UNIVERSITY OF INFORMATION TECHNOLOGY - VNU-HCM

COMPUTER VISION PROJECT : REAL-TIME EMOTION DETECTION USING CNN

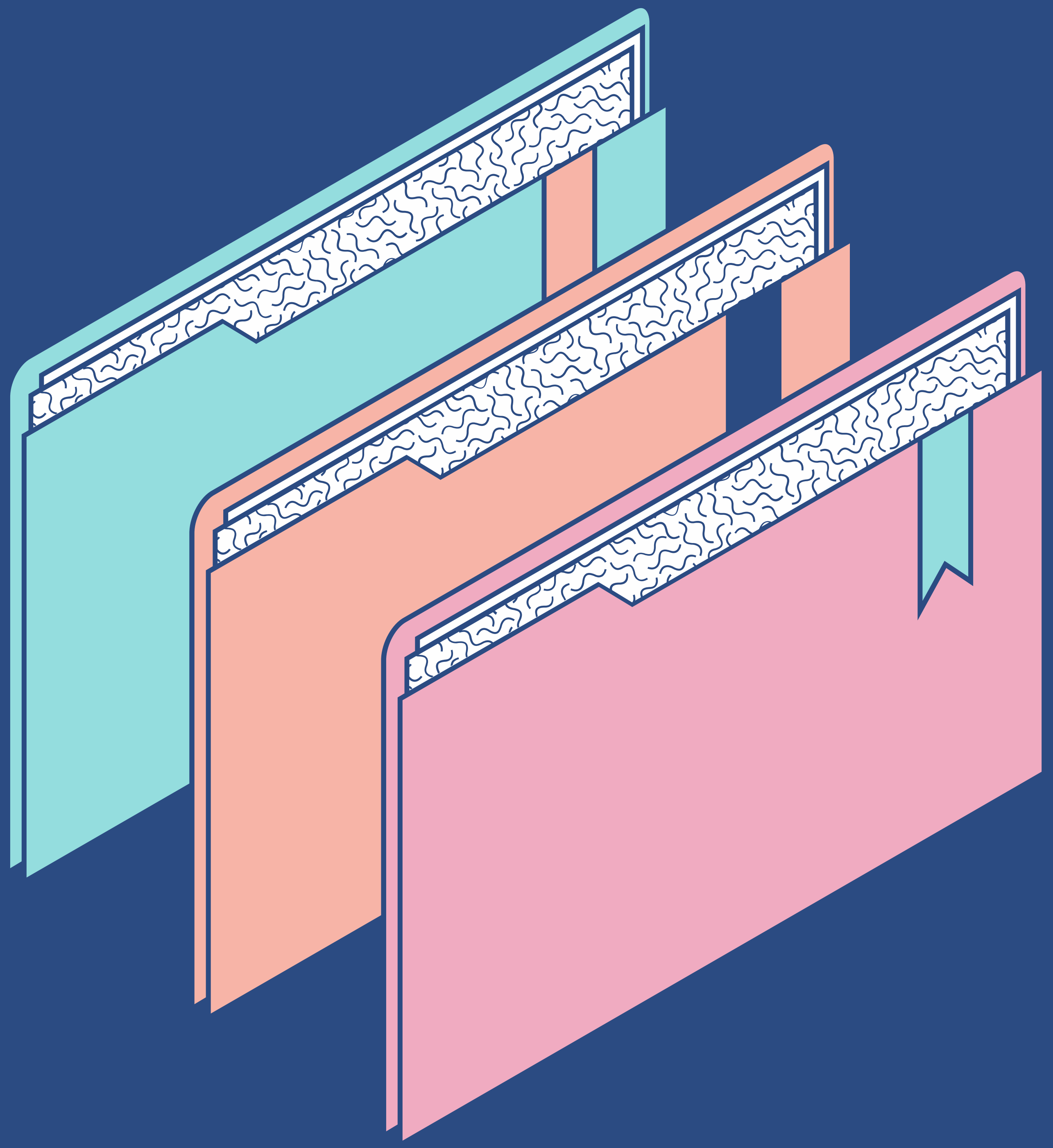
May 2023

MEMBER INFORMATION

Full name: Trần Xuân Minh
Student ID: 21520352

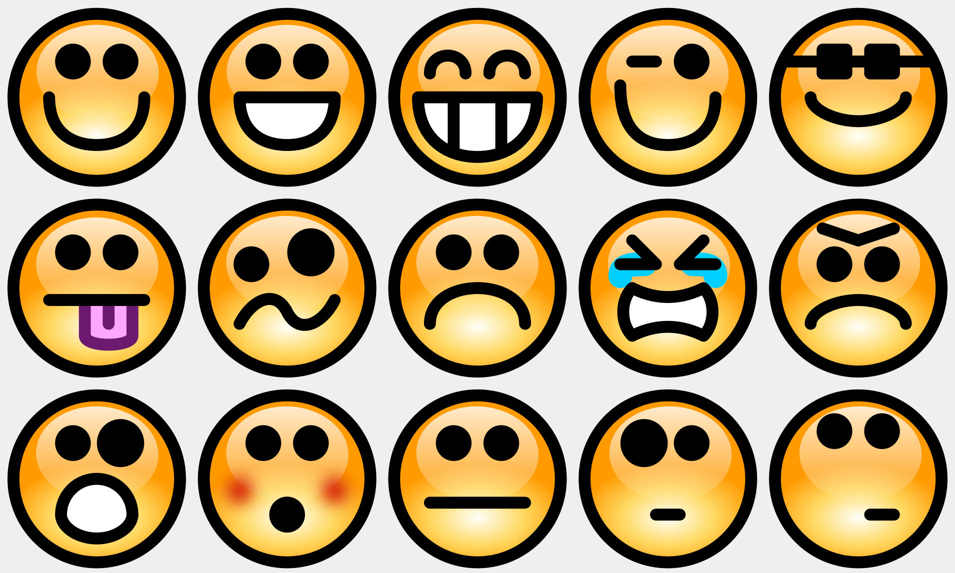
Full name: Lê Thu Hà
Student ID: 21520800

CONTENTS



Part 1: Problem Introduction

DETECTING EMOTION AT REAL-TIME

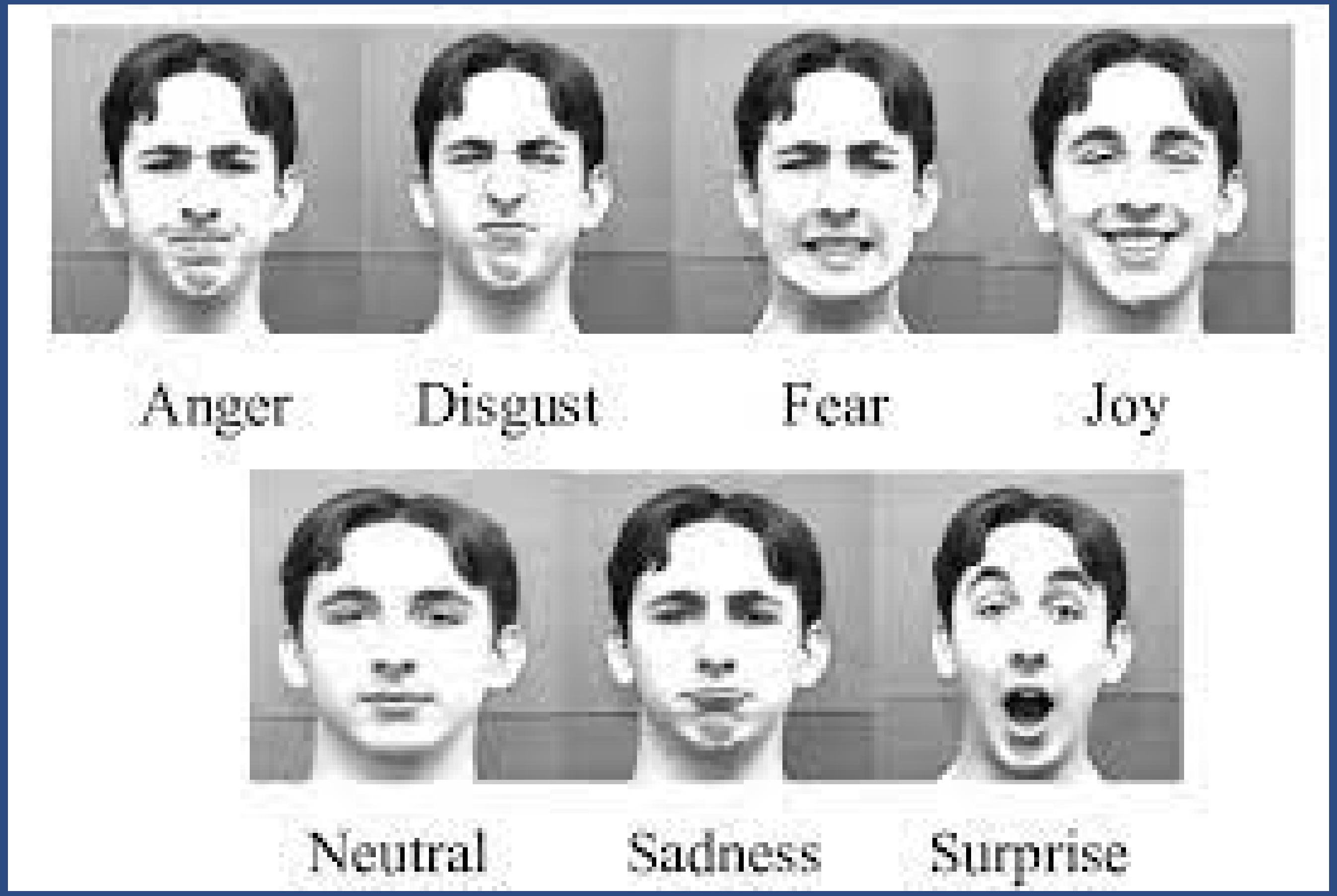


Introduction to Emotions and 7 Basic Emotions

- Emotions are complex psychological experiences that influence our thoughts, behaviors, and interactions.
- Categorizing emotions helps us understand and differentiate between different emotional states.
- The classification of emotions into 7 basic emotions is a widely accepted framework (Paul Ekman's facial-expression research).

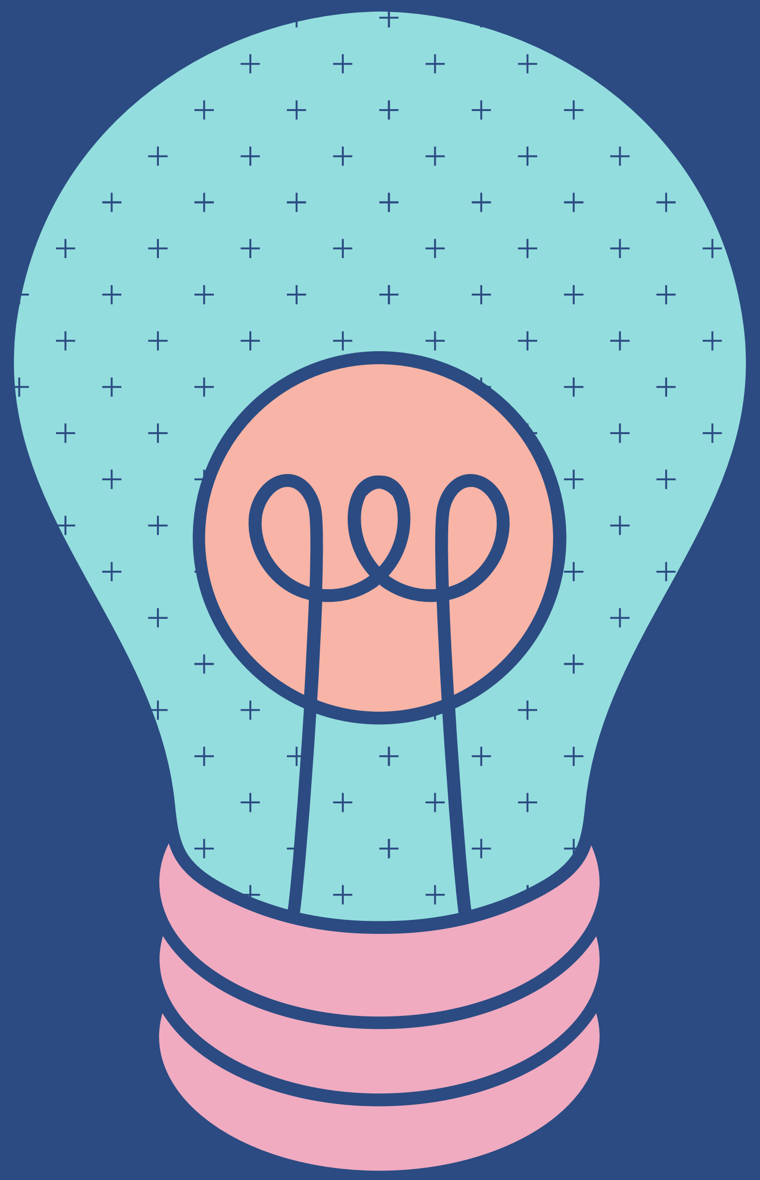
7 Basic Emotions

The 7 basic emotions (happiness, sadness, anger, fear, surprise, disgust, neutral) capture distinct and universal emotional states.



While emotions are diverse, the 7 basic emotions provide a framework for exploring the complexity of human emotional life.

The objective of real-time detecting emotion in this project is to develop a basic system that try to accurately detect and recognize human emotions in real time using computer vision and machine learning techniques which is particularly CNNs.



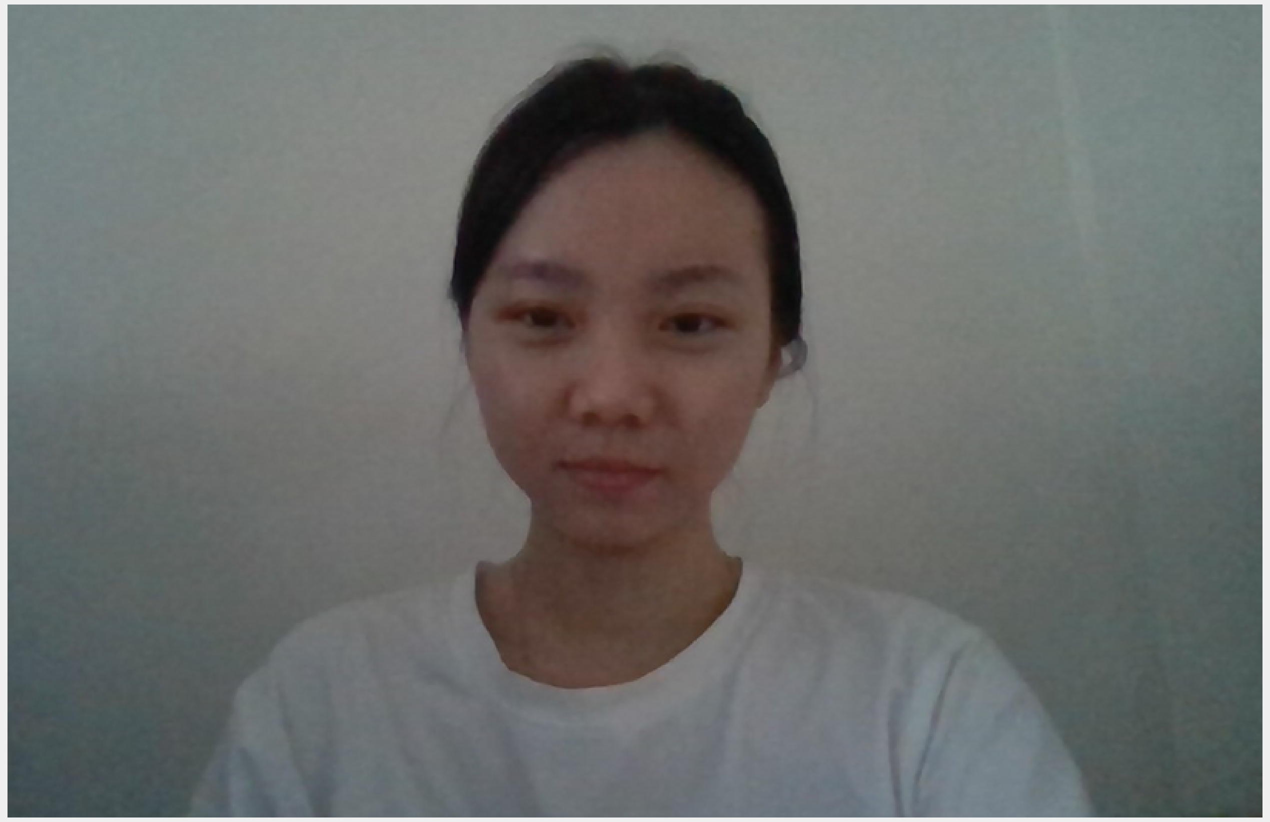
THE PROBLEM OF REAL-TIME DETECTING EMOTION

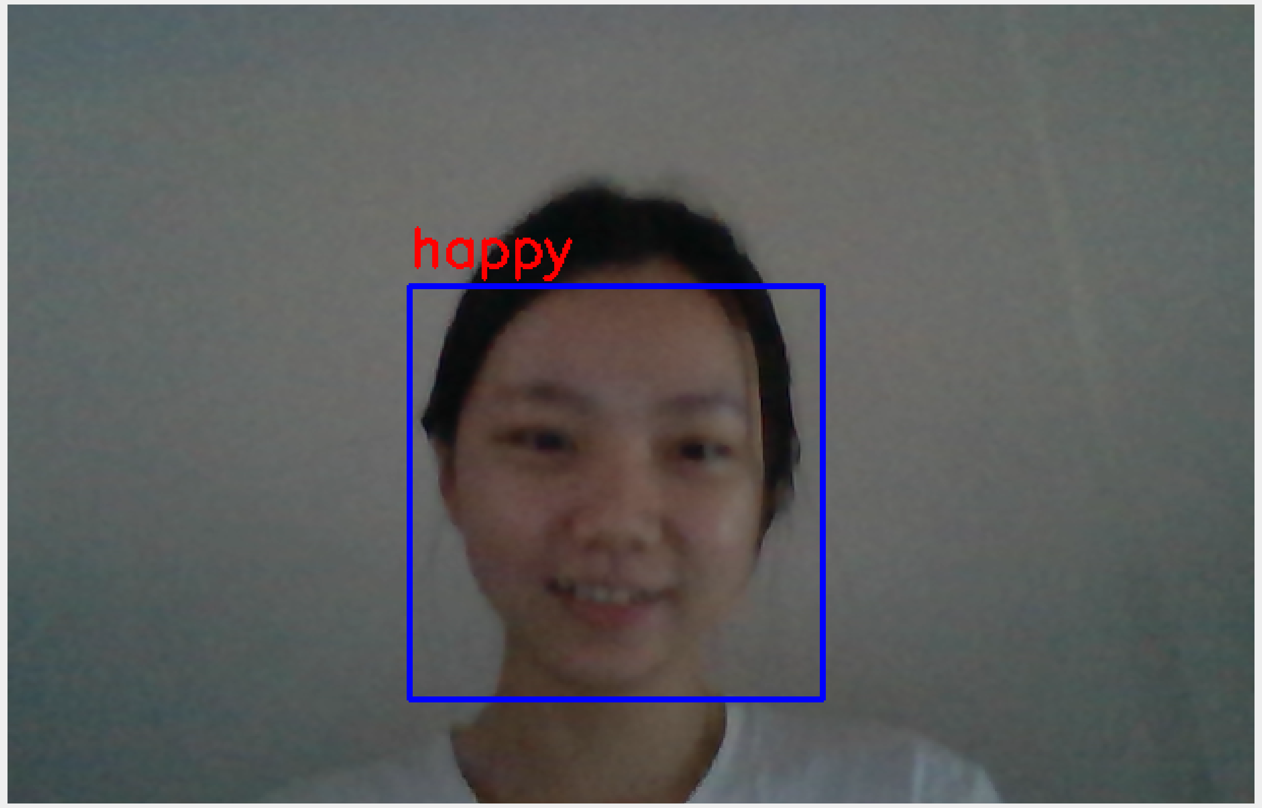
INPUT

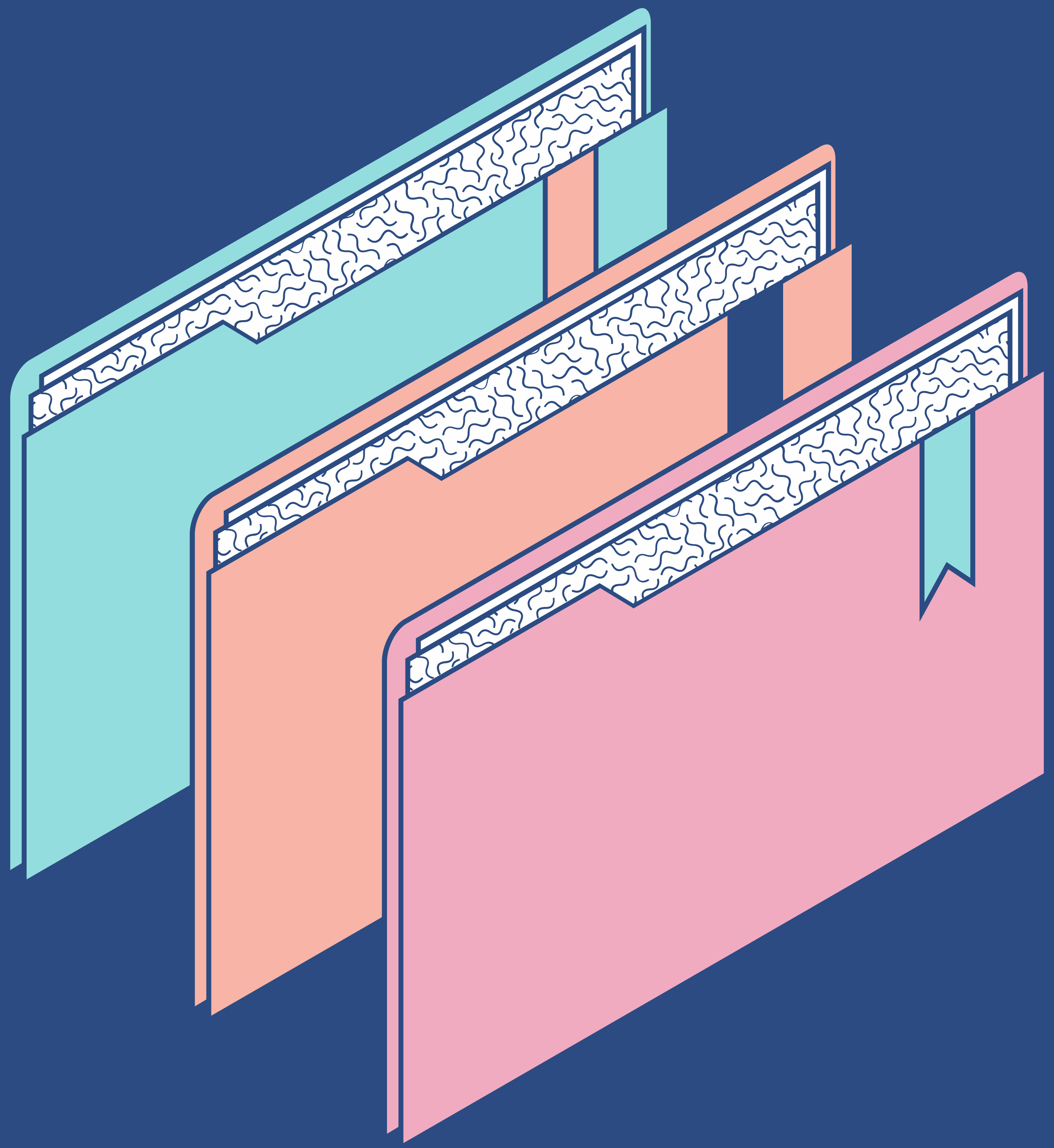
OUTPUT

- Images or videos containing human faces. These are real-time data provided to the system.

- Bounding boxes are drawn around faces to indicate their positions and sizes within the frame, each bounding box is accompanied by a label that represents the recognized emotion.







Part 2: Why we chose this problem?

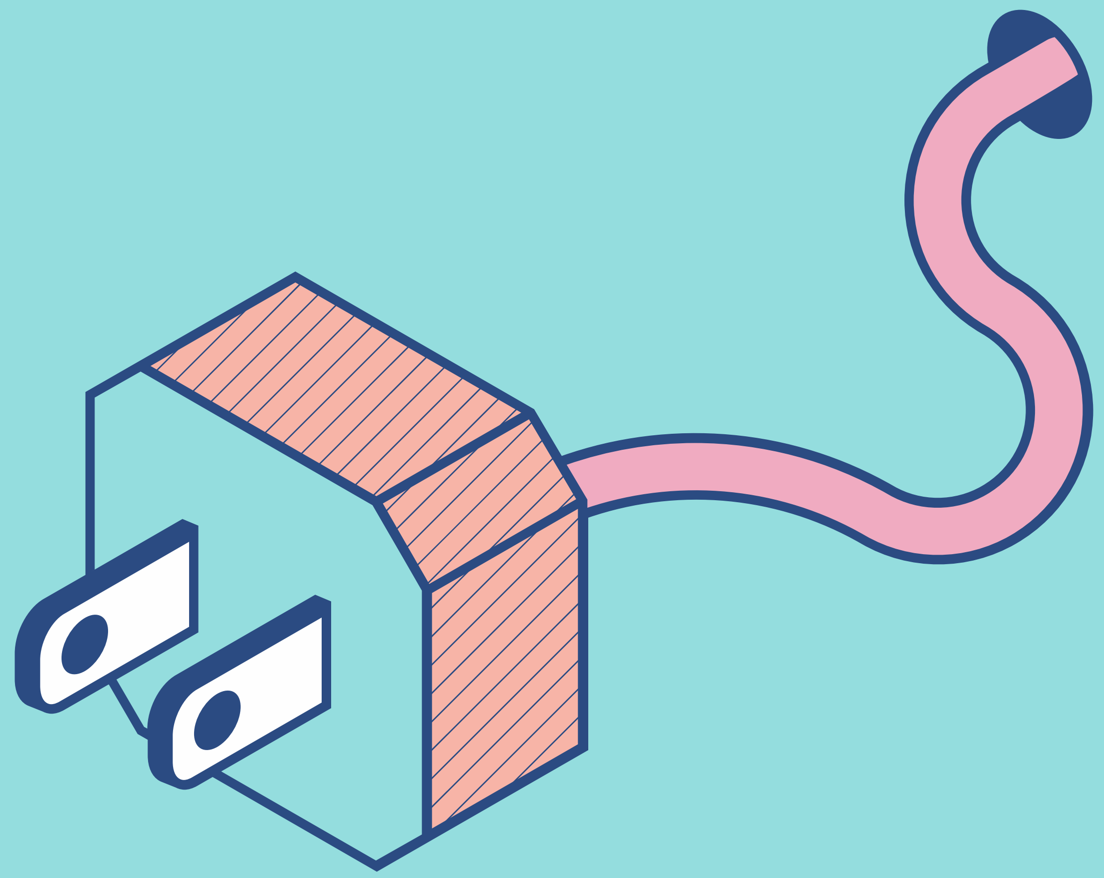
DETECTING EMOTION AT REAL-TIME

Benefits of Real-Time Emotion Detection

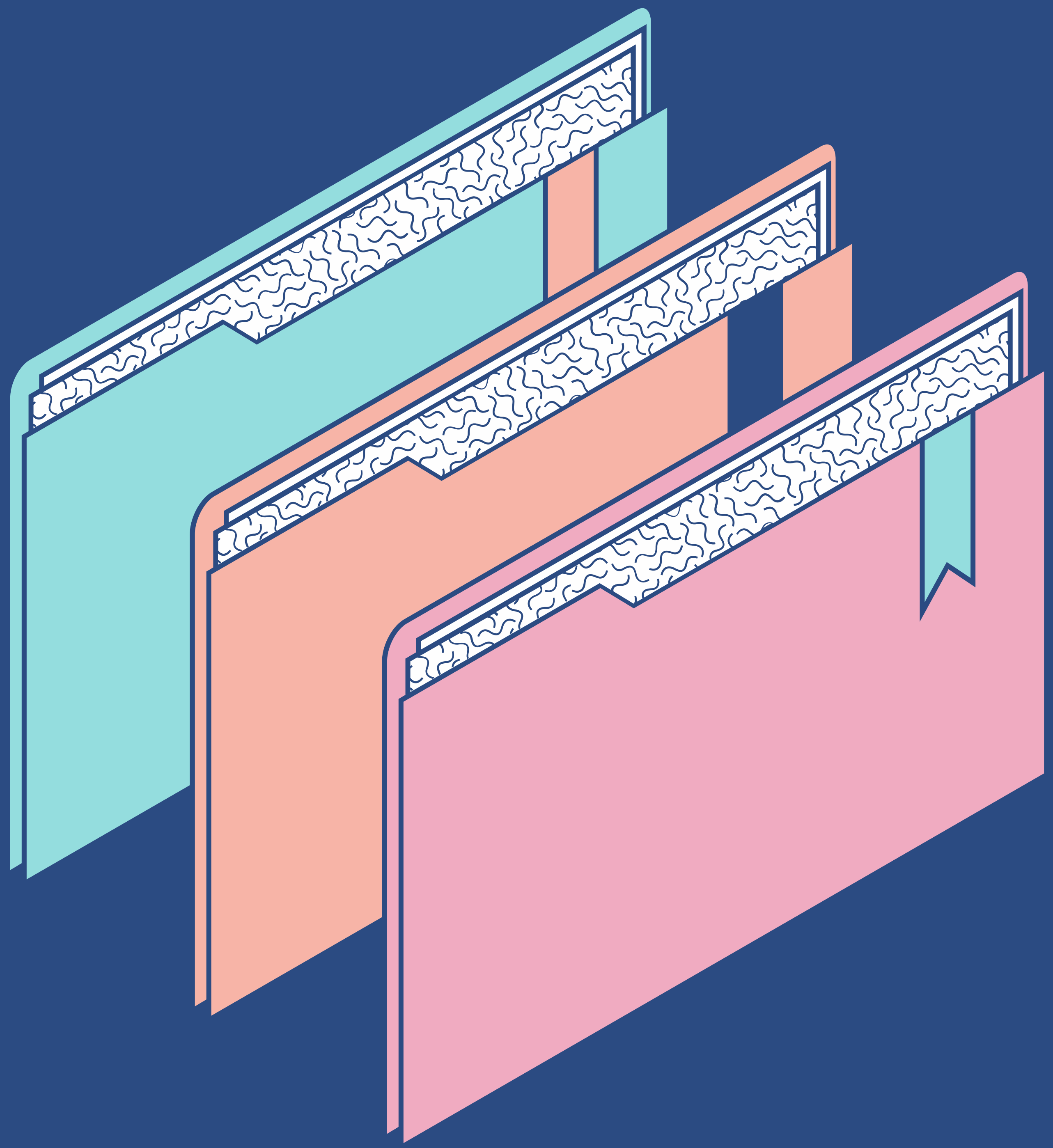
01	ENHANCED USER EXPERIENCE	improve user experiences in various applications
02	IMPROVED MENTAL HEALTH MONITORING	assist in monitoring and managing mental health conditions
03	MARKET RESEARCH AND ADVERTISING	provide valuable insights for market research and advertising campaigns
04	SOCIAL AND PSYCHOLOGICAL RESEARCH	provides researchers with valuable data for studying social dynamics, psychological processes
05	HUMAN-COMPUTER INTERACTION	crucial for building intelligent systems that can understand and respond to human emotions

AND MANY OTHER BENEFITS...

This topic not only provides specialized knowledge for computer science students but also contributes to innovative solutions and progress in the field of artificial intelligence and human-machine interaction.



SO WE DECIDED TO CHOOSE THIS!



Part 3: The method of facial emotion recognition

DETECTING EMOTION AT REAL-TIME

The History of CNN

WHO CAME UP WITH IT, WHO DEVELOPED IT, WHEN AND WHERE?

HELLO

1980s:

- Researchers Kunihiko Fukushima and Sei Miyake developed Neocognitron, a neural network model inspired by the visual cortex of animals. It was developed in Japan.

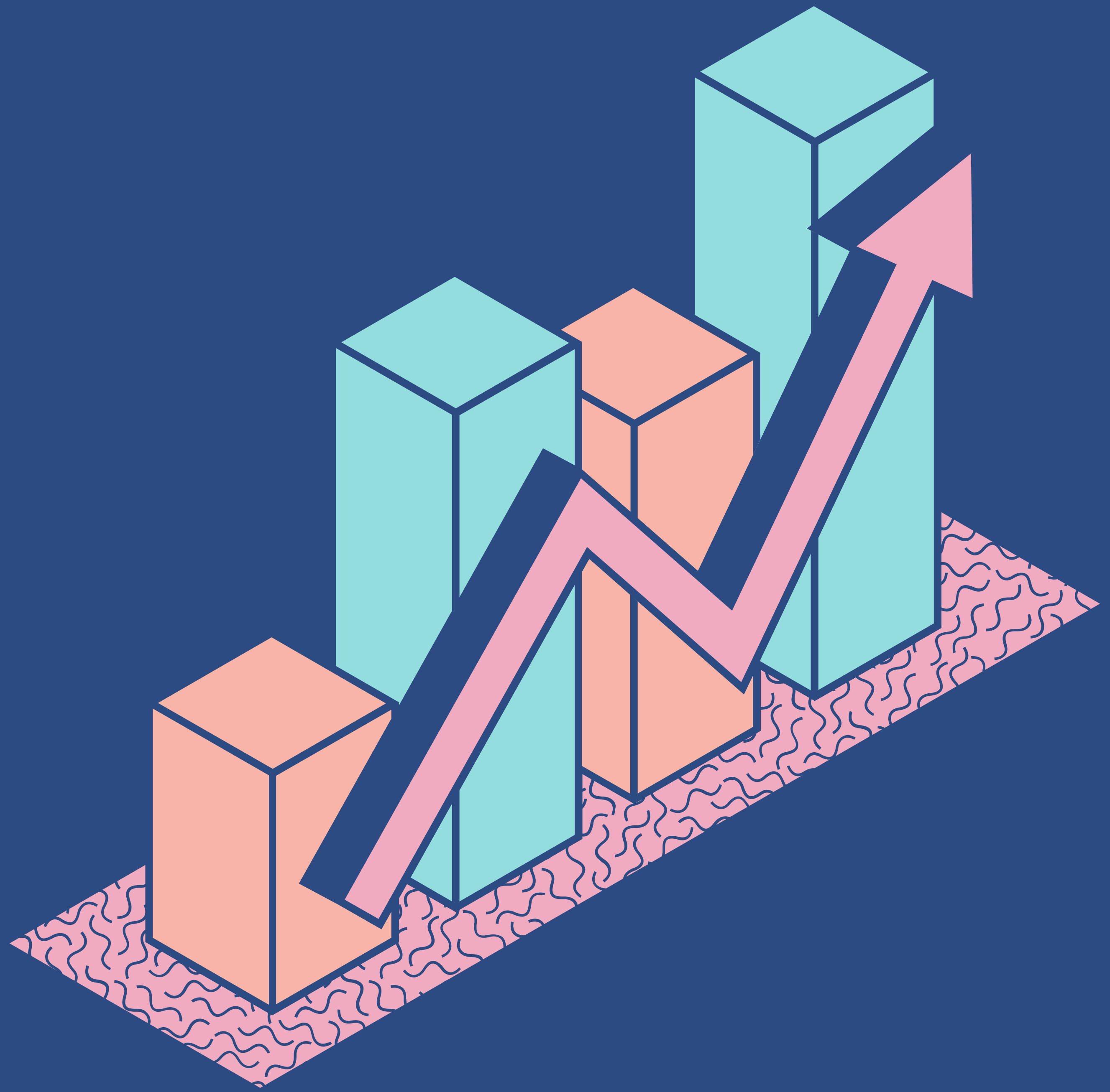
1990s:

- Yann LeCun, Yoshua Bengio, and Léon Bottou introduced Convolutional Neural Networks (CNNs) in 1998 while working at AT&T Bell Labs in the United States. They proposed the concept of CNNs for image recognition tasks, specifically handwritten digit recognition.

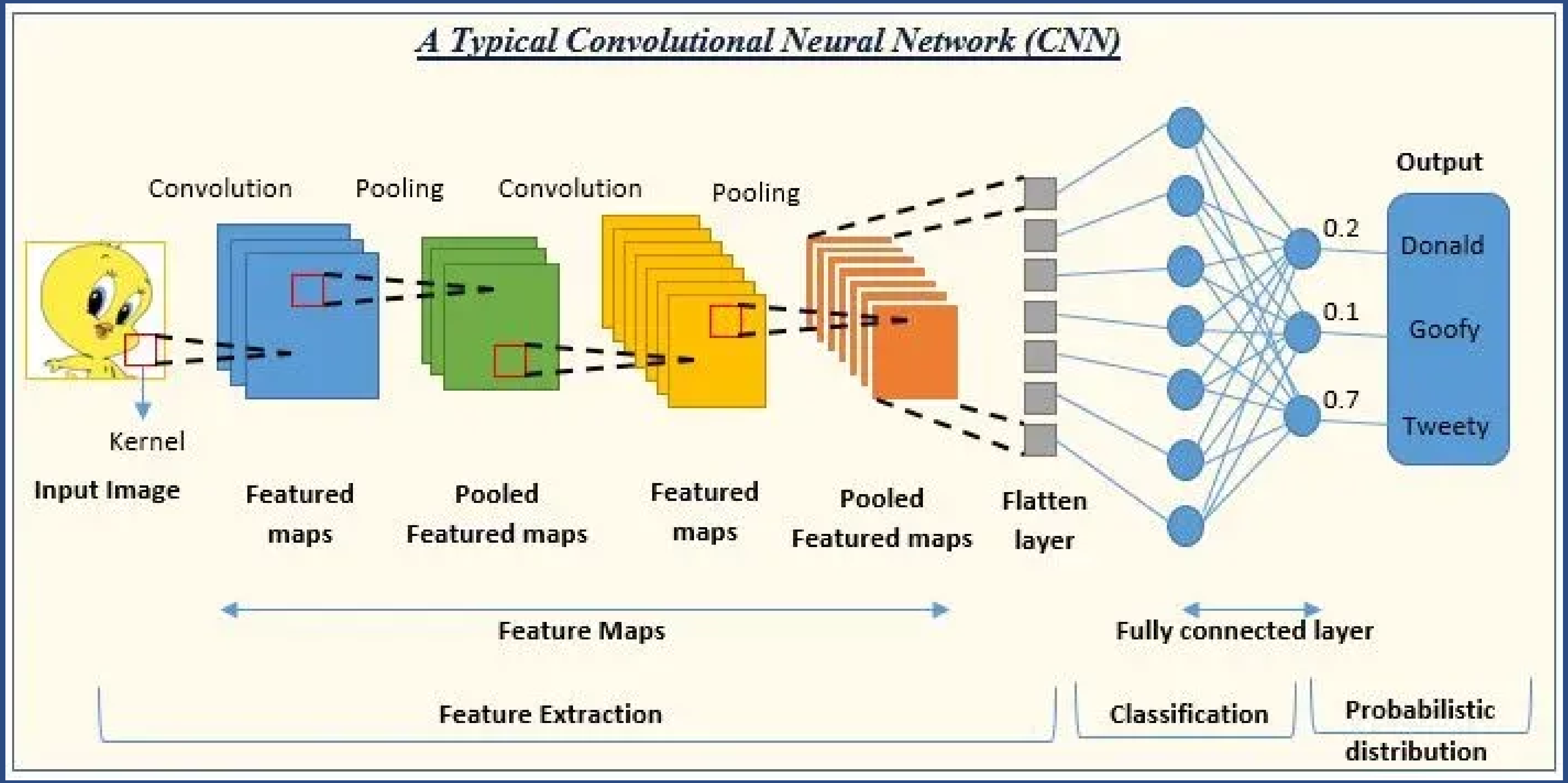
2000s:

- CNNs gained attention in the machine learning community but faced limitations due to the lack of computational power and large-scale datasets. However, research on CNNs continued, paving the way for future advancements.

Throughout the decades, CNNs have evolved from their origins in the 1980s to become a foundational technology in computer vision, driven by the contributions of researchers worldwide.



Basic architecture of CNN



CONVOLUTIONAL LAYERS

Convolutional layers apply filters to the input image, capturing different features and patterns.

POOLING LAYERS

Pooling layers downsample the feature maps, reducing the spatial dimensions while preserving important information.

FULLY CONNECTED LAYERS

Fully connected layers use the extracted features to classify or make predictions based on the learned representations.

CNN ARCHITECTURE IN OUR PROJECT

HOW DID WE APPLY AND DESIGN THE CNN STRUCTURE IN THE PROJECT?

Convolutional layers: 4

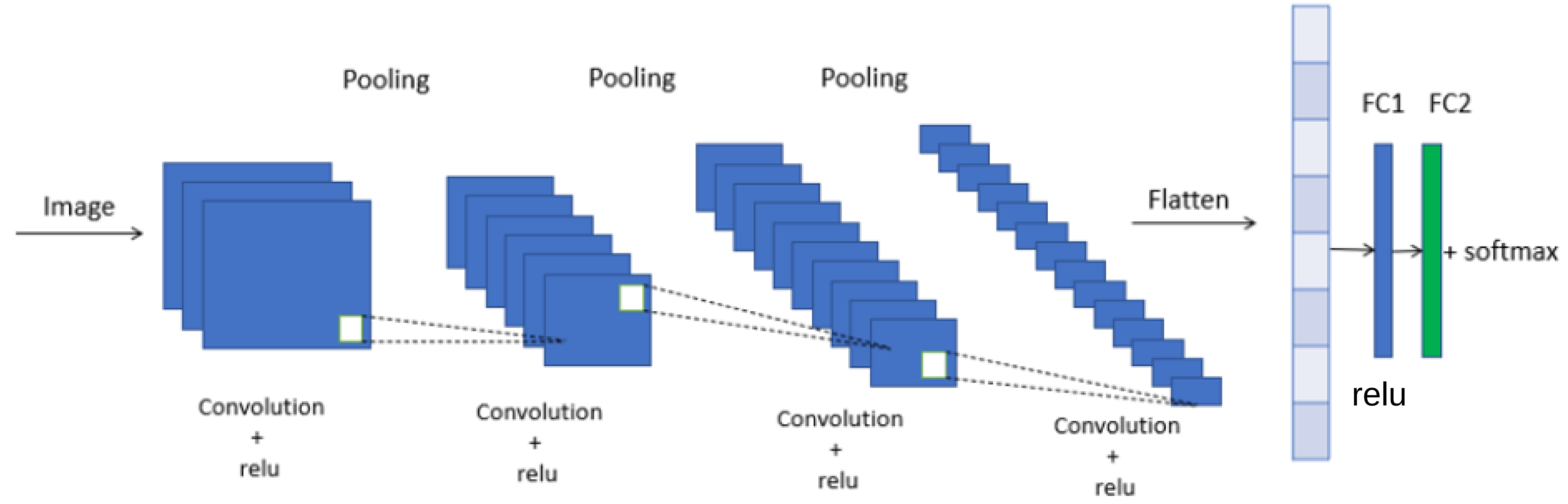
- The first convolutional layer has 32 filters.
- The second convolutional layer has 64 filters.
- The third and fourth convolutional layers both have 128 filters.

Pooling layers: 3

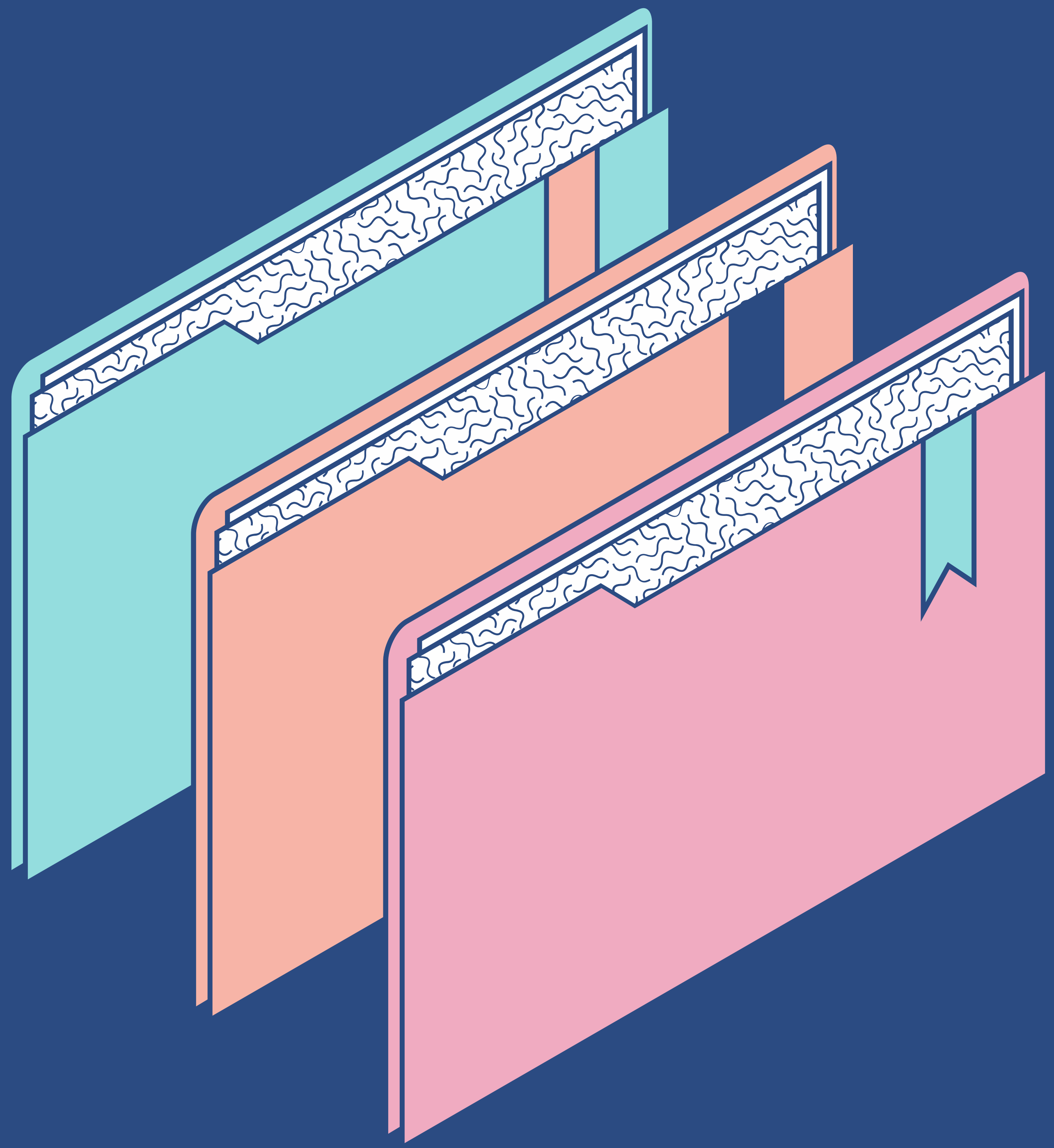
- The first max pooling layer has a pool size of (2, 2).-
- The second and third max pooling layers also have a pool size of (2, 2).

Fully connected layers: 2

- The first fully connected layer has 1024 units.
- The second fully connected layer (output layer) has 7 units, corresponding to the number of emotion categories.



Convolutional layers: 4
Pooling layers: 3
Fully connected layers: 2



Part 4: Experiment

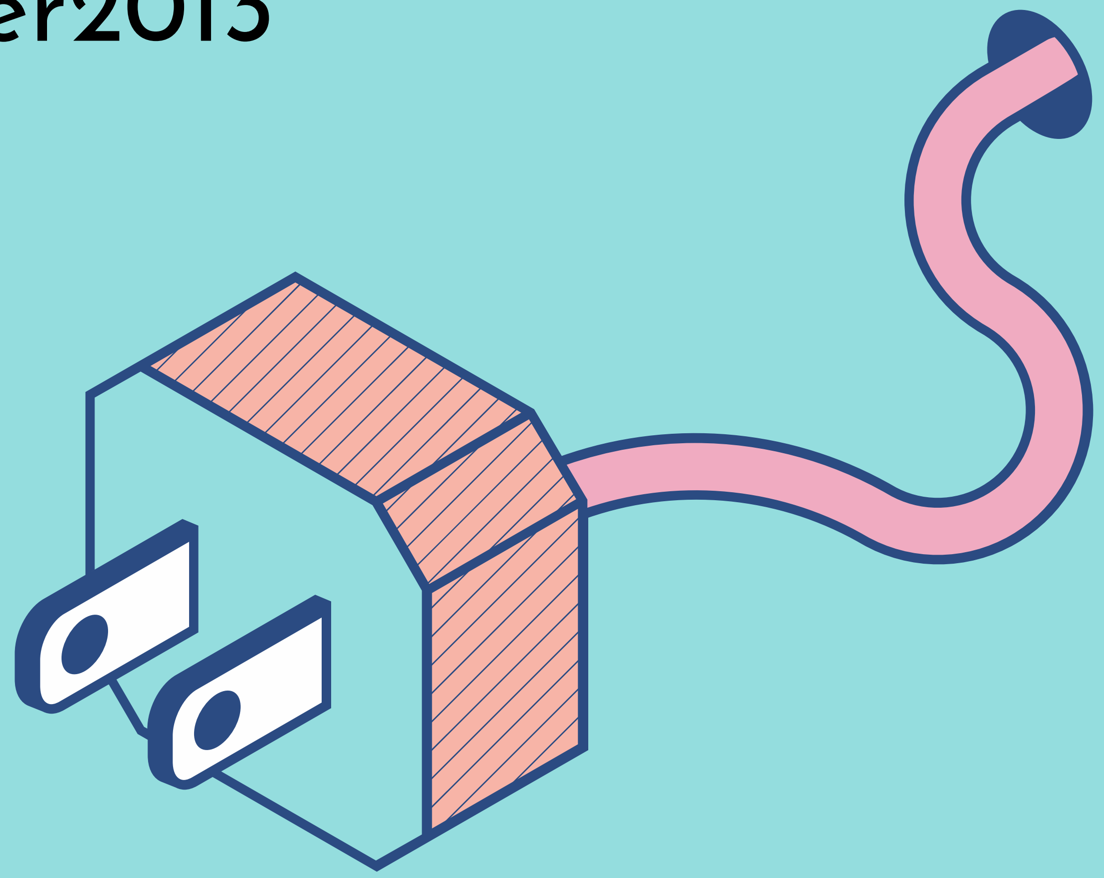
DETECTING EMOTION AT REAL-TIME

DATASET: FER - 2013 (KAGGLE)

https://www.kaggle.com/datasets/msambare/fer2013

FER-2013

Learn facial expressions from an image

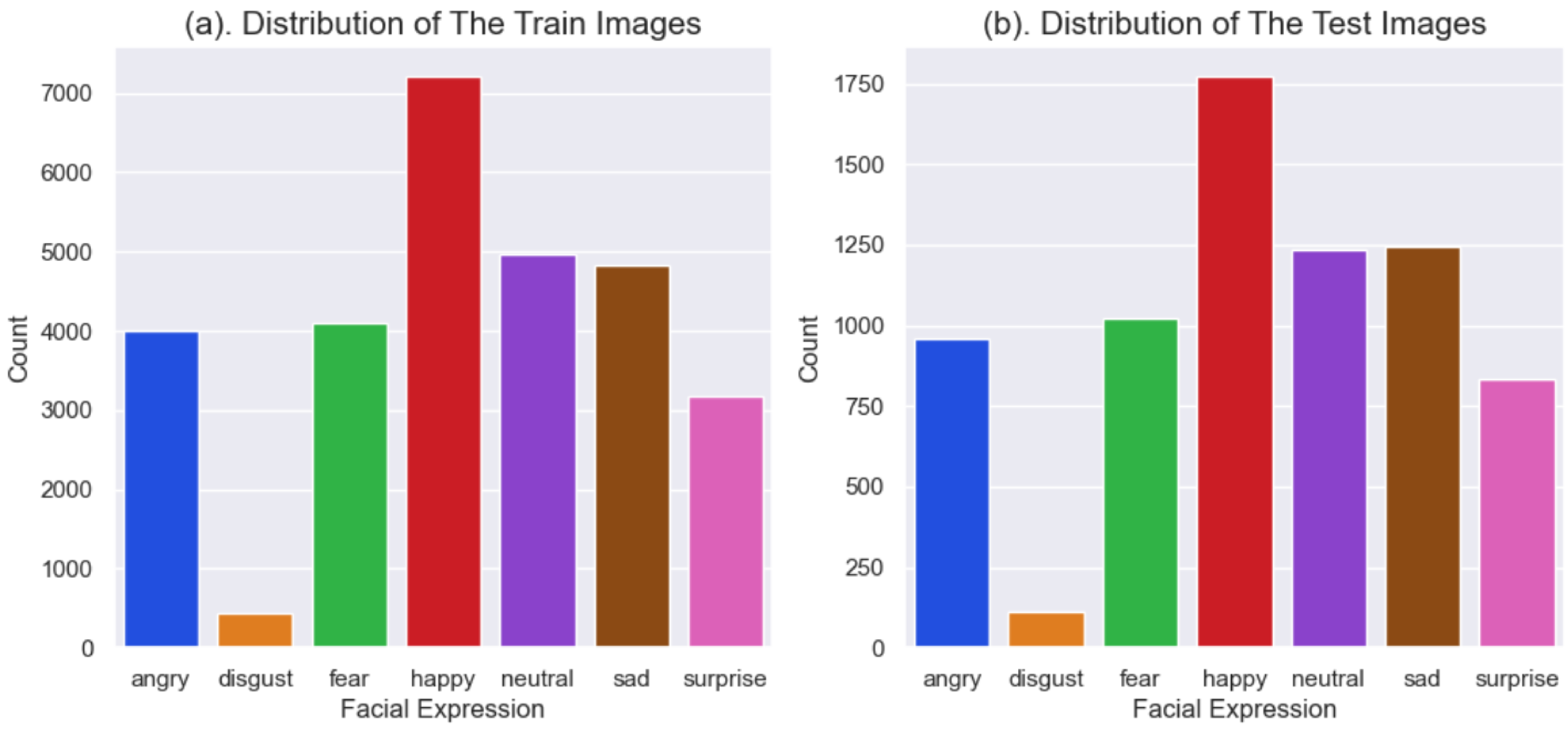


DATASET: FER - 2013 (KAGGLE)

About Dataset

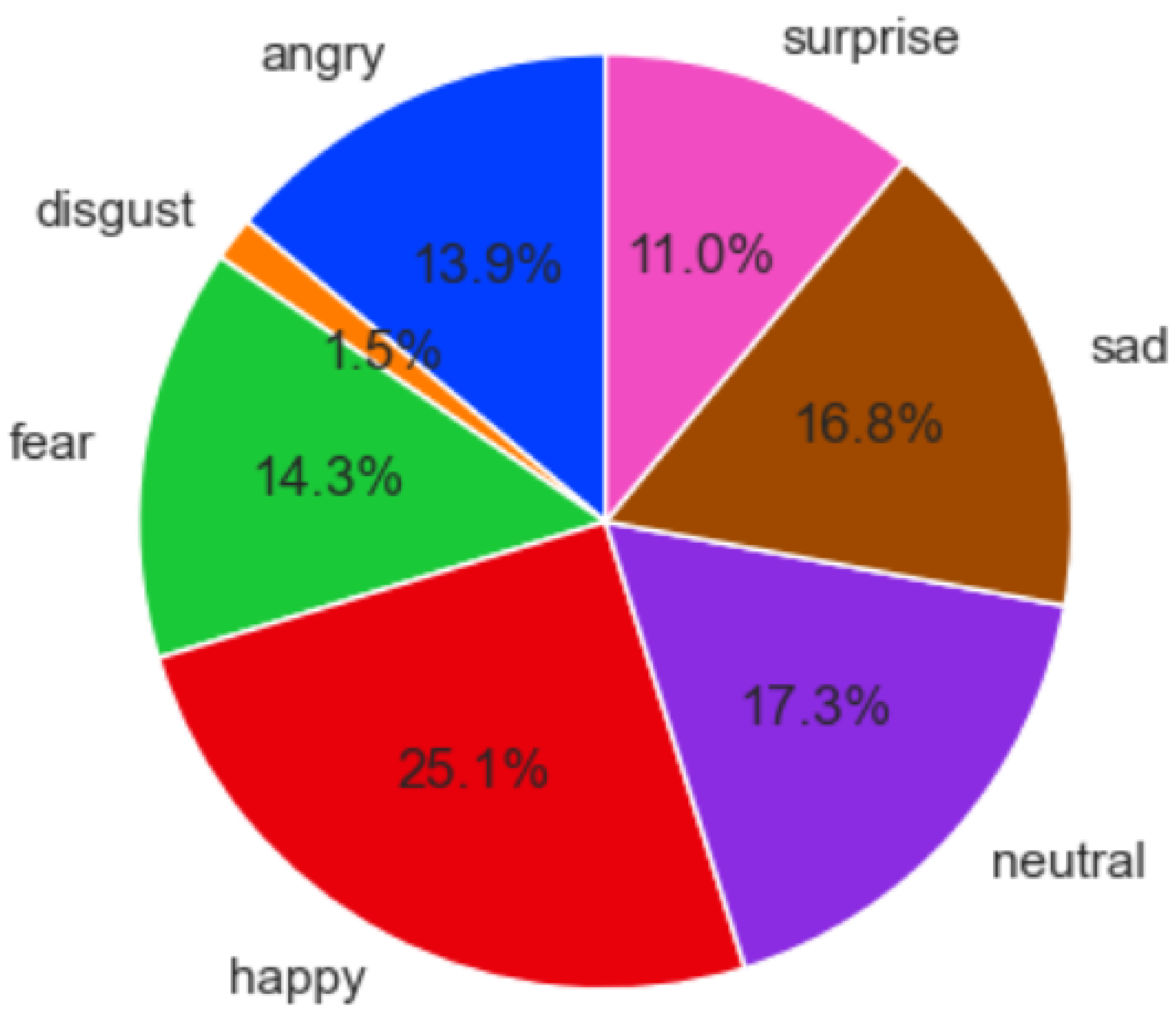
- The data consists of 48x48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centred and occupies about the same amount of space in each image.
- The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

DATA VISUALIZATION

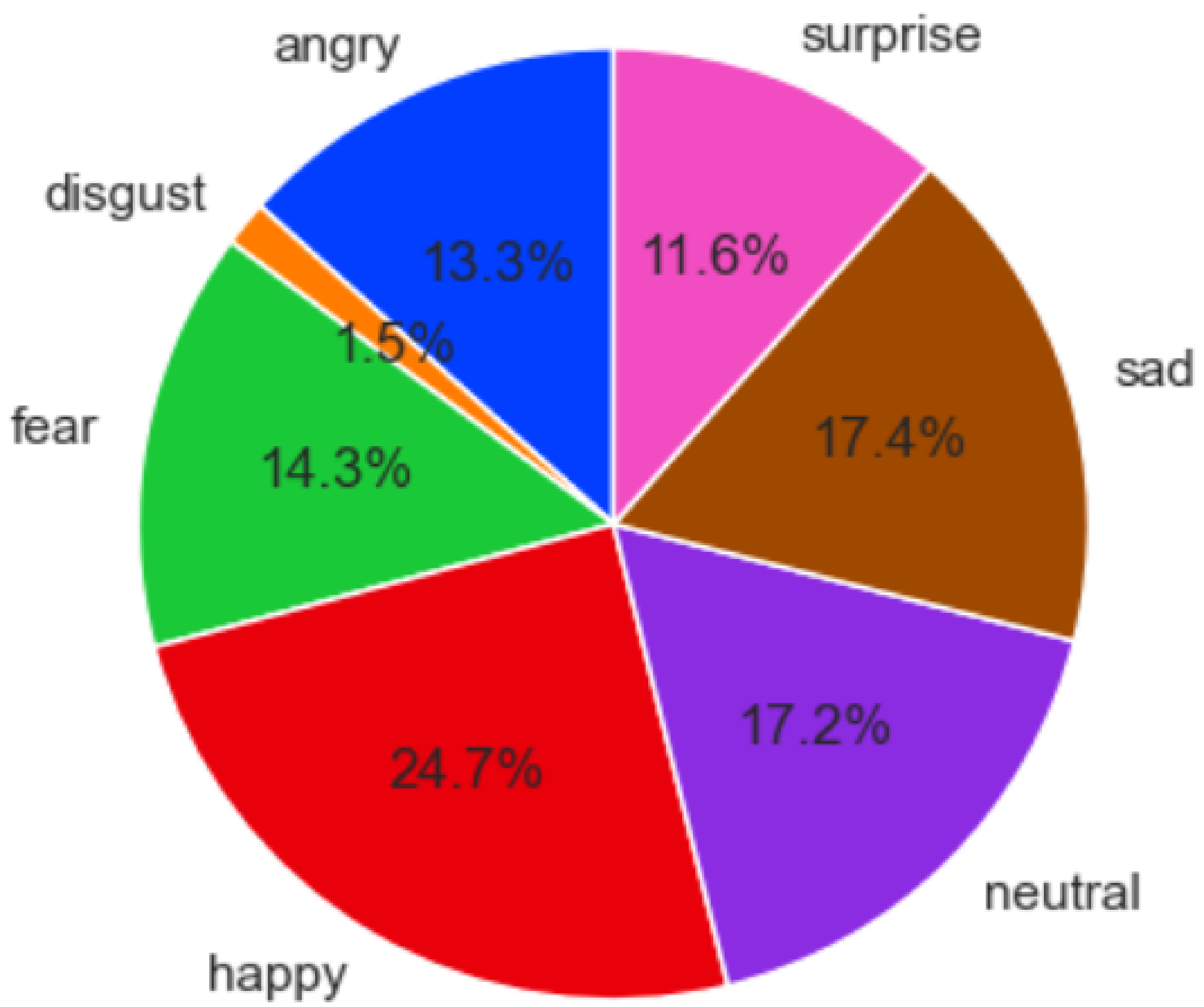


DATA VISUALIZATION

(a). Pie Plot of The Train Images

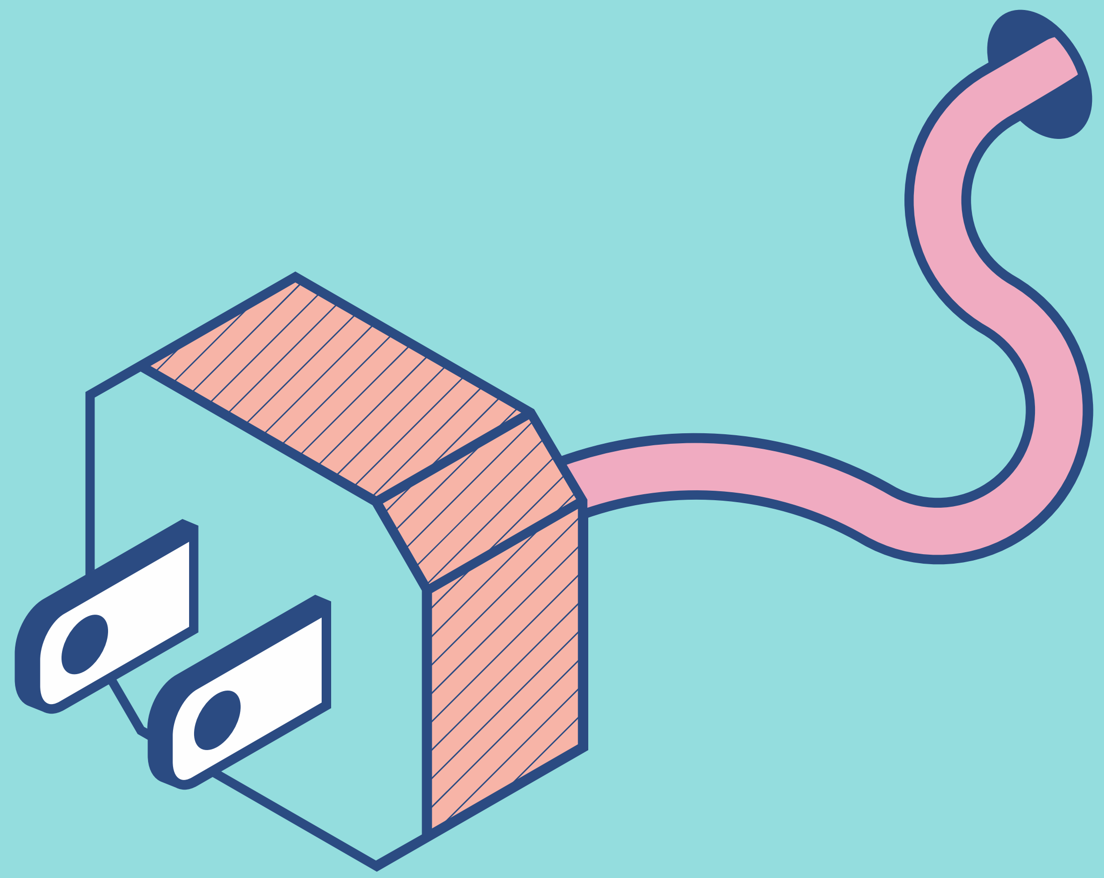


(b). Pie Plot of The Test Images

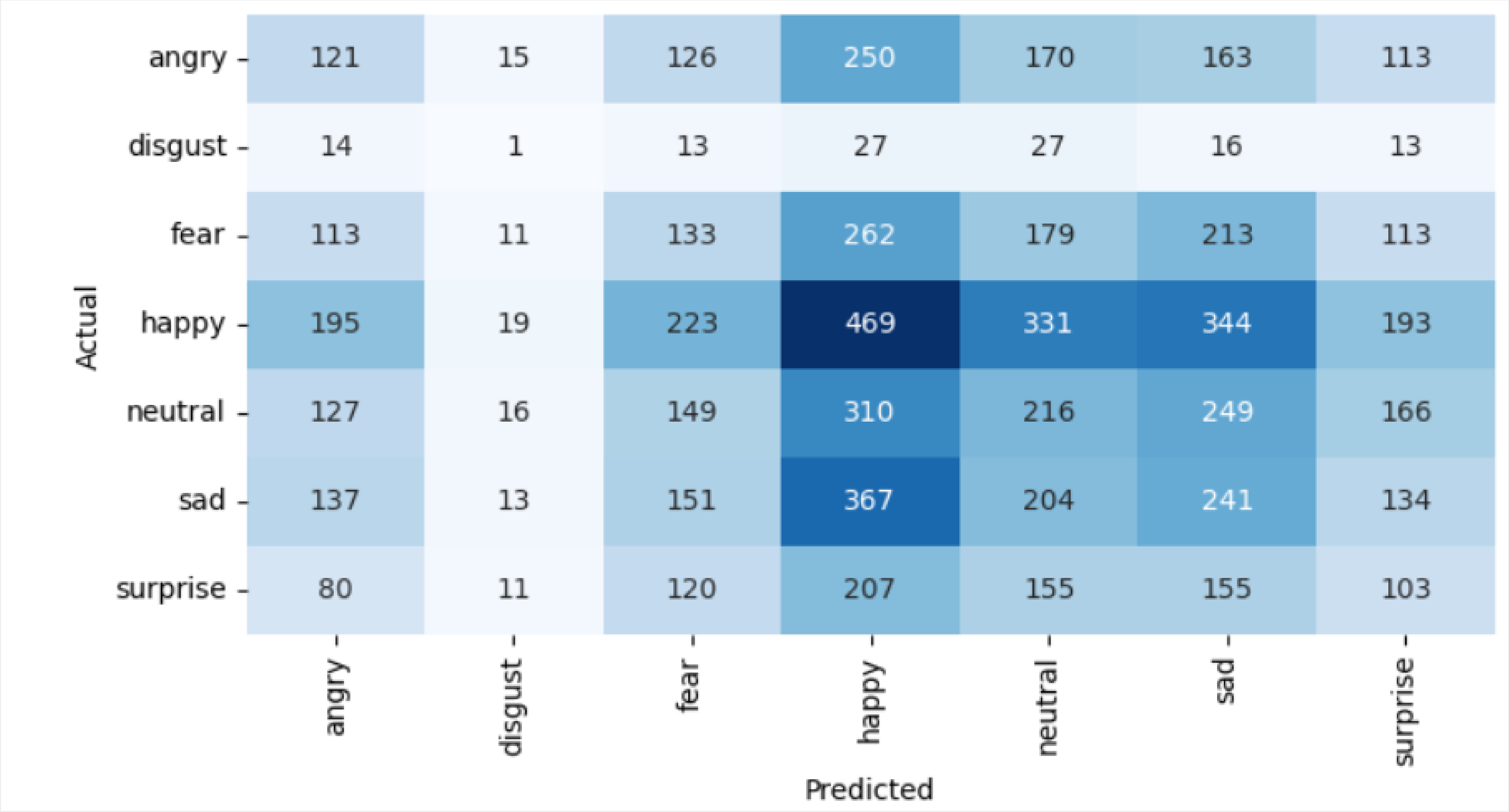


EVALUATION

Confusion Matrix on Validation Data and Accuracy



Confusion Matrix on Validation Data



Accuracy

```
# Evaluate the model
loss, acc = model.evaluate(train_generator, verbose=2)
```

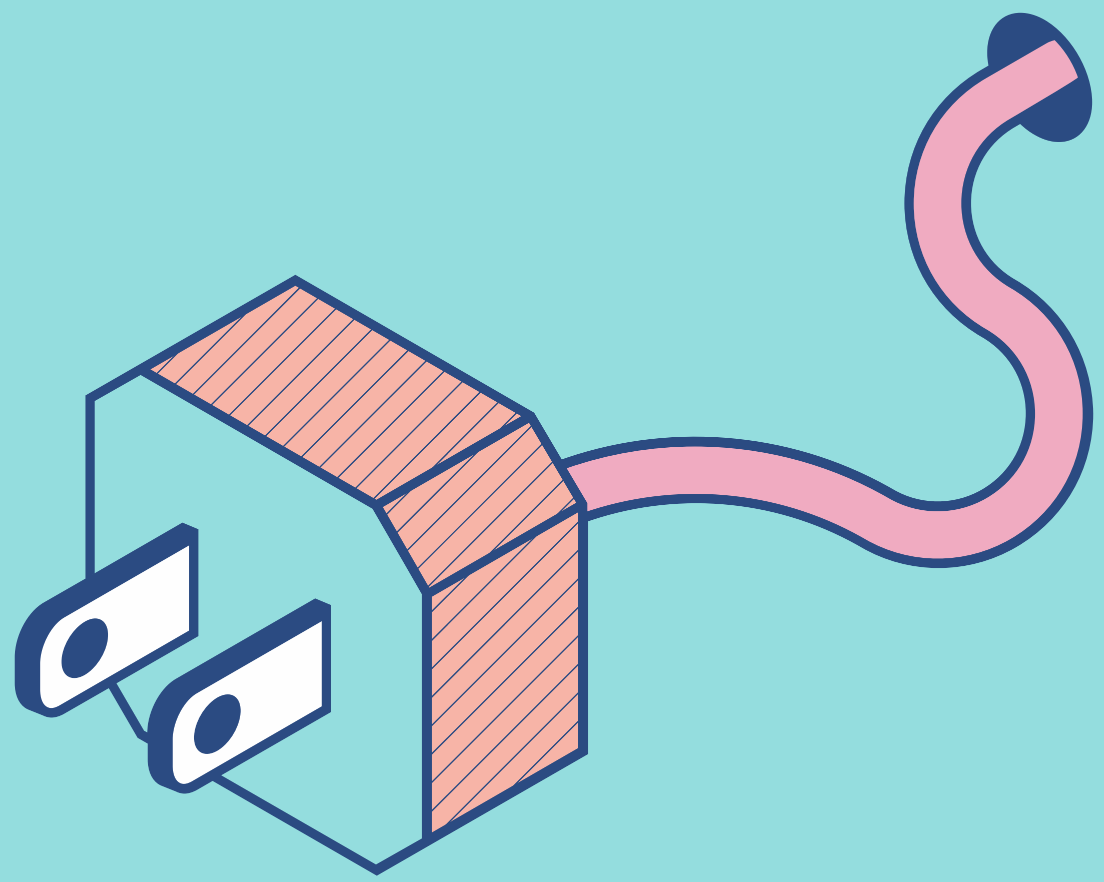
```
449/449 - 411s - loss: 0.1241 - accuracy: 0.9842 - 411s/epoch - 914ms/step
```

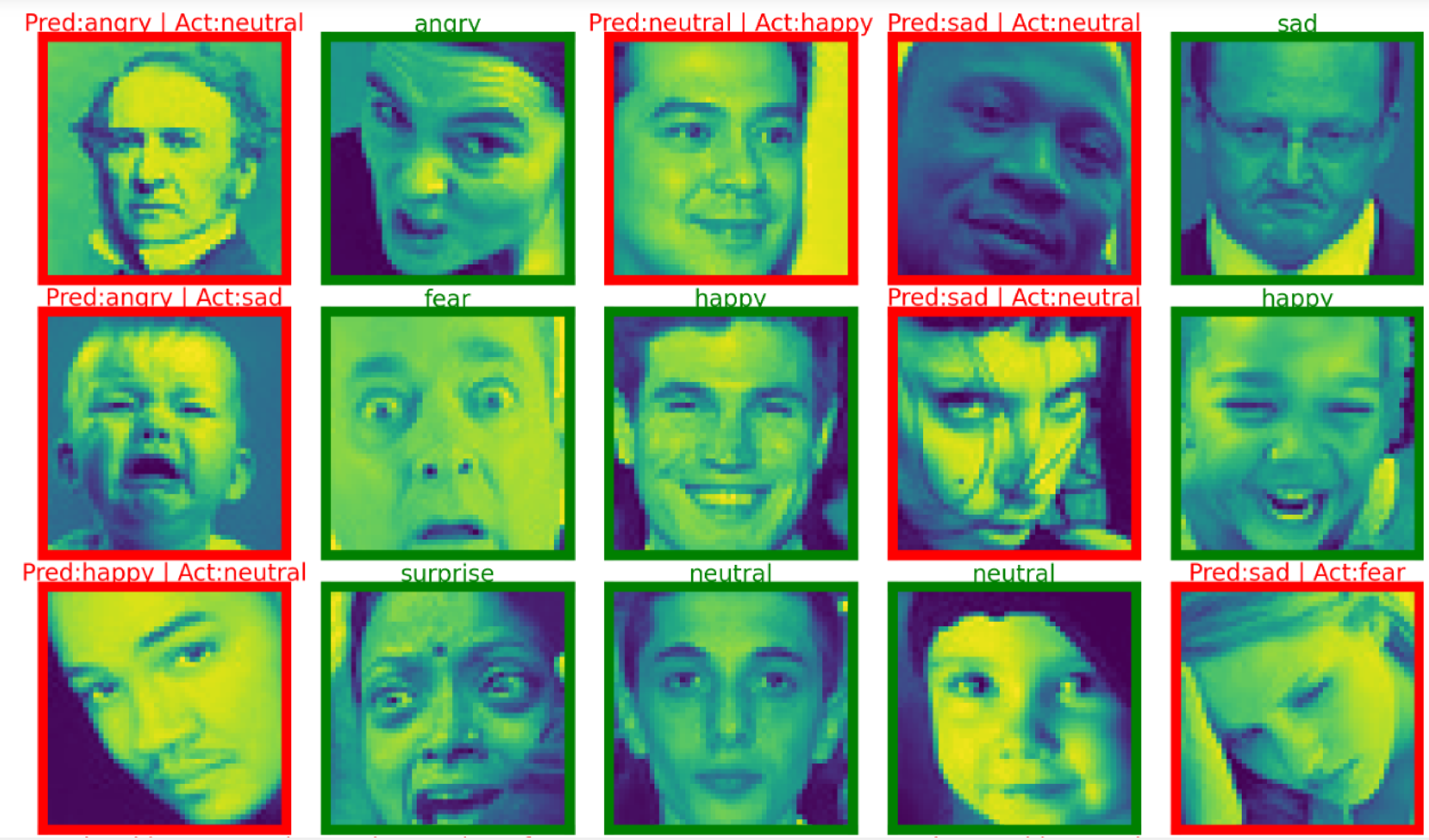
```
# Evaluate the model
loss, acc = model.evaluate(validation_generator, verbose=2)
```

```
113/113 - 95s - loss: 1.2282 - accuracy: 0.6229 - 95s/epoch - 837ms/step
```

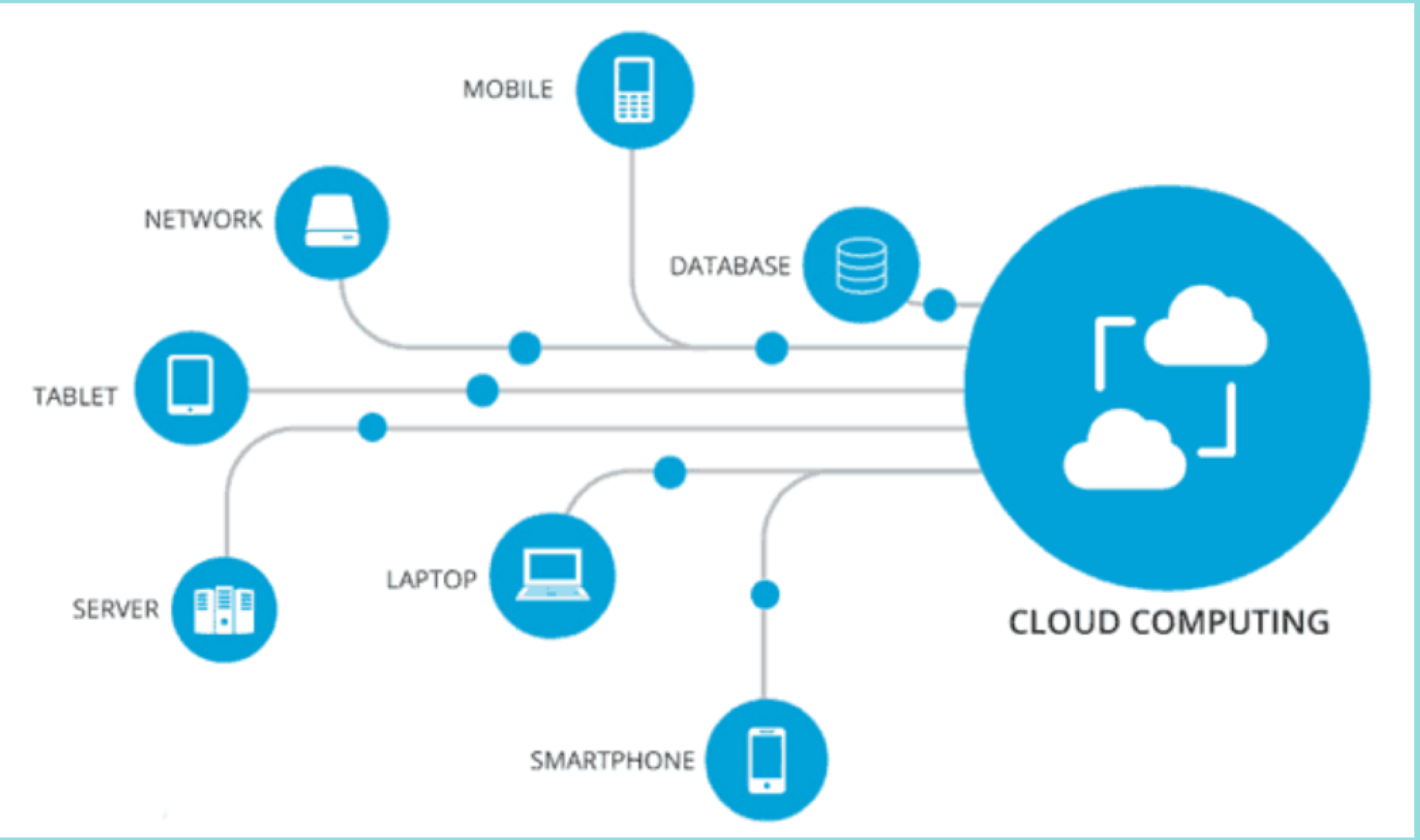
RESULT

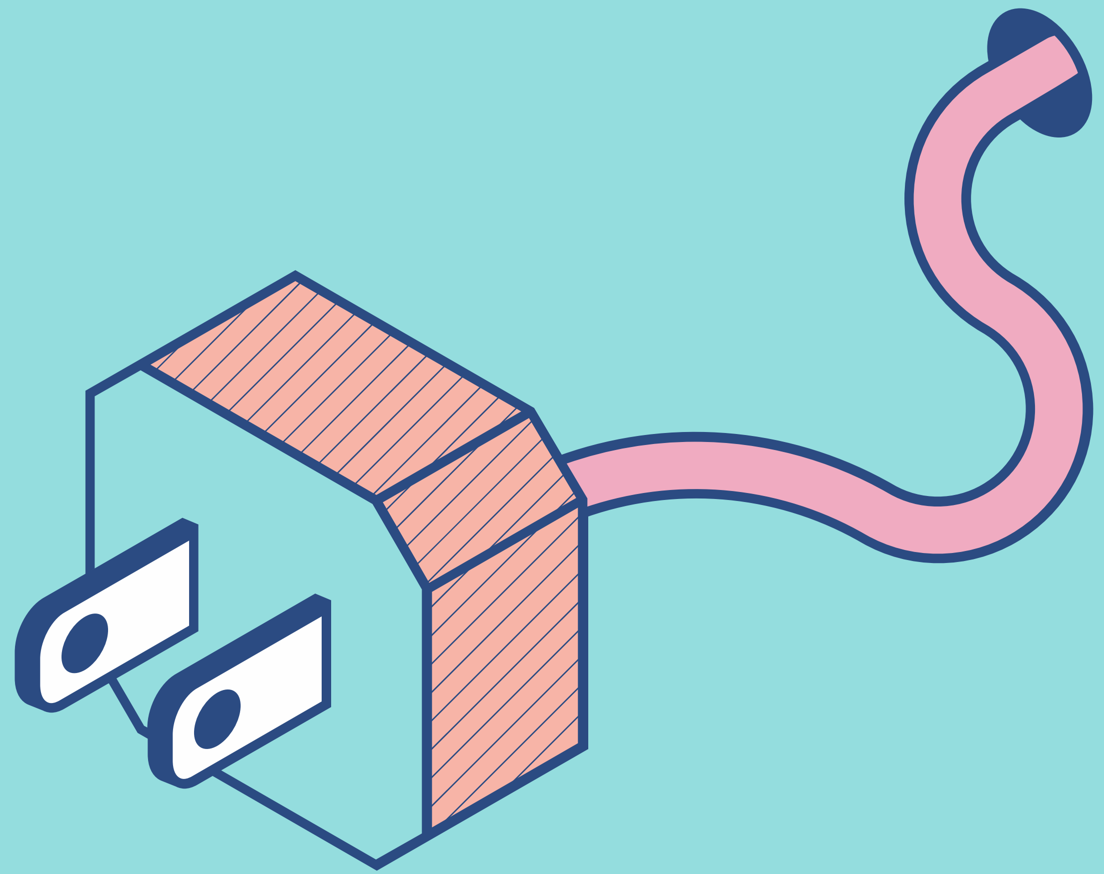
Print Predicted Images on Test Set




Pred:angry | Act:neutral
angry
Pred:neutral | Act:happy
Pred:sad | Act:neutral
sad
Pred:angry | Act:sad
fear
happy
Pred:sad | Act:neutral
happy
Pred:happy | Act:neutral
surprise
neutral
neutral
Pred:sad | Act:fear

REAL-TIME APPLICATION





Load and configure the emotion recognition model:

```
# Load emotion recognition model
model = model_from_json(open("model\\emotion_model.json", "r").read())
model.load_weights('model\\emotion_model.h5')
```

Use cascade classifier for face detection

```
# Load face detection cascade classifier
path_haarcascade = 'haarcascades\\haarcascade_frontalface_default.xml'
face_haar_cascade = cv2.CascadeClassifier(path_haarcascade)
```

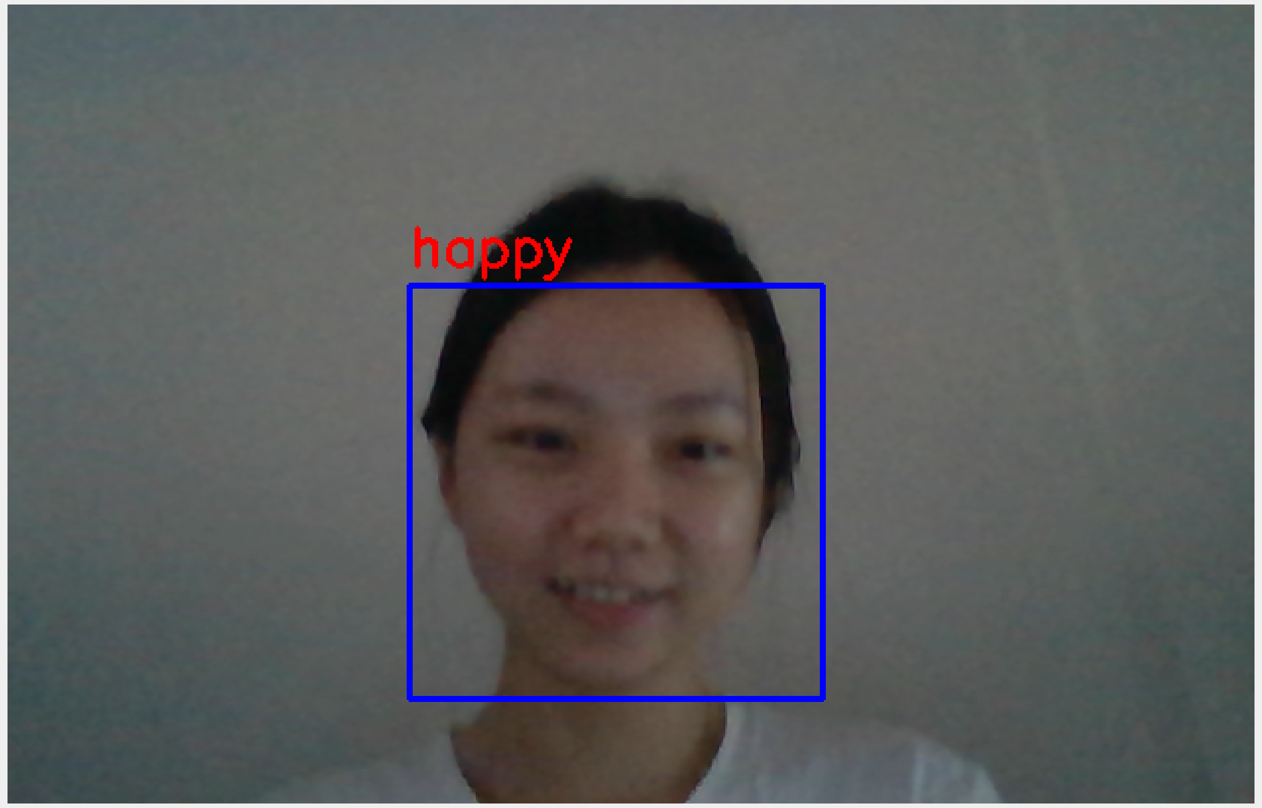
Process each detected face

+ Extracting the Region of Interest (ROI)

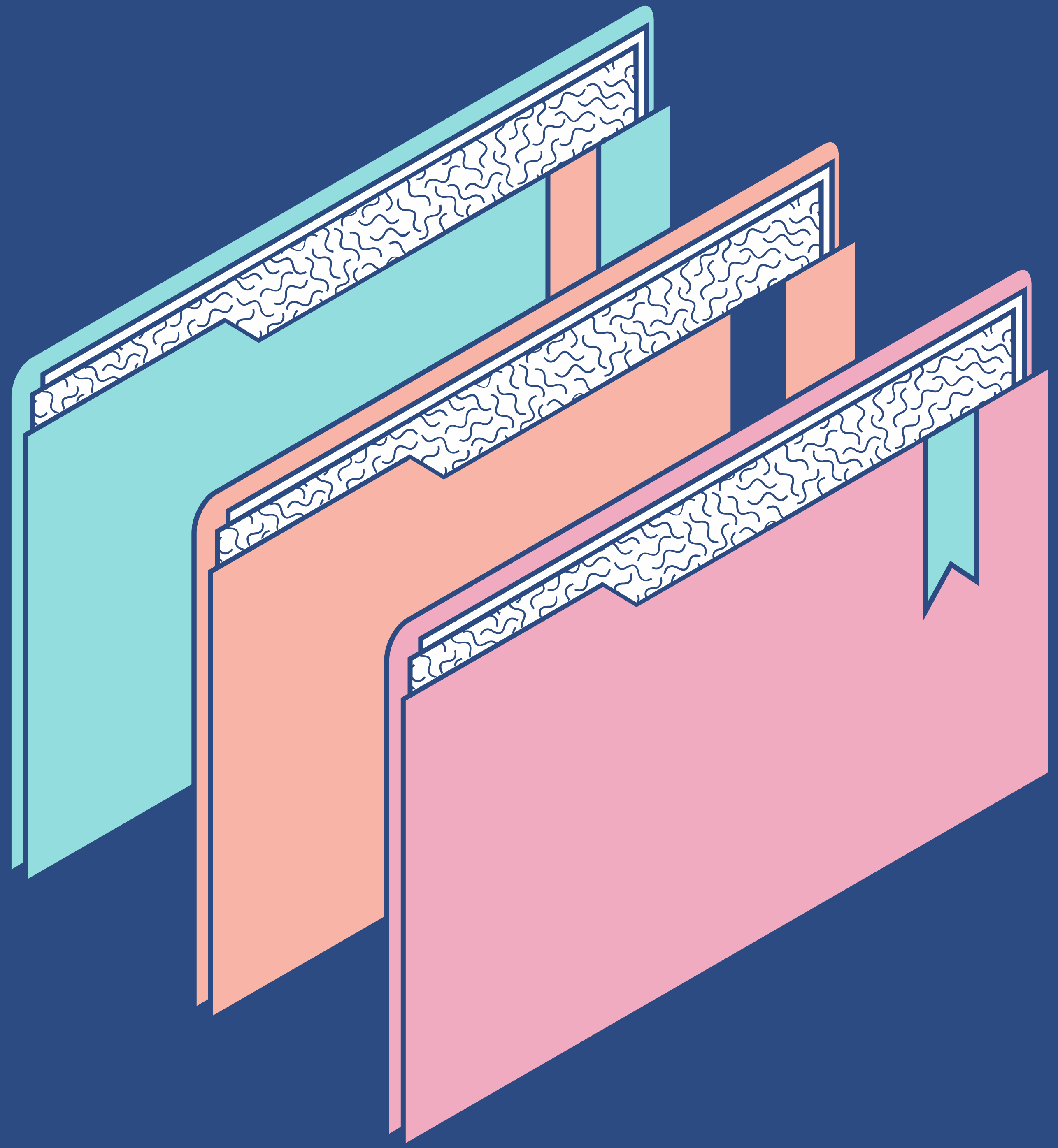
+ Resizing the ROI

+ Normalizing pixel values.

Predict emotions and display the results on the face



Predict emotions and display the results on the input image



Part 5: Conclusion

DETECTING EMOTION AT REAL-TIME

SUMMARY

Convolutional Neural Networks (CNN) for facial expression recognition

Training on labelled emotion dataset

Real-time face detection using cascade classifier

Extracting Region of Interest (ROI) and preprocessing

Displaying predicted emotion on the detected face

SEVERAL CHALLENGES

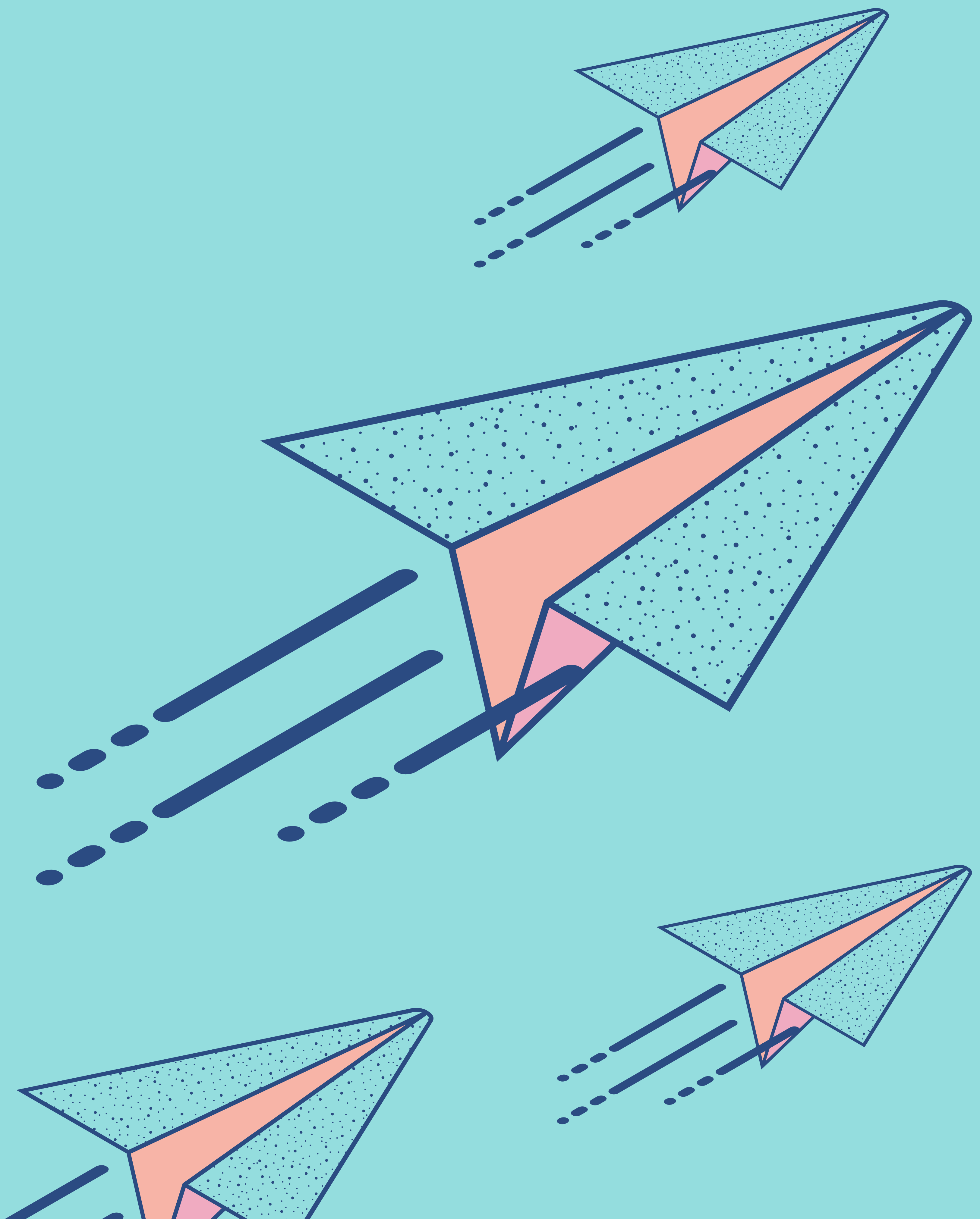
+ Variability in facial expressions
+ Environmental factors
+ Complex emotions
+ Real-time processing

IMPROVEMENT DIRECTION

+ Dataset augmentation
+ Other advanced deep learning architectures
+ Feature engineering
+ Continuous model monitoring and updates

Question Time

Give it to us! Hope you and us learned something new.



THANK YOU!

CNN

CNN: What are advantages of CNN in Emotion Detection?

SPATIAL UNDERSTANDING

CNNs capture spatial relationships and patterns in facial expressions .

HIERARCHICAL FEATURE LEARNING

CNNs learn hierarchical representations of features, enabling them to extract discriminative emotional features from facial expressions

ROBUSTNESS TO VARIATIONS

CNNs are robust to variations in pose , lighting conditions , and facial characteristics , making them suitable for handling diverse facial expressions .

TRANSFER LEARNING

Pre-trained CNN models can be leveraged for emotion recognition , reducing the need for extensive training data and accelerating model development .

SCALABILITY

CNN architectures can be scaled up or down to adapt to different complexity requirements