Question 1

From the dataset provided to the students, two users' video data was selected. The data contained 164 videos that featured both users. Each video was divided into distinct frames, each taking one second to complete. A total of 1640 images were taken from the provided video, and an excel sheet was used to manually label each image. The five categories of boredom, engagement, confusion, frustration, and neural were applied to each image. If no emotion was seen, some of the images were removed. The labeled images were compared to those of a second person using the same dataset. In order to determine the intra-rater reliability using kappa score, a confusion matrix was created. The kappa value was determined to be about 0.83. Therefore, labeling was done again for all the images taken into consideration.

Now that all of the images were prepared, a comprehensive dataset was produced using the Python OpenCV library. To make training the model quicker, all images were changed to gray scale. The sizes of all the images were adjusted to match. Additionally, the pixel values were converted into a 4-dimensional array and scaled between 0 and 1. The dataset was now divided into test and train groups, each with 80% and 20% of the total data. Two convolutional and two maxpooling layers, three dense layers, and one output layer with five neurons made up the convolutional neural network. Dense layers and output "softmax" were utilized with the "Relu" activation function. With "adam" acting as the optimizer, "sparse categorical cross entropy" serving as the loss function, and accuracy metrics, the model was built. Model was trained using 32 batches and 20 epochs, and validation loss metrics were used to set an early stop calling back. After being tested with test data, the models correctly classified the images to the correct emotions with an accuracy of 86%. The images were subsequently classified using the pre-trained model VGG-19. The VGG-19 model, which had a total of 32 layers and was pre-trained using millions of images, was the best model for classifying images because it was straightforward, less complicated, and simple to use. The model was fed a dataset, and the first run produced accuracy of 91%. Images were plotted along with the labels that the model assigned to each one. The model was also tested using some images from a different dataset, and it correctly identified 3 out of 5 images.

Overall, a full model was created from scratch by creating the dataset to be used and trained on.