1. **Initial thoughts on the problem and how to chose approach.**

* Given problem consists of images of 41 people to identify whether the person is not drunk (sober) or mild drunk or drunk. For each person 4 different images before and after drunk are present in the data set.
* But in this work, given problem considered as the binary classification such as drunk or not.
* Hence, images taken before drinking considered as one category called sober and all the images taken after drinking considered as another category called drunk.
* Therefore, 4 images for sober and 12 images for drunk, hence the dataset is imbalanced.
* In this work, transfer learning is used, such that, popularly used CNN models that are VGG16 and ResNet50 are used in this work but the top layers from the models are removed only feature vectors are considered for classification.
* Generated feature vectors are classified using three different classification methods, that are Logistic regression, Decision tree and Multi-Layer perceptron

1. **Brief description on what folders were selected for training and why.**

* All the images are loaded into two feature extraction unit called VGG16 and ResNet50 to generate two feature vector size of 1 X 1024 for each image.
* For each feature vector of image, the corresponding label is assigned whether drunk or not as 0 for sober and 1 for drunk.
* Then finally, the generated feature vector and the corresponding label is split into training and testing set. And trained using multiple classifier on the training set and tested on the testing set.
* Three classifiers are used such as Logistic regression, Decision tree and Multi-Layer perceptron since these classifiers achieve better results for classification.
* Data also can be trained using other classifiers such as SVM, Ensemble classifiers etc and results can be compared.
* Finally the results are recorded that, the accuracy of training and testing and also classification report stating the confusion matrix and precision and recall values.
* The above process feature extraction and Training can be done in different category. In first category data is trained without any modifications.
* In second category, hand images are removed since recognizing the emotion and classifying drunk or not from the hand image difficult.
* Since the results from the 1st and 2nd category overfit and also False Negative and False Positive values are high due to unbalanced dataset.
* Therefore, to make the dataset balanced and also reduce FN and FP values data augmentation is introduced into the dataset.
* Third category, data augmentation increases the data size by making rotation and zoom over the existing data and makes the data equal over drunk or sober and finally trained with the increased dataset.
* Fourth category is uses the data augmentation and removes the hand images to train the dataset and results are recorded.

1. **Inferences from the results and conclusion**

* Different combinations of features and classifiers are executed and results are recorded. For each combination, following metrics are considered to evaluate the models that are training accuracy, testing accuracy, and confusion matrix including precision, recall, F1-score.
* For example, feature vector generated by VGG16 is trained using Logistic regression classifier model and the training accuracy is 99.421% and testing accuracy is 66.92%.
* Precision shows that, “for all instances classified positive, what percent was correct?” sober is detected 31% and drunk is detected 76% correctly.
* Recall shows that “all instances that were actually positive, what percent was classified correctly” which shows that, the actual sobers are classified as sober is 24% and actual drunk are classified as drunk is 81%
* F1 score is a harmonic mean of Precision and Recall and achieves 27% for sober and 79% for drunk.
* Same procedure is followed for the combinations of different features with different classifiers and also with and without data augmentation.
* Results shows that ResNet50 features achieves better F1 score and accuracy compared to the VGG16 features.
* **Especially ResNet50 features with Logistic regression classifiers produces better F1 score of 37% and 86% for sober and drunk respectively and accuracy of 77%. Therefore combination of ResNet50 and Logistic Regression without removing hand images and without data augmentation produces better results.**
* Removing the hand images does not increases the results.
* The above results shows that the model overfits and also low F1 score for sober but works well for drunk case due to increased error in the FN value.
* **False Negative shows that, the person is actually sober but classified as drunk. It is not good according to the law. Since, Sober should not be punished for drunken case. Hence, FN value should be reduced.**
* To overcome these issues, Data augmentation is introduced and dataset size is increased and also balanced.
* **Finally, with augmentation and by removing hand images and with ResNet50 features and Logistic Regression classifier, the model generates overall better results compared to the other models. The results are, F1 score for 71% and 69% for sober and drunk respectively and accuracy of 70%**
* Even though accuracy is 70% which is less compared to the model discussed above which generates 77% accuracy, due to introduction of data augmentation the FN value reduced which is good for the classification model