Nicholas Newton, Tim Fitzpatrick
CV Part 4 Deliverable (Final Report)
Project: Face Recognition

A. Description of the test database you collected or downloaded: What is the size of the database? What is different when compared to the training and validation subsets? Why do you believe these differences are sufficient to test your final programs? **(2 points)**

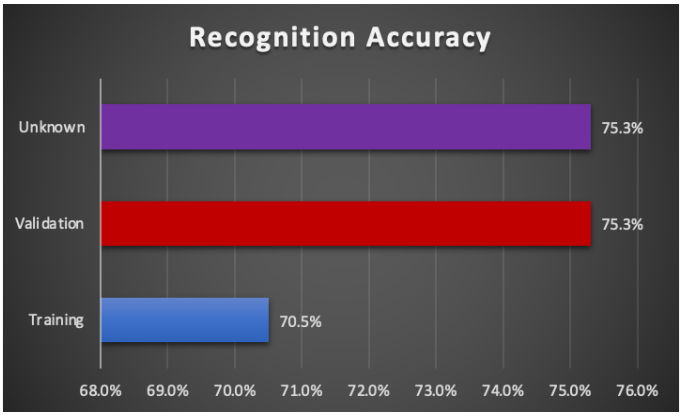
   The database of face samples we collected comes from a project in the UMass Computer Vision Department called Labeled Faces in the Wild (http://vis-www.cs.umass.edu/lfw/). This image set was then filtered down to samples with individuals containing multiple image sets. From these user image sets, we divided the candidates into testing, validation and unknown datasets. This database after removing single image candidates contained 908 training candidates, 412 validation candidates, and 360 unknown candidates.

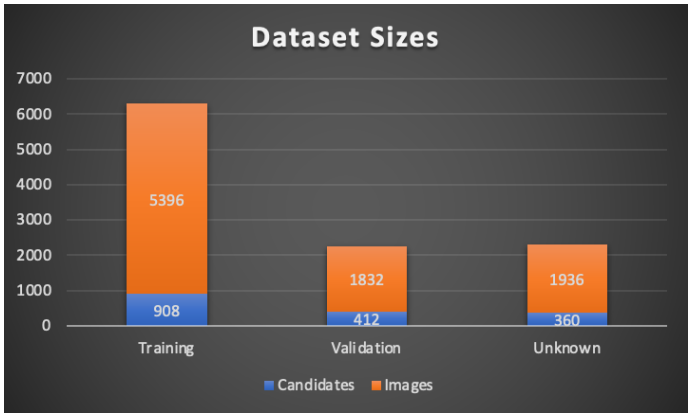
B. A classification accuracy achieved on the test set. Use the same metrics as in part 3. **(3 points)**

   Our training data contained 1936 unique images. These were among 360 candidate sets. From our model, 75.3% of the images were identified correctly. Accuracy was calculated based on the number of images correctly identified.

C. Most of you should see worse results on the test set when compared to the results obtained on train/validation sets. You should not be worried about that, but please provide the reasons why your solution performs worse (with a few illustrations, such as pictures or videos, what went wrong). What improvements would you make to lower the observed error rates? **(5 points)**

   Based on the results of our unknown data, we actually were able to hit similar marks on our accuracy of the data. The training data most likely took the biggest hit in accuracy because increasing the volume of images made it more difficult for the model to differentiate between initial features extracted from the dataset.Validation data and the unknown data found similar accuracies with around the same sample sizes. To improve the accuracy, we would look more into methods to classify images further using more specific indicators.





D. Imagine you want to present your final program to a friend (or investor). And this presentation should be short and illustrative (that is: show what you did and how good it is). Prepare a short video, or a short slideshow with pictures, presenting how your final program works. **Here are good examples prepared** by former CV students. **(5 points)**

   Video Link: emailed and on GitHub

E. Push your **final (to be graded)** codes realizing what you mention in the report to GitHub, along with instructions how to run them (either Adam or Siamul will do it to see how the final solution works on test data). Your program(s) should pick one example from the test set (please attach this sample to your codes) and present the processing result. We should be able to run your programs **without any edits** -- please double check before submission that the codes are complete. **(5 points)**

   Github Link: https://github.com/nnewton2/cv_face_recognition

   https://github.com/nnewton2/cv_face_recognition/tree/main/Final_Delivery