2.1. As explained in class, the provided third-party face recognition library is able to extract a 512-dimensional float feature vector from a given face image, as well as calculate the angular distance between two feature vectors, using ArcFace (https://bit.ly/3J8Tgtc). The expected behavior for the software is to generate small distances for two face images that depict the same individual (genuine pair), and large distances for two images that depict different individuals (impostor pair).

Leveraging the content of **only** the "dataset" folder within the provided data, the third-party face recognition library, and metrics learned in class, please determine **what is a good angular distance threshold to separate genuine from impostor pairs**. While providing your answer for the distance threshold, please explain in detail how you computed it. (1 point)

The angular distance threshold that corresponded with EER(Equal Error Rate) for this dataset was found to be 1.6202.

* In order to determine a suitable angular distance threshold that discriminates genuine pairs from impostor pairs, first we needed to generate genuine pairs and impostor pairs with associated angular distance scores. In order to do this, I first needed to generate all possible genuine pairs, as the dataset we had only contained 9 images for each subject, with 10 distinct subjects. Each subject therefore has 36 possible genuine pairs, and with 10 subjects this leads to only 360 possible unique genuine pairs. As this was not a large task, I generated distance scores for all 360 possible genuine pairs. The imposter pair set was much larger, but to keep the dataset balanced, I generated 360 distance scores for 360 unique imposter pairs. The angular distance scores were generated utilizing ArcFace, and the libraries covered in class. After generating the 720 distance scores along with their associated labels in an array, I was able to utilize the compute_fmr_fnmr_eer function covered in class to get the EER and the associated angular distance.

* The optimal threshold was identified at the point where the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR), known as the Equal Error Rate (EER). For this specific dataset, the EER was found to be 11.67%, with the corresponding angular distance threshold calculated at 1.6202. This threshold was chosen as it represents the point of equilibrium where the system is equally likely to falsely accept an impostor as it is to falsely reject a genuine user, thereby providing a balanced security measure for the face recognition system.

* **A caveat I would mention, especially for this system, is that since we were using Viola-Jones for the facial detection, there is a significant chance that there were some genuine pairs/images that the facial detection system failed to pick up on the face due to the background and as such reported much larger distance scores for the genuine pairs than would otherwise be the case. I suspect that this was the case, and as such a more appropriate threshold for this system might in fact be closer to 1 based on the histogram.**

2.2. **What is the AUC** of the face recognition system you are using? In addition, please provide a graph with the system's **ROC curve**. (1 point)

The facial recognition system's AUC for this dataset was found to be 0.95.
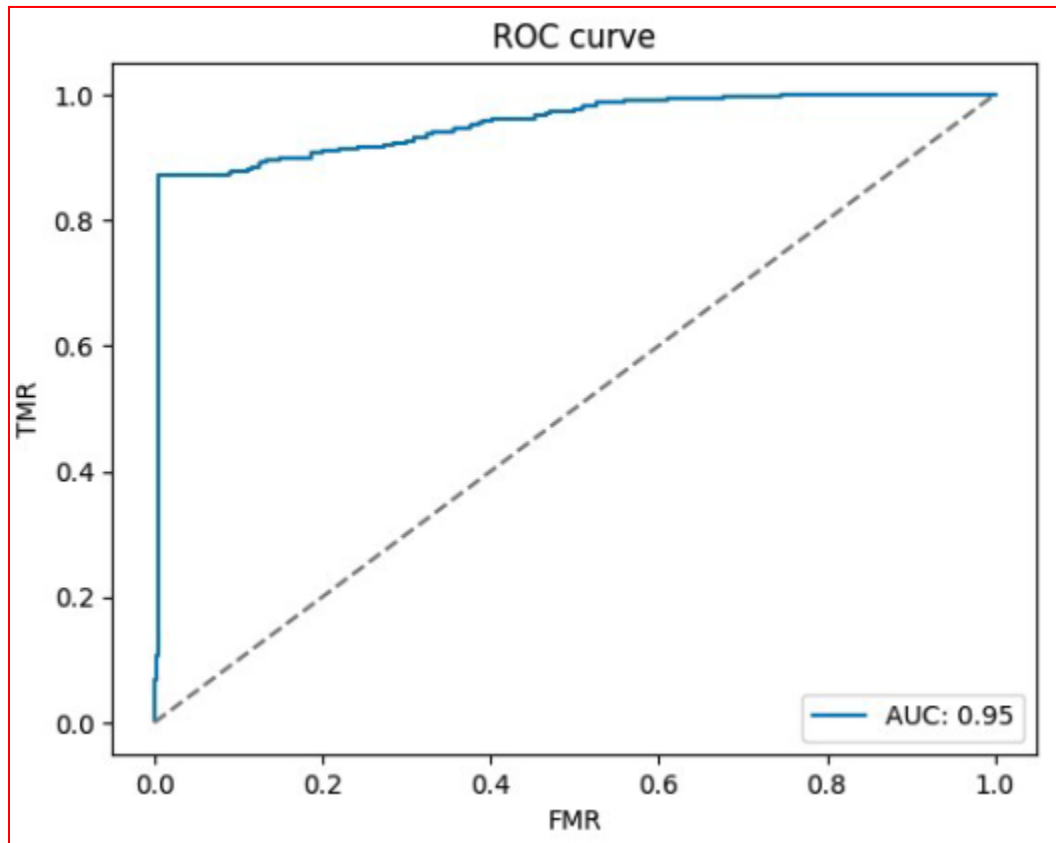
2.3. By leveraging the face recognition system and the distance threshold previously computed, and by either capturing your face with your webcam or providing an image with your face, find within the "dataset" folder what is the individual that is the most similar to you. **Please provide the subject ID and the angular distance between your face and theirs**. In your opinion, do you have anything in common with this subject (e.g., gender, ethnicity, age, etc.)? If yes, what is it? (2 points)

The subject ID that the system found my image to be most similar to was subject 10. The angular distance was found to be 1.4331976.

The subject that returned as the most similar to my picture does share some similar characteristics to me. For one, we seem to share the same ethnicity, as I believe the subject in the image is Indian. In addition, we share somewhat similar facial features, especially around the eyebrow region.

2.4. By leveraging the face recognition system and the distance threshold previously computed, **please provide the subject ID** (or "**UNKNOWN**", if the individual does not have a face within the "datasets" folder), as well as **the respective angular distances** that supported your decision, for each one of the 15 images provided within the "queries" folder. (6 points)

Query 6706: Unknown
Query 6653: Subject 5, distance score of 0.10225881
Query 6510: Subject 10, distance score of 1.5110078 (this was wrong, it should be unknown)
Query 6012: Unknown
Query 4387: Subject 2, distance score of 0.15844831
Query 5314: Subject 8, distance score of 0.21567674
Query 4535: Subject 3, distance score of 0.24399015
Query 7745: Subject 2, distance score of 1.4956621 (also wrong, should be unknown)
Query 7076: Subject 6, distance score of 0.11525211
Query 7633: Subject 4, distance score of 0.074939854
Query 9708: Unknown
Query 9395: Subject 10, distance score of 0.07755465
Query 4507: Subject 7, distance score of 0.09682656
Query 4168: Subject 1, distance score of 0.13350847
Query 7549: Subject 9, distance score of 0.2223371