augmentation to further improve the model's performance
- Training on a larger dataset that would contain more diverse and representative data
- Expansion of types of emotions including 'anger', 'fear' and 'disgust'
- Cloud Computing as a cost-effective solution: only pay for the computing resources used

EXECUTIVE SUMMARY

EDA

The dataset is composed of small size grayscale images of faces expressing four types of emotion: happy, sad, neutral or surprised.

ARCHITECTURE

Using Transfer Learning did not prove beneficial.

A CNN model is the preferred approach.

SOLUTION

A Complex CNN model achieving 82% accuracy and that is relatively computationally efficient.

LIMITATIONS

Lower accuracy on classes 'neutral' and 'sad', on images with facial occlusion.
Training requires high computational resources.

BENEFITS

Accuracy, robustness, flexibility and scalability.

RECOMMEN-DATIONS

Data Augmentation, training on larger dataset.
train on more classes, Cloud computing

RISKS AND CHALLENGES

- Ethical and privacy concerns
- Risk of bias
- Computational requirements and latency

THANK YOU

FE VALVEKENS

fe.valvekens@gmail.com
www.linkedin.com/in/fevalvekens/

“In this new economy, three groups will have a particular advantage: those who can work well and creatively with intelligent machines, those who are the best at what they do, and those with access to capital.”

Cal Newport, Deep Work, 2016

APPENDIX

1 - SAMPLE OF IMAGES PER CLASS

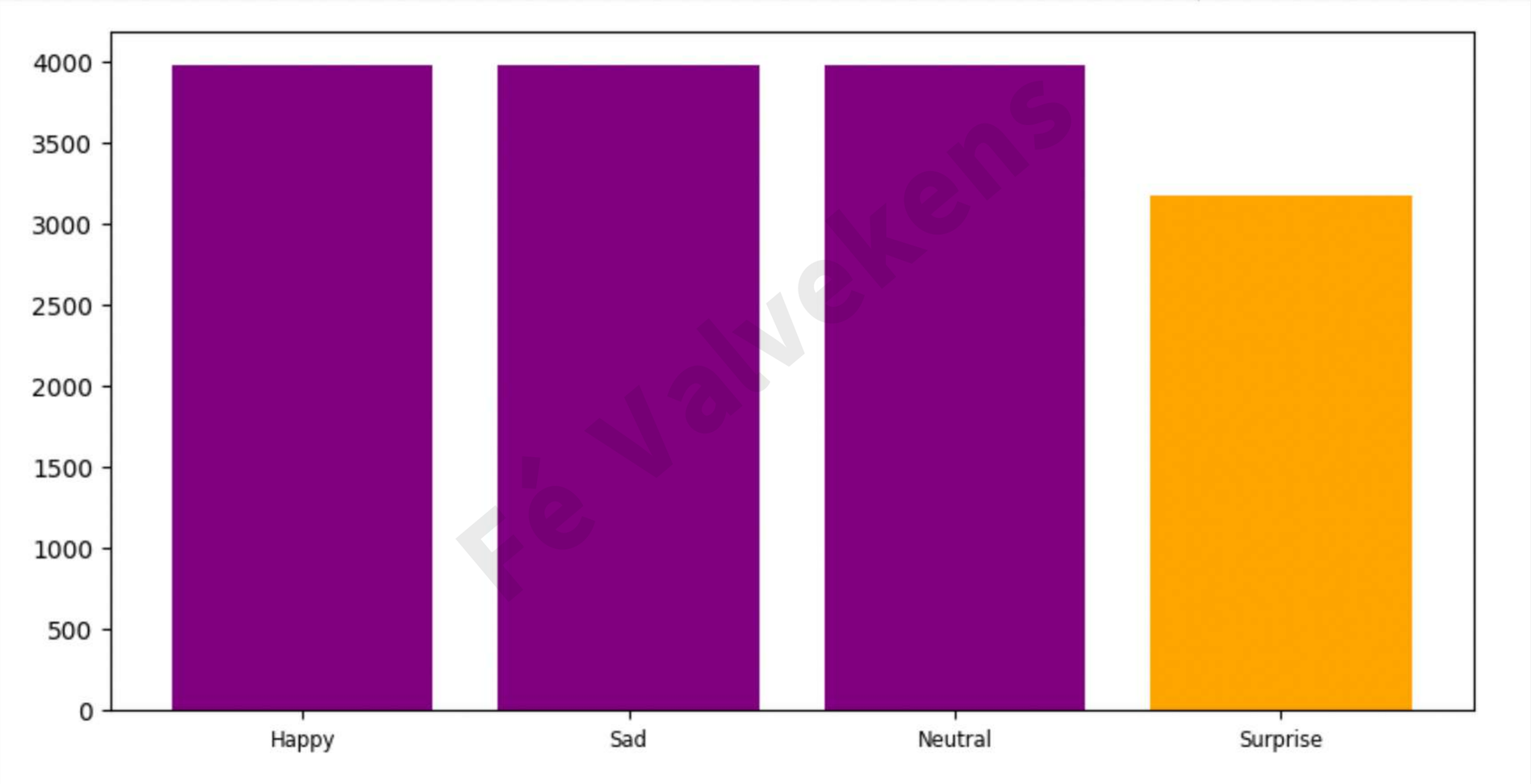
happy

sad

surprise

neutral

2 - CLASS DISTRIBUTION



3 - CNN PERFORMANCE

CNN Architecture: Model Performance Summary

	Training Accuracy	Validation Accuracy	Test Evaluation	Computation Time *
CNN 1 (g)	60%	68%	**64%**	16m
CNN 1 (rgb)	54%	64%	**58%**	15m
CNN 2 (g)	68%	81%	**75%**	11m
CNN 2 (rgb)	58%	62%	**65%**	11m

** using Google Colab on Macbook Pro M1 chip*

4 - TRANSFER MODEL PERFORMANCE

Transfer Learning Architecture: Model Performance Summary

	Training Accuracy	Validation Accuracy	Test Evaluation	Computation Time *
VGG16 (rgb)	54%	75%	**56%**	8m
ResNet V2 (rgb)	63%	65%	**60%**	14m
EfficientNet B07 (rgb)	38%	43%	**50%**	6m

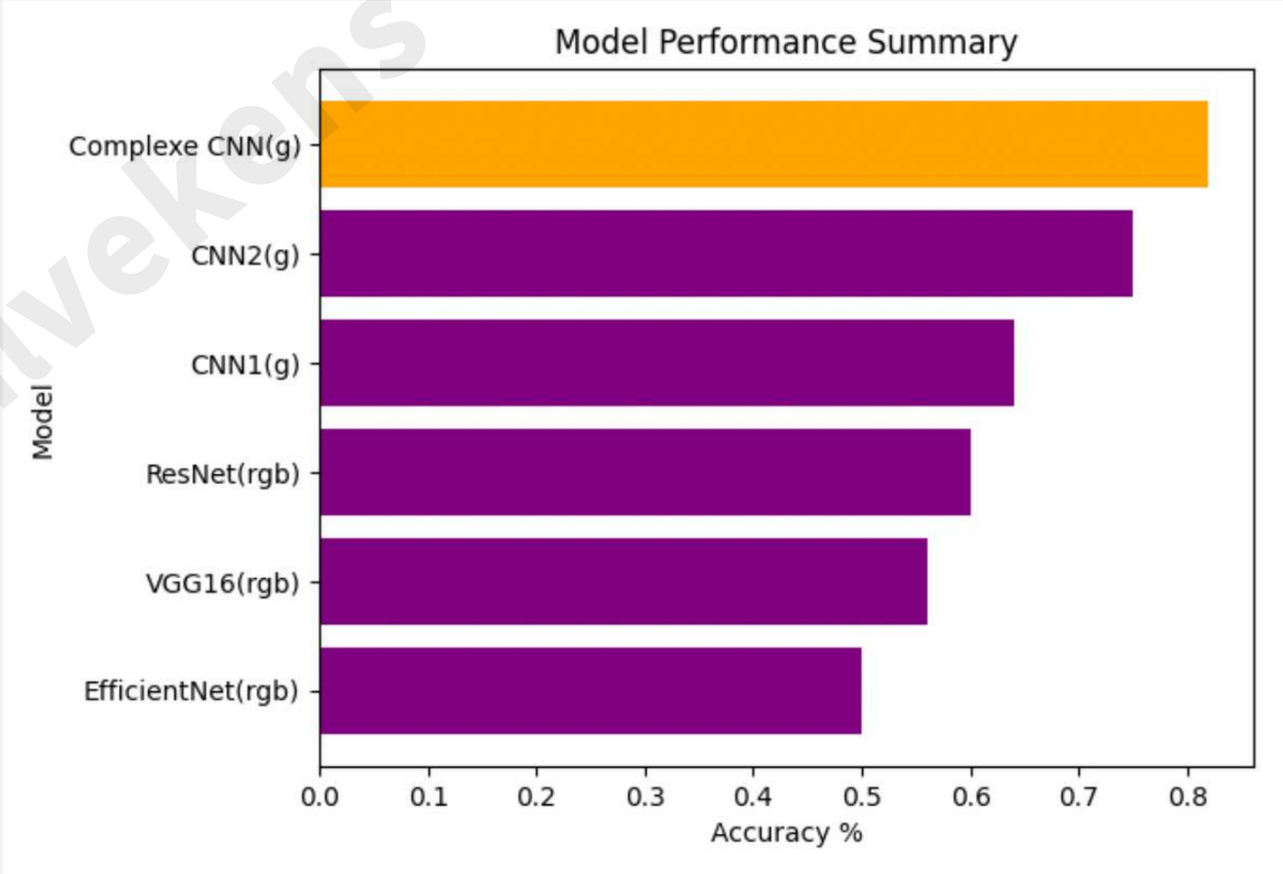
** using Google Colab on Macbook Pro M1 chip*

5 - COMPLEX CNN PERFORMANCE

Complex CNN: Model Performance Summary

	Training Accuracy	Validation Accuracy	Test Evaluation	Computation Time *
Complex CNN (g)	81%	75%	**82%**	8m
Complex CNN (rgb)	81%	68%	**74%**	10m

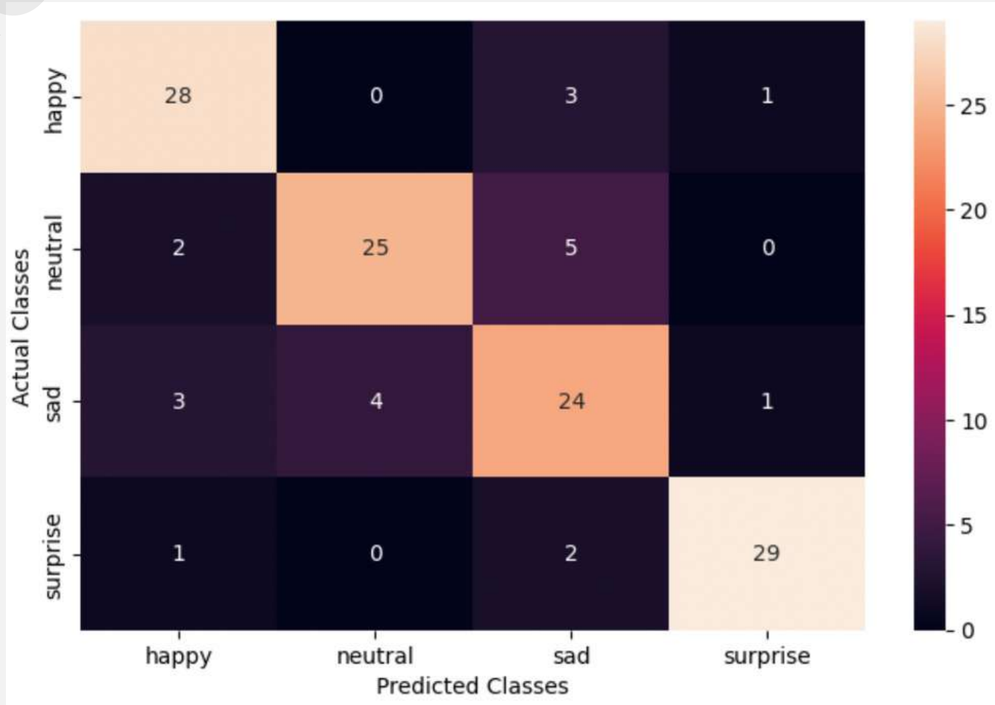
** using Google Colab on Macbook Pro M1 chip*



6 - CLASSIFICATION REPORT AND CONFUSION MATRIX

	Precision	Recall	F1-score
happy	82%	88%	85%
neutral	86%	78%	82%
sad	71%	75%	73%
surprise	94%	91%	92%

The classification report summary for the Complex CNN



Complex CNN model

7 - MISCLASSIFIED IMAGES

actual: happy
predicted: sad

Figure (c)

actual: neutral
predicted: sad

Figure (d)

actual: sad
predicted: happy

Figure (e)

actual: surprise
predicted: sad

Figure (f)

actual: sad
predicted: neutral

Figure (g)

actual: happy
predicted: sad

Figure (h)