**Emotion Detection Using Deep Learning**

**G Kundana Sree- E21CSEU0313, Divisha Gandhi-E21CSEU0296**

**Abstract:**

An essential yet complex area of research in social communication is the recognition of human emotions from images. Beyond conventional image processing techniques, deep learning techniques have grown into highly efficient instruments in this field. An artificial intelligence system that can identify emotions through facial expressions is created and demonstrated in this study. It investigates the emotion detection process, which consists of three main steps: emotion classification, feature extraction, and face detection. In this article, a novel convolutional neural network (CNN)-based deep learning architecture designed especially for emotion recognition in photos is presented.

This project tries to explore how computers can recognize people’s emotions from their facial expressions. Additionally, the project will apply advanced techniques called deep learning with a particular focus on two popular methods – VGG16 and ResNet. Therefore, we aim to design a smart system capable of accurately predicting people’s emotions based on their natural images with faces. We will then analyse the methods using a dataset of multiple pictures with faces and hopefully obtain a feasible system that can distinguish people’s emotions in real situations. This could be very beneficial in several areas, such as healthcare, technology and marketing.

**Introduction:**

As humans, our faces are like windows to our emotions, displaying a rich array of feelings ranging from happiness and astonishment to sorrow and fury.

Numerous efforts have been made to train computers to recognize emotions from facial images. Typically, large image datasets representing a range of emotional states are fed into convolutional neural networks (conv nets). However, it's crucial to acknowledge the limitations of convolutional nets. When tackling such problems with machine learning systems, it's essential to consider human cognitive processes. For example, on the fer2013 dataset, the maximum accuracy reached is 60%. How to increase this accuracy to 90% is the subject of this article. We will dive into the intricate details of building a convolutional neural network and training a machine learning model using OpenCV as we explore the fascinating field of real-time facial emotion recognition in this article. Along the way, we'll uncover the profound implications of this technology, particularly its role in improving interactions between humans and computers, as well as fostering emotional intelligence.

What if computers could read our emotions by looking at our faces? That is our goal with the project. The reason behind this Introduction is to create a safer and more rewarding field in detecting emotions from facial expressions, making a better computer, knowing what a person is feeling, and ensuring healthcare. Although some old methods can detect facial emotions, they have high error rates. For this reason, we turn to deep learning, smart technology inspired by the way our minds function. The two deep learning methods we will look into in this project are VGG16 and ResNet. These methods have the potential to detect a pattern from a picture, whether a facial expression or anything else.

**Problem Statement:**

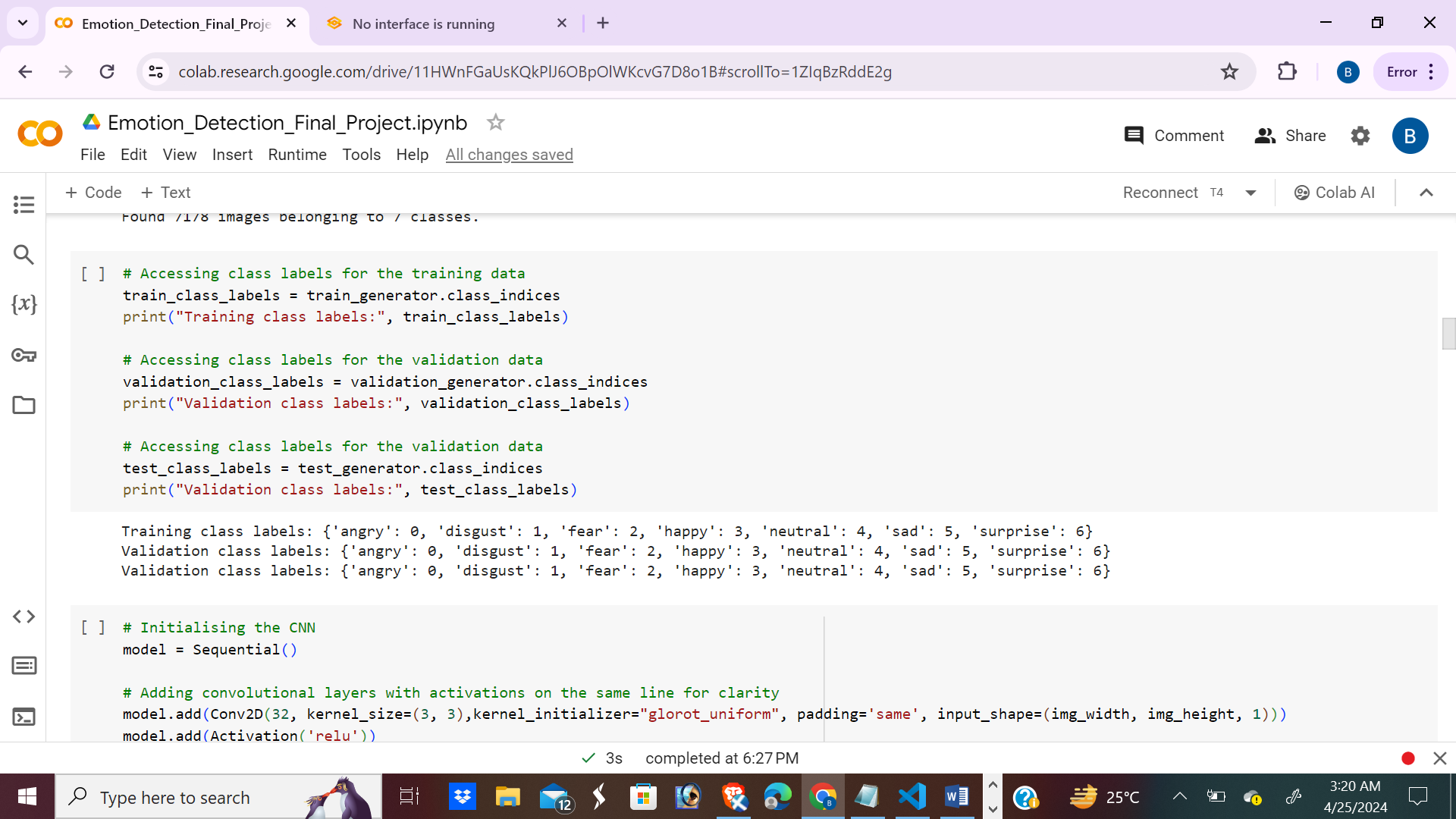
Conventional methods for detecting emotions often suffer from inaccuracies and limited scalability. This project seeks to overcome these limitations by leveraging deep learning techniques to develop robust emotion recognition systems. We delve into different patterns from diverse data sources and we aim to create automated systems capable of accurately understanding and responding to human emotions. This endeavour holds promise for advancing fields like sentiment analysis and human-computer interaction.

The challenge of our project is to build a system that can detect emotions in a face in a computer with high accuracy. Furthermore, this system must work in a wide spectrum of situations and faces. Faces may have been captured with different lighting or be at different angles for example. Our system should also be able to work even if faces are captured with these variations. We also want to compare our two models VGG16 and ResNet, thus we will run several tests to see which model is better for this task. For instance, we will test how many cases they predict correctly and how well they are at detecting different faces.

**Dataset:**

The project makes use of the FER2013 dataset, which was sourced from Kaggle, specifically the Mood Dataset. This dataset is valuable as it contains a wide range of facial expressions categorized into seven primary emotions: Happy, Angry, Sad, Neutral, Fear, Surprised, and Disgust.

The data is preprocessed to ensure the model's accuracy in real-world situations. In this phase, pixel values are used to adjust the size of the images. Further, numerical labels are mapped to corresponding emotions using a dictionary. This approach is frequently used in machine learning, particularly for tasks involving sentiment analysis and emotion recognition. The model's training and assessment are made easier by the association between each numerical label and a particular emotion.



**Load Data:**

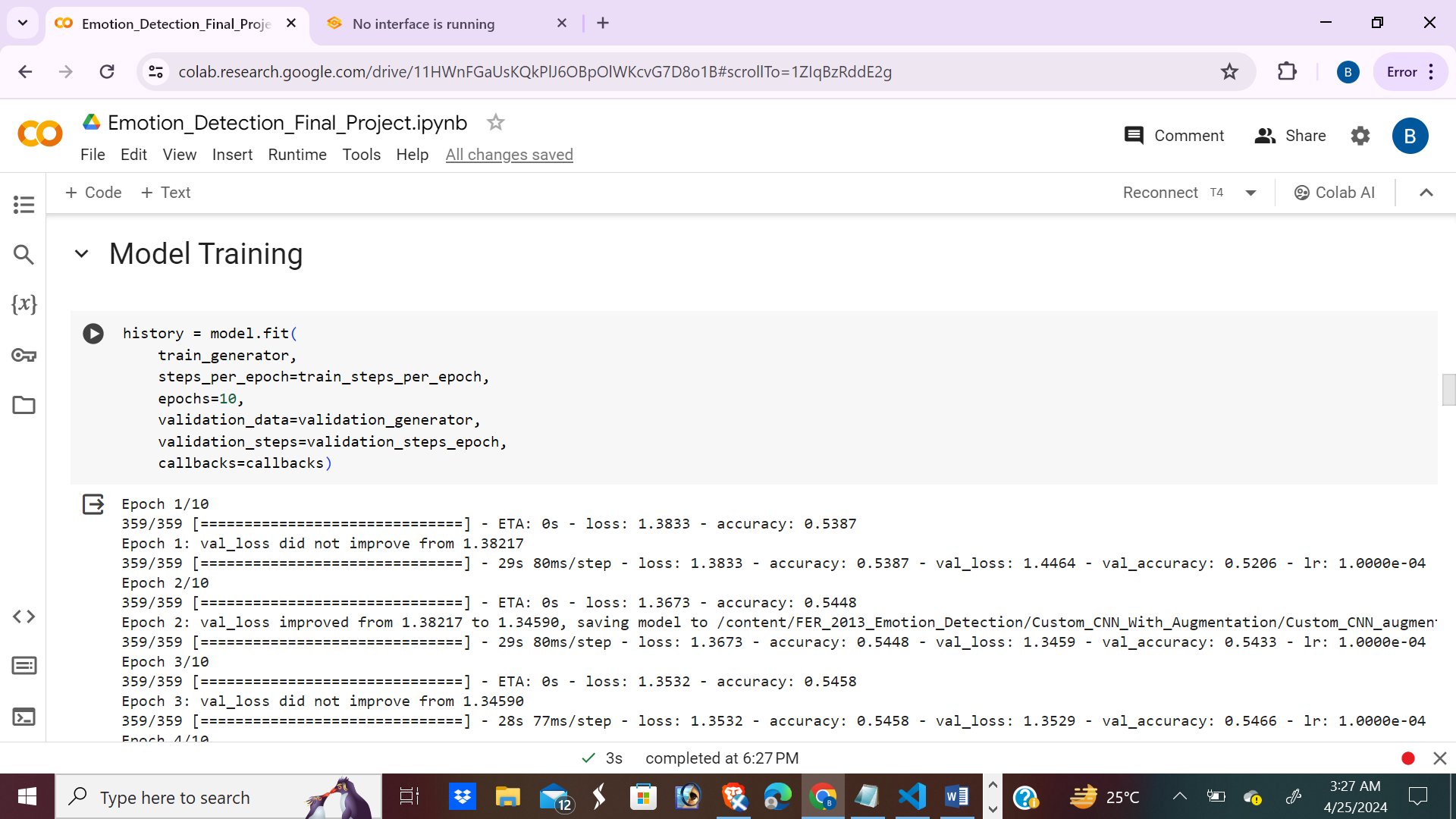
Let's delve into the process of importing our dataset.

To begin, we specify the path to our dataset. Then, we organize the dataset into seven subfolders, each representing a distinct category. We iterate through the main folder to access every subfolder within our list of subfolders, aiming to read the images contained within.

During this process, we extract the index of each folder and assign it to a label. Following a predefined logic discussed earlier in our planning phase, we transform these labels into new ones.

Iterating through the subfolders, we read images one by one. Each image undergoes a transformation, being converted to grayscale and resized to dimensions of 48x48 pixels. As we progress, we append the pixel values of each image to a list named "images" and their corresponding new labels to another list called "new\_labels".

Moving forward, we convert both lists, "images" and "new\_labels", into numpy arrays. Subsequently, we encode the labels into three distinct classes, a crucial step prerequisite for our convolutional neural network (CNN) implementation.



**CNN Model:**

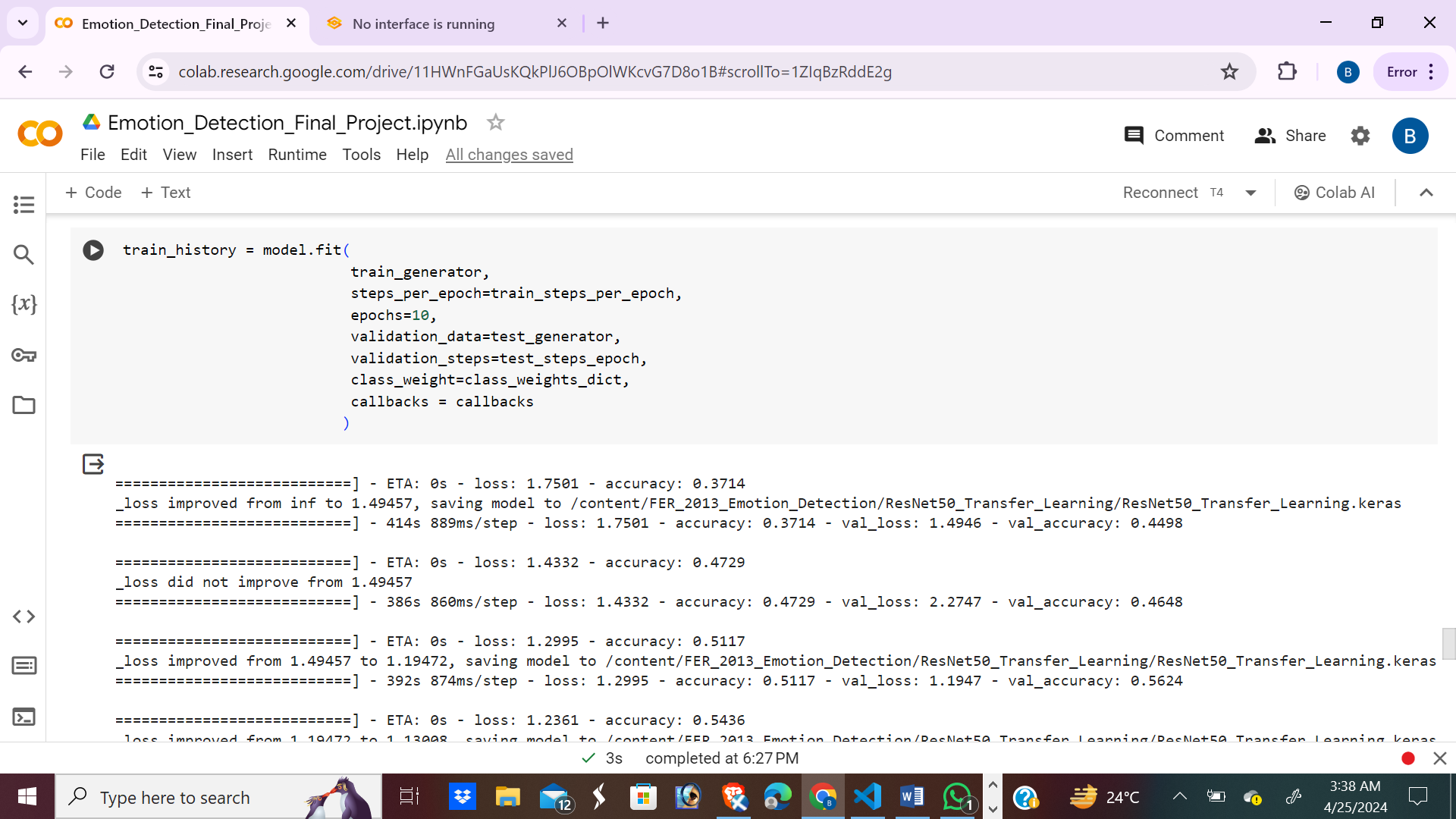
Now, we come to the critical part, where we define our CNN architecture which

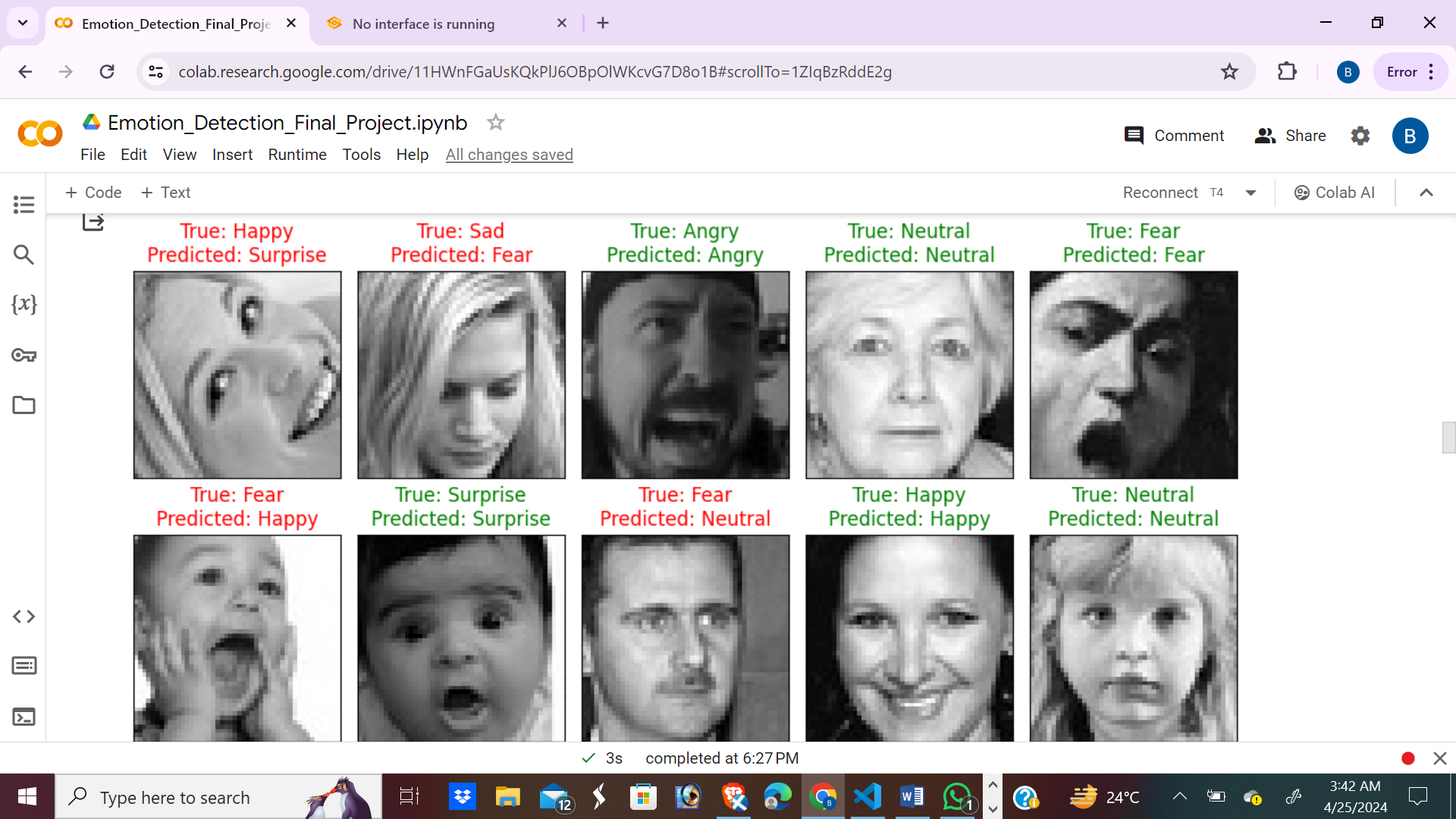
utilizes convolutional and pooling layers to learn significant features from the facial expressions, an input 2D Convolutional layer (with 32 filters) paired with a 2D MaxPooling layer.

We also used VGGNet and ResNet50 to check the accuracy. VGGNet typically uses 3x3 convolutional filters with a stride of 1 and padding to maintain the spatial dimensions.

ResNet50 specifically refers to a ResNet model with 50 layers, consisting of multiple residual blocks.

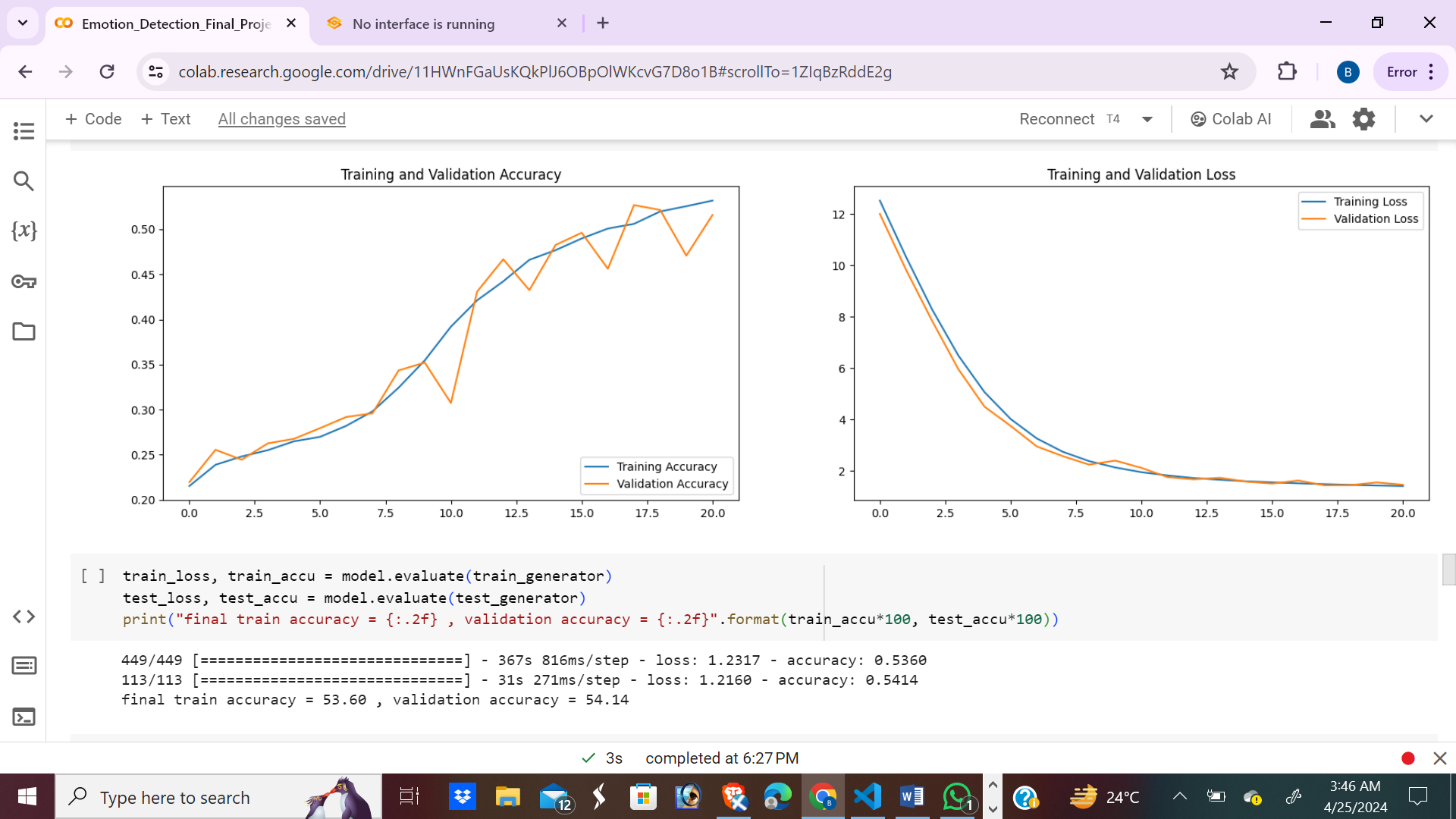
Next, we proceed to construct the CNN architecture as defined earlier. Additionally, we implement a model checkpoint mechanism to save our model periodically throughout training. This ensures that our model's progress and performance are preserved as it iterates over 50 epochs.



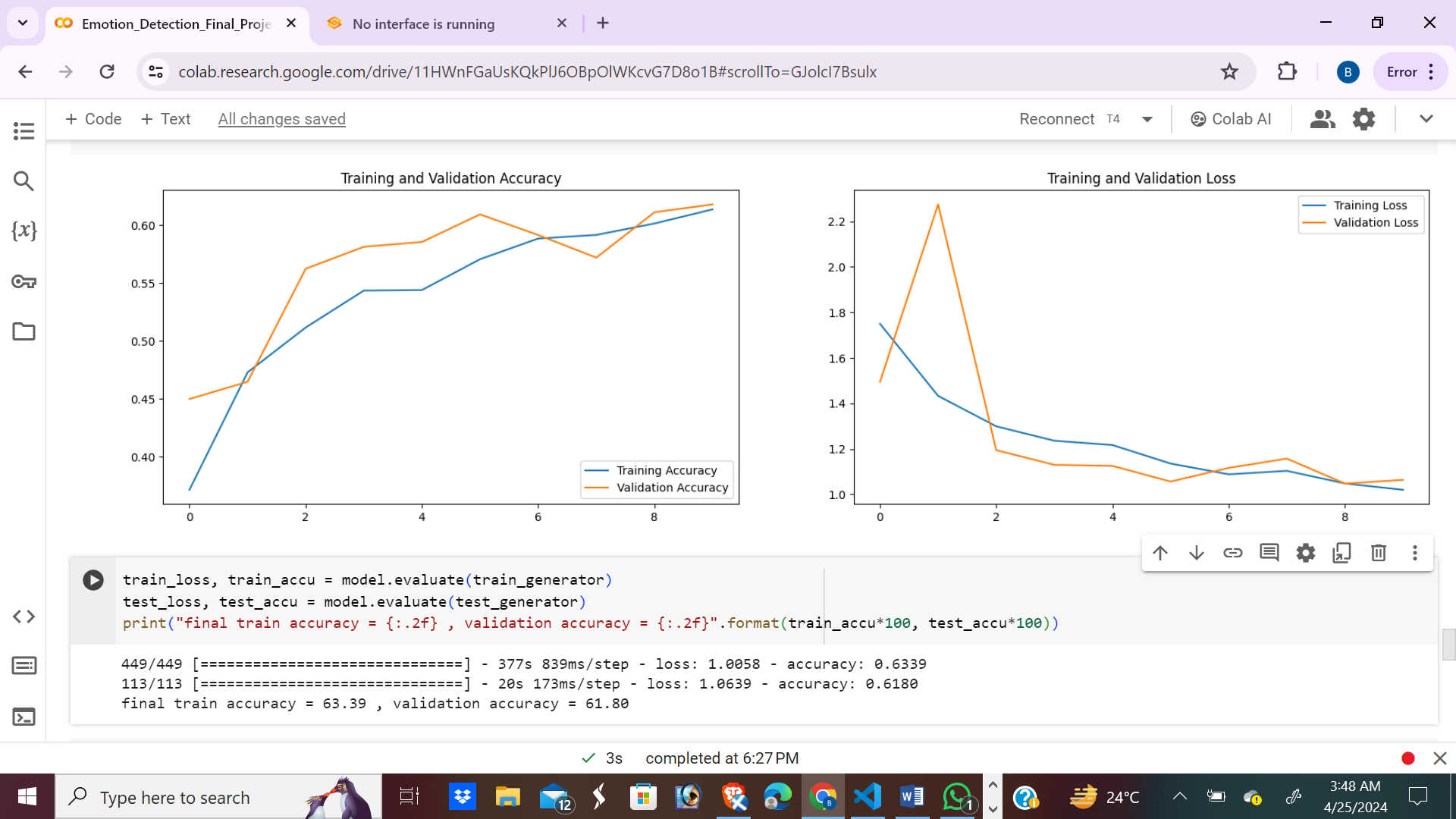


We also used VGGNet and ResNet50 to check the accuracy.

* VGGNet typically uses 3x3 convolutional filters with a stride of 1 and padding to maintain the spatial dimensions.



* ResNet50 specifically refers to a ResNet model with 50 layers, consisting of multiple residual blocks.

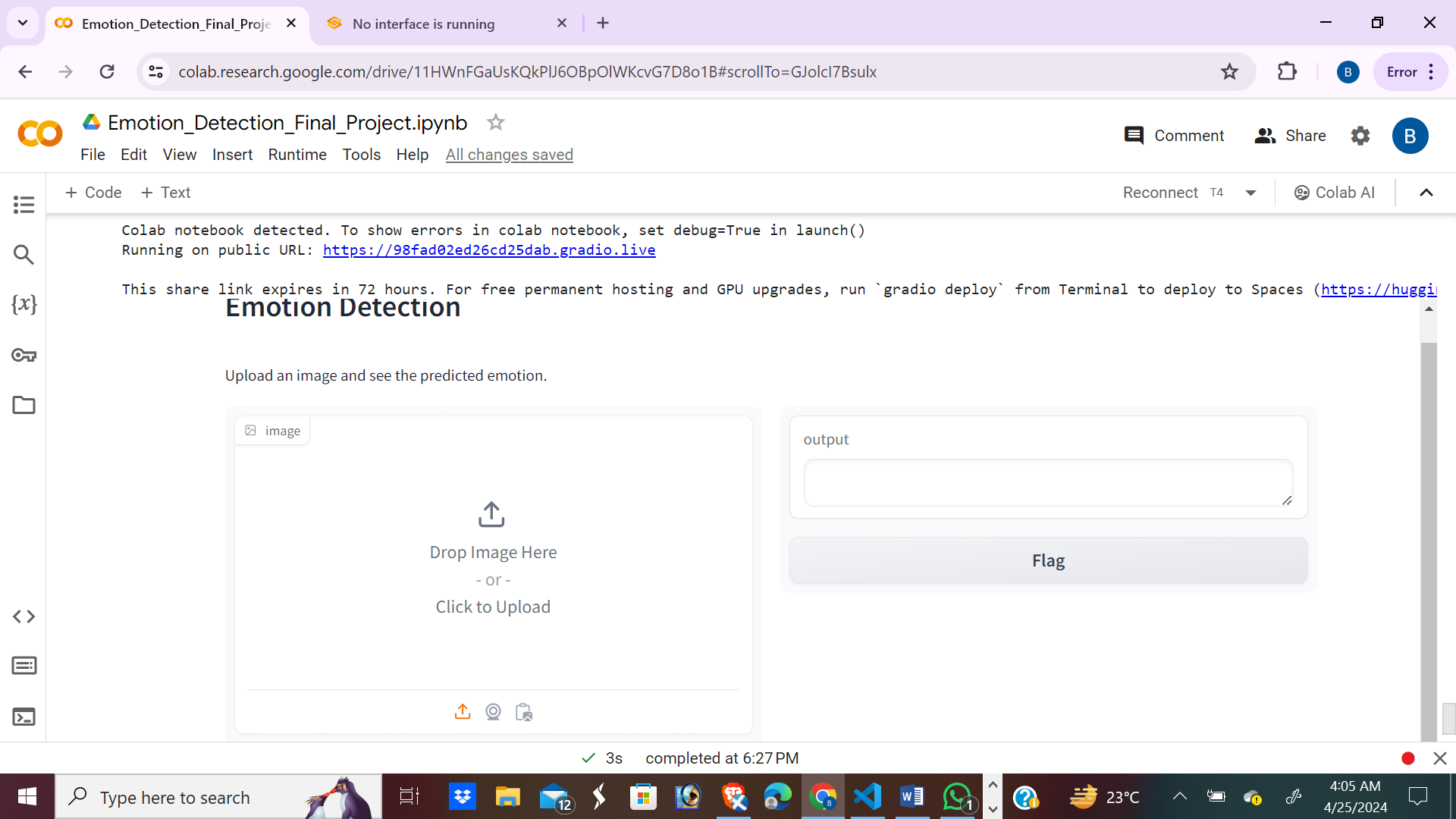


From the above we can say that ResNet50 has higher accuracy when compared with VGGNet. So we could ResNet is best choice.

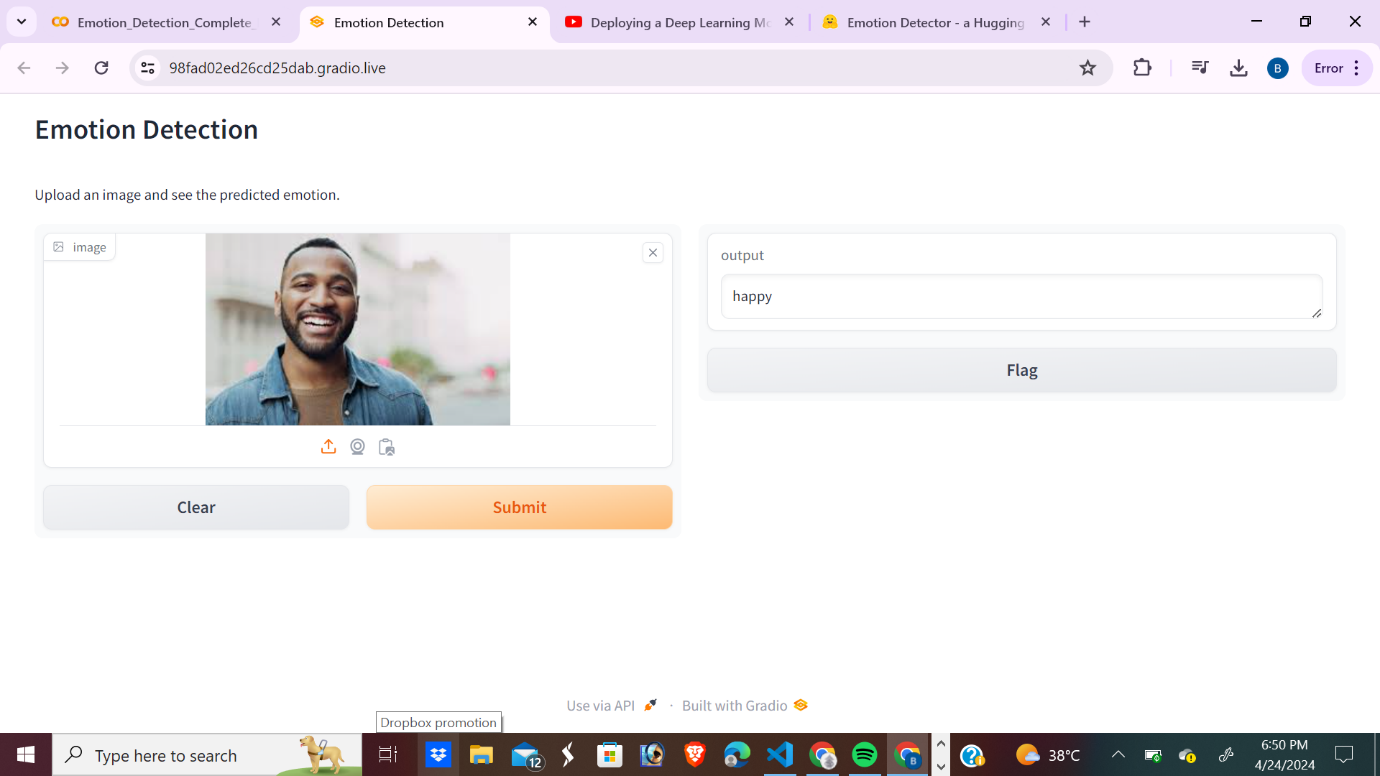
**Deployment:**

The next stage is to incorporate a deep learning model that can predict emotions into an intuitive graphical user interface after it has been refined. The goal is to create an interactive application that predicts a person's emotions. Install the required libraries into the virtual environment's root directory before beginning the project.

The model\_path, the model, and the associated tokenizer are the three essential components that need to be defined in order to set up the setup. Then, to handle any unique tokens in the text, we will create a function named "preprocess." After tokenizing the input text and producing predictions, the emotions function will be in charge of turning these scores into emotion labels. At last, we'll launch the Gradio interface class, which has an input textbox and a label for the output."



**Output:**

****

**Literature Survey:**

Many academics are trying to teach computers how to recognize emotions on faces. They used to depend on antiquated techniques that weren't particularly effective in capturing the nuance of facial emotions. However, with deep learning, things are now different. Deep learning techniques such as VGG16 and ResNet have demonstrated significant potential in identifying patterns in images, particularly emotional ones. Numerous studies have demonstrated that these approaches can be effective in accomplishing this objective; however, further refinement is warranted, particularly when confronted with challenging circumstances such as partially obscured faces or those in various poses.

**Models Used:**

1. **1. VGG16:** VGG16 functions as a kind of blueprint for instructing computers on how to interpret images. Its architecture is such that it gradually acquires the ability to identify patterns. It's a little more basic than other approaches, yet it works really well for many jobs, such as facial expression recognition.
2. **ResNet:** Another innovative technique that excels at comprehending images is ResNet. Its ability to handle intricate, multi-layered images is what sets it apart. This aids in its learning of increasingly intricate patterns, which is quite helpful for jobs like precisely identifying emotions.accurately.

**Conclusion:**

Our experiment demonstrates that, with the help of techniques like VGG16 and ResNet, computers can learn to recognize emotions from facial expressions rather successfully. These techniques have demonstrated a great deal of promise in identifying emotional patterns in images. ResNet fared better than VGG16 overall even though both techniques performed admirably. This is because ResNet can handle more complicated patterns. Still, there are a few issues we need to address, such as ensuring our system functions well with a variety of faces and circumstances. All things considered, our work advances computer intelligence and makes them more useful for deciphering human emotions.

**References:**

[1] Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems . Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. pages 1097–1105, 2012.

[2] Face expression recognition and analysis: the state of the art . Vinay Bettadapura. arXiv preprint arXiv:1203.6722, 2012.

[3] Xiaodong Cui, Vaibhava Goel, and Brian Kingsbury. Data augmentation for deep neural network acoustic modeling. IEEE/ACM Trans. Audio, Speech and Lang. Proc., 23(9):1469–1477, September 2015.

[4] In IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2013. Daniel McDuff, Rana El Kaliouby, Thibaud Senechal, May Amr, Jeffrey Cohn, Rosalind Picard. Affectiva mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions collected in the wild.

[5]Zisserman, A., and Simonian, K. (2014). Deep convolutional networks for large-scale picture recognition. The preprint arXiv is arXiv:1409.1556.   
[6] Sun, J., Ren, S., Zhang, X., & He, K. (2016). Deep residual learning for the identification of images. In Computer Vision and Pattern Recognition Conference Proceedings, IEEE (pp. 770-778).   
Bengio, Y., Goodfellow, I., and Courville, A. (2016). In-depth knowledge (Vol. 1). Cambridge MIT Press.

[7]Mahoor, M. H., Hasani, B., and Mollahosseini, A. (2017). A database for valence, arousal, and facial expression computation in the wild is called AffectNet. Affective Computing Transactions, IEEE, 10(1), 18–31.   
[8]Cohn, J. F., Tian, Y., and Kanade, T. (2000). extensive database for the study of facial expressions. The Fourth IEEE International Conference on Automatic Face and Gesture Recognition Proceedings, pp. 46–53.