Final Project Report

IST 718 – Sec1 Winter 2022

Classmate Computer Vision

This paper seeks to recommend the best possible facial recognition algorithm by applying a traditional approach (with Euclidean distances) and a more modern take (using the entire face) to identify members of our IST 718 class and rendering various models such as kNN, Random Forest, Naïve Bayes, and a Convolutional Neural Network (CNN).

With a 98.5% accuracy, the CNN model scored highest and is our final recommendation to be used.

Lei Cheng

Katie Haugh

Garen Moghoyan

Sandy Spicer

Karl Treen

Introduction to Facial Recognition

Long considered the stuff of science fiction, biometric facial recognition has permeated our daily lives, for better or worse. Several industries have adopted and benefitted from the advancements made in the field over the past 60 years, including law enforcement, border control, retail, mobile technology, and banking and finance.

Facial recognition technology began in the 1960s, with academics using computers to recognize the human face. Because an intelligence agency initially funded their efforts, much of the work was never published. However, it was later revealed that the initial work involved manually marking various "landmarks" on the face such as eye centers, mouth, etc. These were then mathematically rotated by a computer to compensate for pose variation. The distances between landmarks were also automatically computed and compared between images to determine identity.

From the 1970s through the 1990s, law enforcement agencies developed their own facial recognition systems. These were crude compared to the technology today, but these systems' work led the way to modern facial recognition programs.

The United States military's Defense Advanced Research Projects Agency (DARPA) influenced the commercial facial recognition market, which led to the Face Recognition Grand Challenge (FRGC) launch in 2006. Their goal was to promote and advance face recognition technology designed to support existing face recognition efforts in the U.S. Government. The FRGC evaluated the latest face recognition algorithms available. High-resolution face images, 3D face scans, and iris images were used in the tests. The results indicated that the new algorithms were ten times more accurate than the face recognition algorithms of 2002 and 100 times more accurate than those of 1995, showing the advancements of facial recognition technology over the past decade.

In 2010, Facebook began implementing facial recognition functionality that helped identify people whose faces may feature in the photos that Facebook users update daily. Facial recognition technology advanced rapidly from 2010 onwards. Today, it is used on mundane tasks like unlocking cell phones, paying for purchases through online wallets, accessing credit card or bank accounts, and identifying and tagging familiar faces on social media accounts.

How does facial recognition technology work?

While there are various approaches to facial recognition, they all have similar foundations. First, a picture of a person's face is captured from either a photo or video. There may be multiple individuals in the image or just one. The person may be looking straight into the lens or nearly in profile. Next, the geometry of the face is examined. The technology records key factors such as the distance between the eyes or from the forehead to the chin. The program identifies the facial landmarks that are critical in determining the person in the image. Finally, a "facial signature," which is a mathematical formula using the facial landmarks, is compared to a database of known faces in hopes of finding a match.

Specification

The Problem

Computer vision is becoming increasingly popular and necessary technology in today's interconnected world. We constantly use everyday devices such as our phones, tablets, laptops, or cars for their computer vision capabilities. Our team challenged ourselves to create our own computer vision program while improving upon an already existing model to compare and contrast the two's performance.

The Hypothesis

We predicted that our models would be able to recognize the faces of all our classmates in IST 718 with a high degree of accuracy. Knowing that access to data was limited to students who spoke up during class, we hypothesize that the traditional method's accuracy would be greater than 87.5%, and the modern approach's accuracy would be greater than 92.5%.

The Data

More than 20 videos were downloaded from the 2U Zoom platform that contained members of our class. The goal was to use these videos to extract individual photos of each student. The recordings were of IST 718, as well as other classes we had shared with our current classmates.

Only videos recorded in "speaker view" were collected, meaning the person speaking takes over the screen.

The collected recordings were placed into a "raw footage" folder where we could then process each via the FinalProject_ImageProcessing.ipynb program. This script runs through a series of steps on each video to extract the images we needed for our analysis.

The program takes in a set of inputs:

- Path of video(s) to process or entire folder containing multiple videos
- Skip rate
- Max number of images per person
- Names of people in our roster
- Path of where to save the images

From there the program:

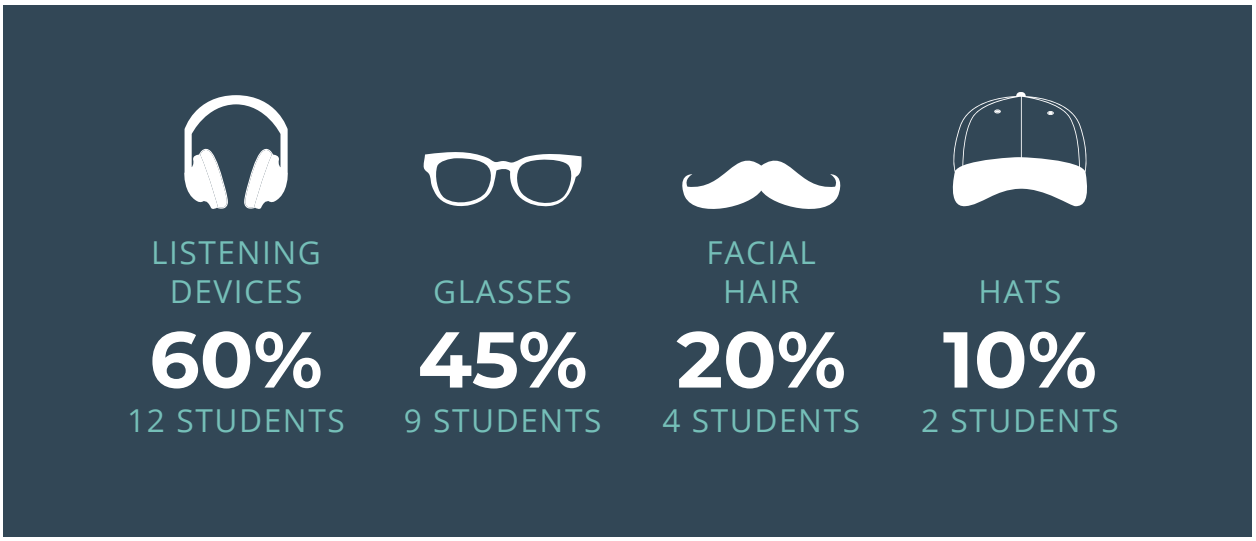
1. Reads the **name** from bottom left corner of the video frame using pytesseract
2. If the **name** is in roster, grab video frame as image
3. Names the image with an abbreviated person name + video name + frame
4. Saves the image as jpg into designated folder

Due to some students having busy backgrounds, there were several instances where the program was unable to extract a name or the appropriate name from the images. This required us to continue collecting videos after each class to ensure we had an adequate number of photos of each classmate.

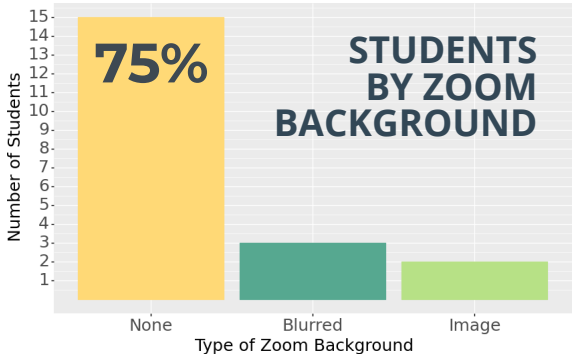
Observation

We were able to collect data on 20 students in the class:

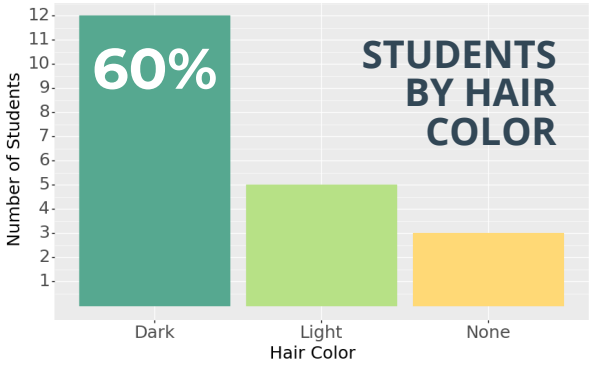
- Twelve (60%) had listening devices such as headsets and earbuds,
- Nine (45%) wore glasses,
- Four (20%) sported facial hair,
- Two (10%) wore a hat.

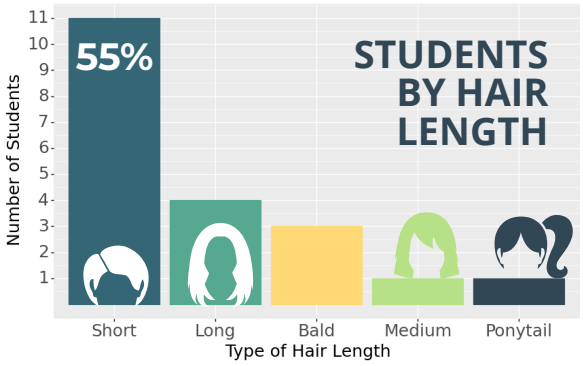


Additionally, we looked at Zoom background amongst the class. Fifteen had no set background, while three blurred their background, and two used an image such as the Syracuse logo.

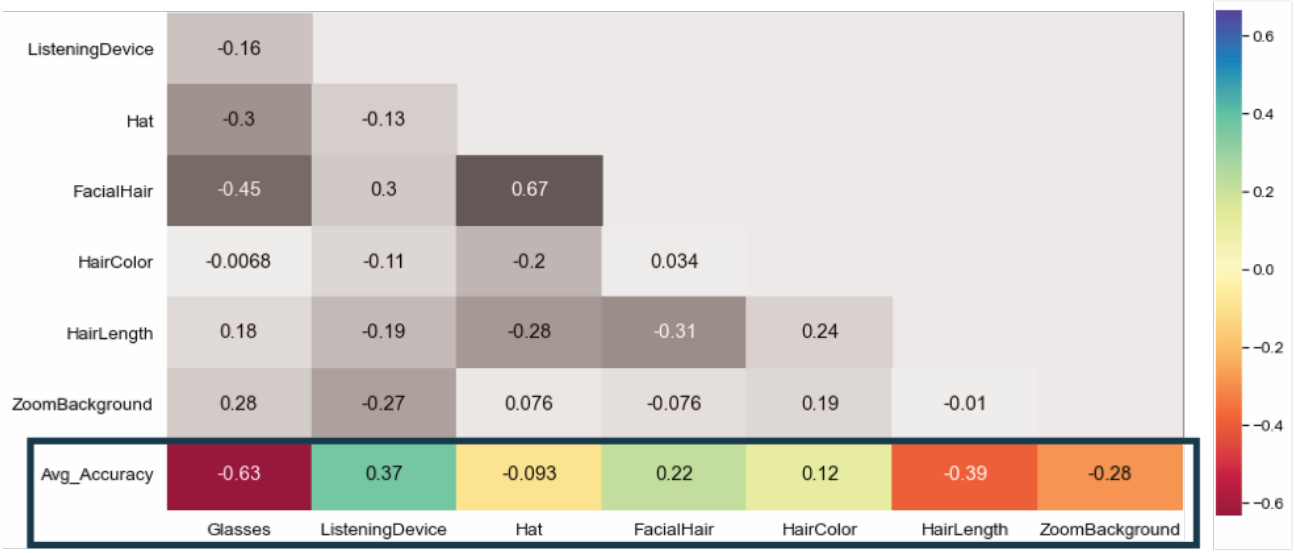


Hair color and length was another feature we took note of during our initial investigation. Twelve students in our data had dark hair, five sported light hair, and three were bald. Of the 17 who had hair, eleven had short hair, while four had long. Only one student had medium length hair and only one wore a ponytail.





From there we decided to run a correlation to look if any of these variables had a relationship with the average accuracy of our models. Glasses was the only variable that had a moderately strong relationship with Average Accuracy, but the relationship is a negative one at -63%. This may suggest that glasses have a negative impact on our model's accuracy.



Analysis

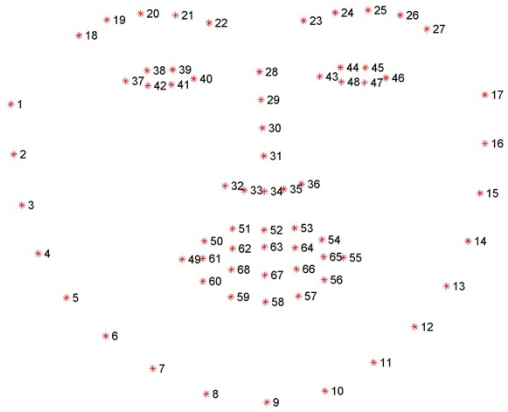
Our group explored two avenues for facial recognition. The first approach uses Euclidean distances between landmarks on the face to identify the individual, while our other approach utilizes the entire face.

Approach 1: Euclidean Distances

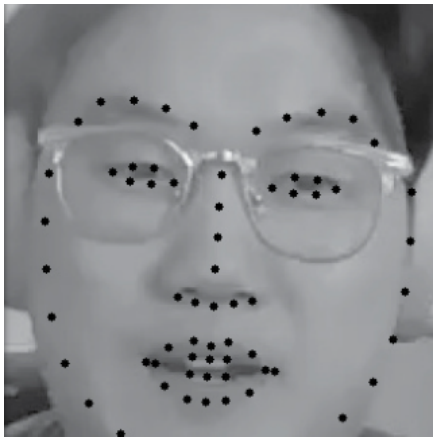
We prepared the previously captured images via the 2_FaceRecog_ToCSV_new.ipynb program to start this approach. This script runs through a series of steps on each image to recognize the face and mark with points for our analysis.

For each image extracted from the "Image Processing" program:

1. The image is converted to grayscale.
2. Using CV2 and a pre-programmed face detector XML file, the program detects the face and crops it, discarding excess pixels. The pre-programmed face detector is from OpenCV and contains trained classifiers for seeing a face when it is front-facing.
3. The image is then resized to 330 x 330.
4. Using dlib's pre-trained shape predictor, each face is marked with 68 landmarks, as shown in the figure to the right. The shape predictor was pre-trained on the ibug 300-W dataset.
5. The locations of the points on the face are then added to a data frame and later exported to a CSV.
6. If the shape predictor was successful in recognizing a face and placing the points on the face, the program continues on to calculate the Euclidean distance between the nose and the other 67 landmarks on the



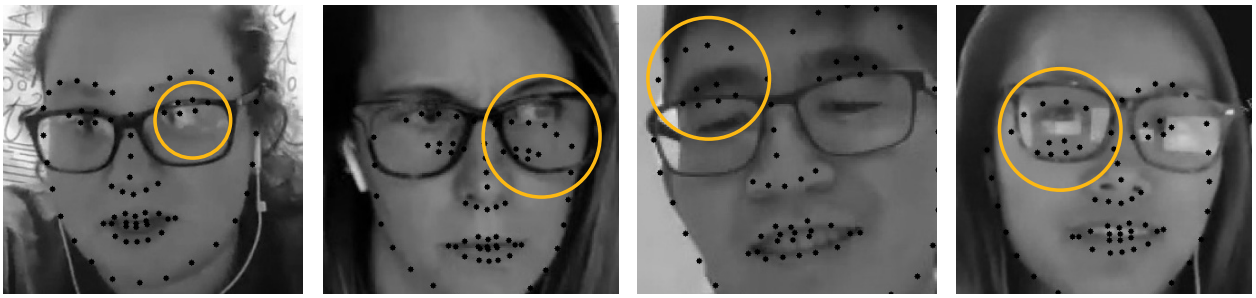
Source: https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/



face and also the Euclidean distance between the corner of the right eye and the other 67 landmarks.

7. These distances are then compiled into a data frame and exported to a CSV file.
8. The program then selects the predetermined number of images (110) of each person to use for the training and test data.
9. Using an 80/20 split, the image data is divided into train and test sets to be saved separately.

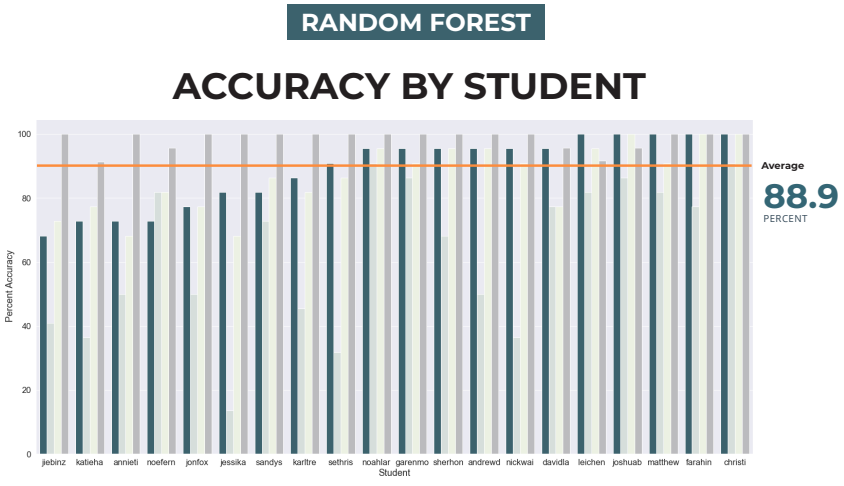
When the shape predictor determined the location of landmarks on the face, it struggled with students who had angled their faces and even more so with those who wore glasses.



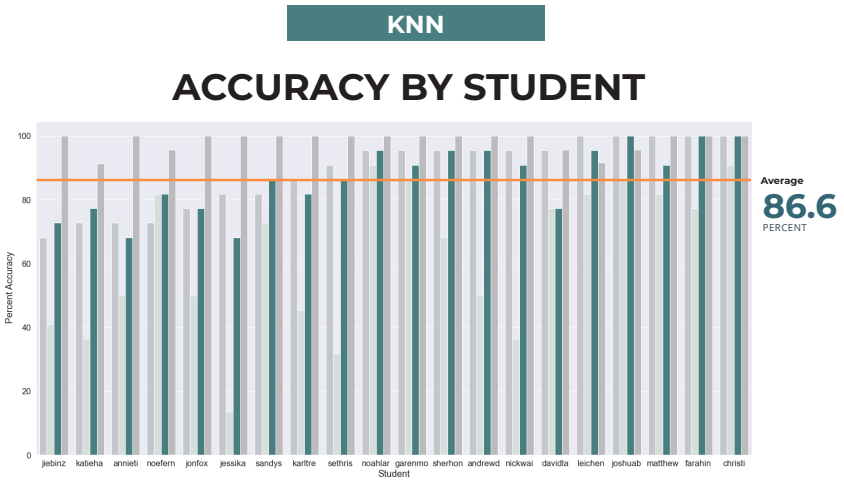
Three machine learning classification techniques, which include Random Forest, kNN and Naïve Bayes were explored to determine which was most efficient in recognizing the faces of our fellow classmates.

Random Forest

Random forest models work well because a large number of relatively uncorrelated models operating as a committee will outperform any of the individual constituent models. For our random forest model, we designated the model to build 80 trees and the model produced an accuracy of 88.9%.

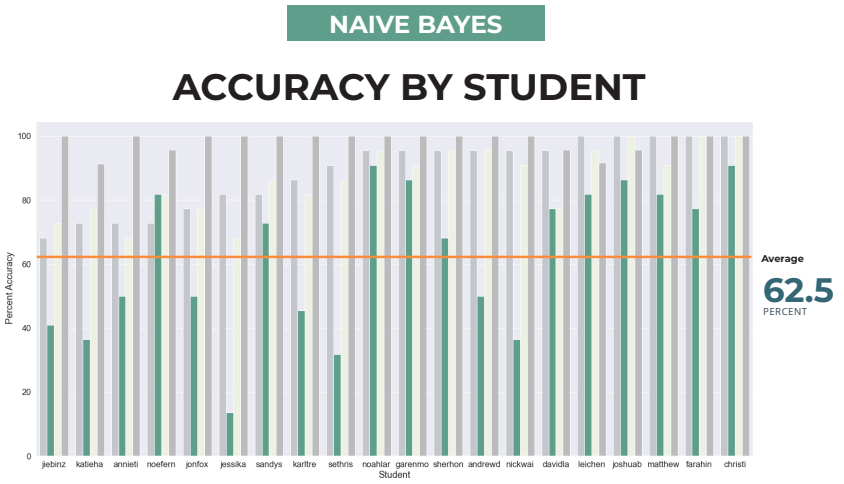


kNN



K – Nearest Neighbors (kNN) is an instance-based algorithm, which compares neighboring data points' similarities. There is no need to build a model, tune parameters, or make additional assumptions. The algorithm is quite versatile and can be used for classification, regression, and search. However, the algorithm gets significantly slower as the number of examples, predictors, or independent variables increase. For our kNN model, we tested between 3 and 9 neighbors and found 3 as the optimal number to be the best balance between accuracy and overfitting. The model produced an average accuracy of 86.6%.

Naïve Bayes



Naïve Bayes is a supervised machine learning algorithm that handles high-dimensional data very well, so we thought we would see how it performed on our facial recognition problem. Additionally, it is easy to parallelize and performs better than more complicated models when the dataset is small, such as ours.

Unfortunately, our naïve bayes model produced the low accuracy of 62.5%, which was the lowest of all the models.

Approach 2: Modern Approach using the Entire Face

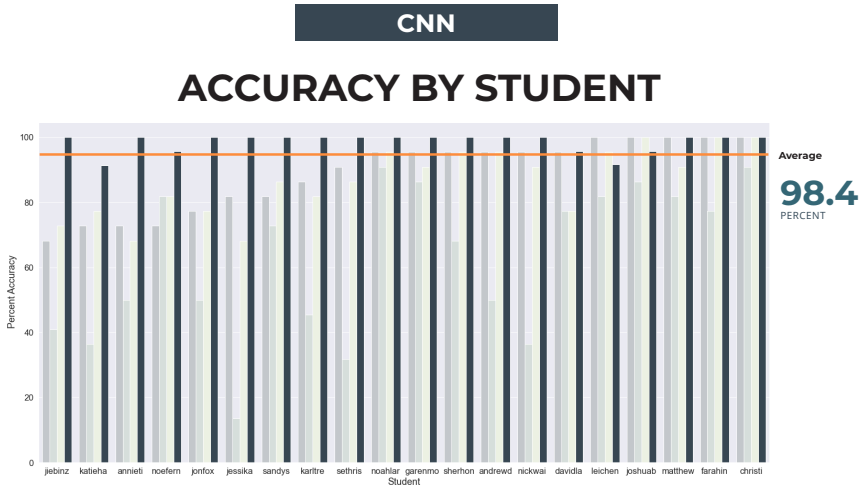
The Convolutional Neural Network (CNN) is a type of neural network algorithm typically used for many computer vision applications. Images, when used in a machine learning setting, have high "dimensionality". Each pixel in an image would be a feature which means a small 50x50 image will have 2500 features.

CNNs are effective in reducing features within a dataset without sacrificing model quality. The convolution layer compresses the image by a factor equivalent to the kernel size using matrix operations. By continuing to use more convolution layers in a CNN, one is able to significantly reduce the dimensionality.

Therefore it makes sense to choose CNN algorithm for facial recognition (necessitates higher resolution) models.

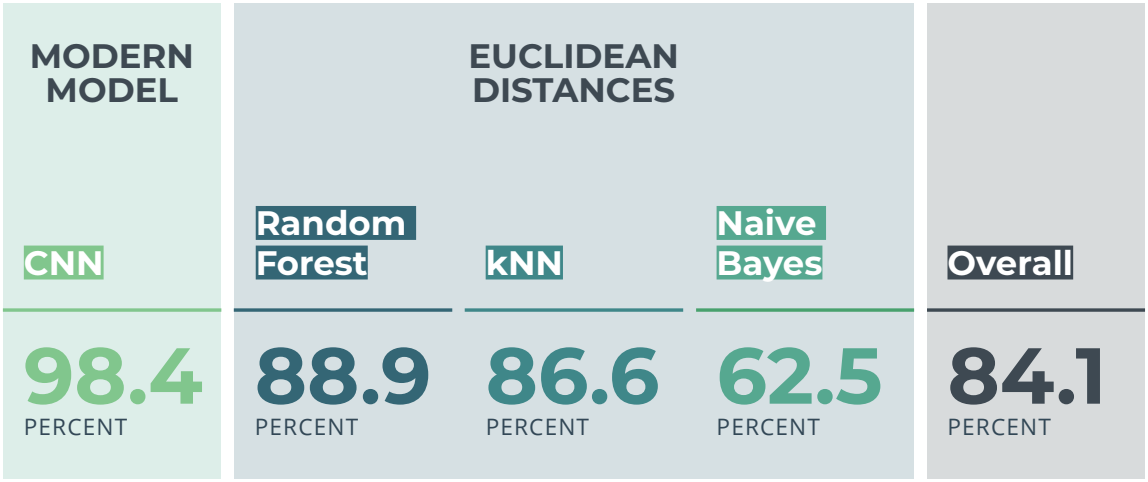
In the modern approach, we used the collected and cropped images and fed them into Python using the Keras ImageDataGenerator function while keeping the testing and training sets consistent with the genuine method. This function allowed us to add zoom, sheer, and flipping to the images to enrich the dataset and decrease overfitting. Once the images were collected and aggregated into a dataset, the model was created and initialized. The Keras model had:

- Convolutional layer of (5,5) kernel size & max pool of (2,2)
- Convolutional layer of (3,3) kernel size & max pool of (2,2)
- Dense layer with ReLu activation function
- Dense layer with Softmax activation function.
- The loss category was categorical cross-entropy, and the optimizer was Adam.
- Finally, the model was trained over 25 epochs and 32 steps per epoch.
- The final model accuracies were a 98% training accuracy and a 98.4% validation accuracy. This means the model is extremely accurate with little over/under fitting.

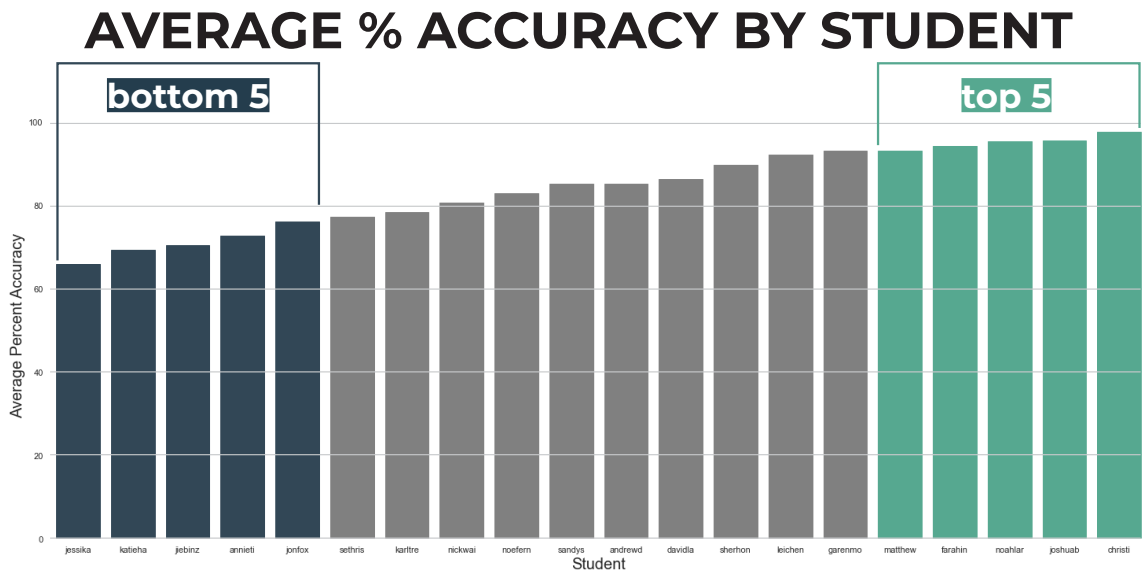


Recommendation

The various models' accuracies are summed up in the graph below.



To take a deeper dive into the accuracies, we observed the average accuracy by student. We noticed that the bottom five students all wore glasses right off the bat.

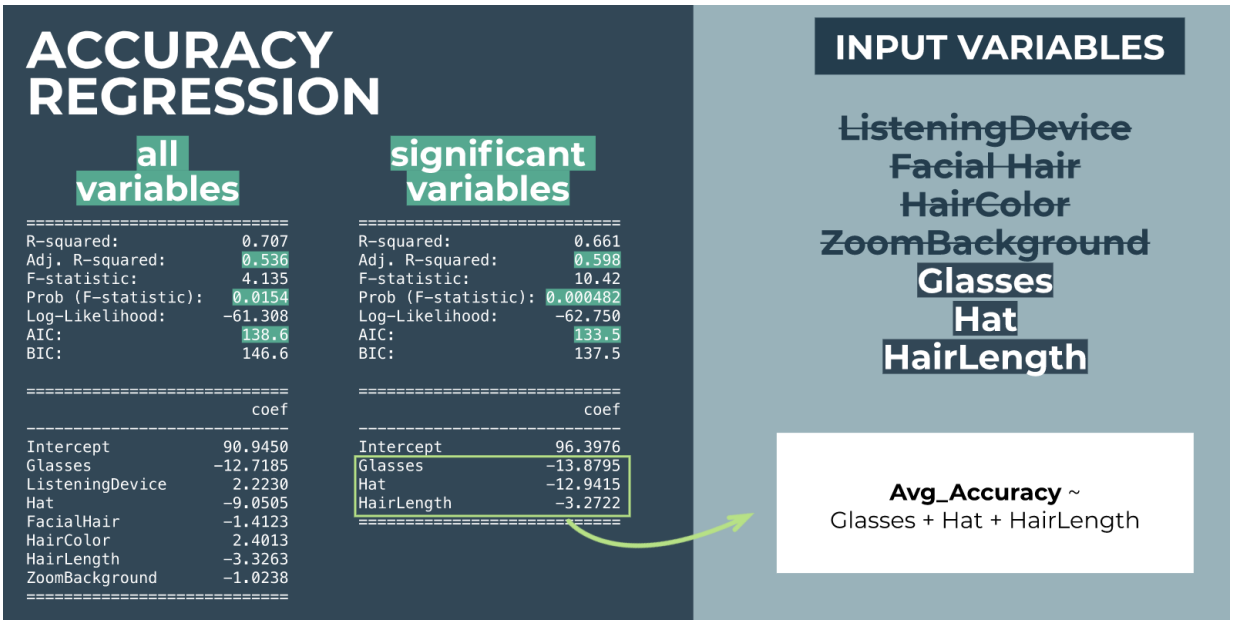


To investigate further, we analyzed possible factors impacting the accuracies by running various regressions; first by using all the variables, namely:

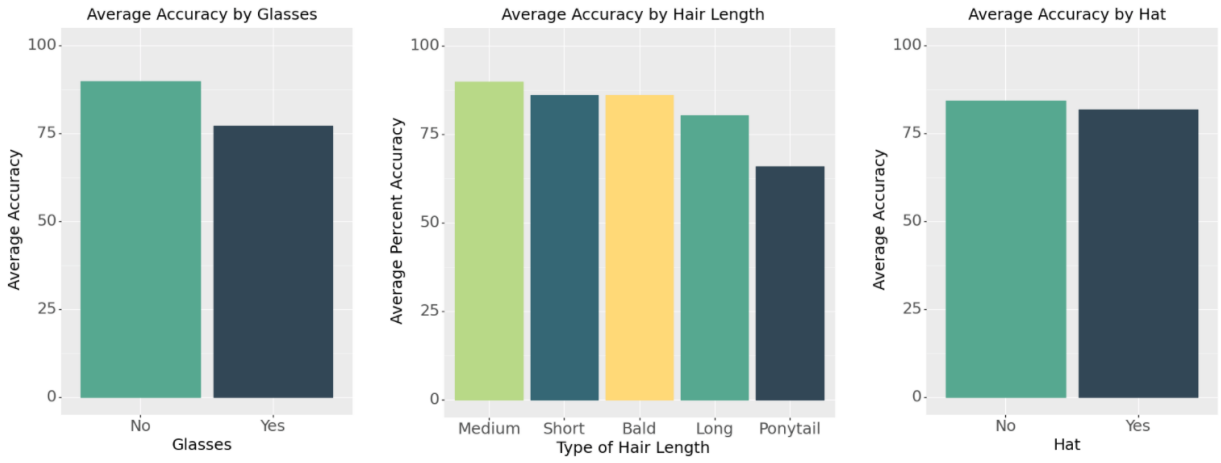
- Glasses
- Listening devices

- Hat
- Facial hair
- Hair color
- Hair length
- Zoom background

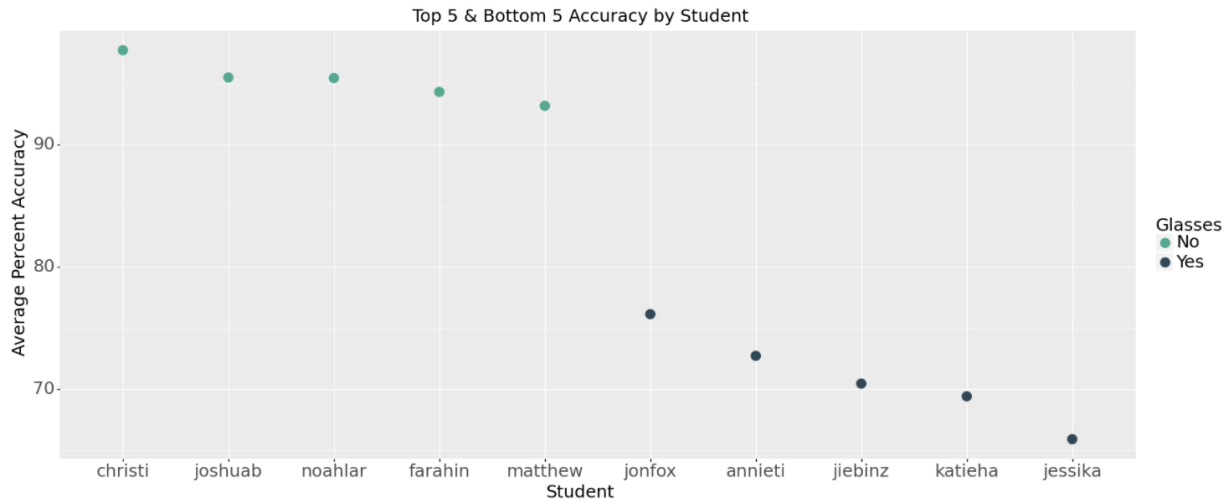
The ensuing regression had an R-squared of 70.7% with glasses, hat and hair length being the most significant variables. We then isolated those three to run a new regression and obtained a R-squared of 66.1%.



The graph below shows the impact that the glasses, hair length and hat variables have on the average model accuracy.



The variable that impacted accuracy most negatively was glasses, while the story for average hair length and whether a student was wearing a hat or not is not as clear. Those without glasses' had an average accuracy of 87.5%, while those with glasses were sitting at 75% accuracy. Indeed, the five lowest accuracies belonged to bespectacled people in the Zoom meetings.



Upon diving deeper, we realized the glasses impeded the capture of the pictures by incorrectly identifying the facial landmarks, most notably the eyes. This led to the low accuracies we recorded.

For businesses wishing to implement facial recognition, the model we recommend is the Convolutional Neural Network, cited above, with a 98.5% accuracy. The CNN model efficiently consumes more complex imagery without the need for excessive simplification, and its brain-like structure is very effective for image recognition. The model is not limited by one fixed strategy,

and it can be fine-tuned at different node levels. It has superseded the traditional approach and paved the way for even faster and more accurate models.

Anyone with a high school education, a camera, and a laptop can now build a workable facial recognition system. With cameras in both public spaces and private homes, computer vision is increasingly difficult to avoid. For concerned individuals wishing to avoid facial recognition, we recommend not posting facial photographs to social media, wearing dark or reflective glasses in public, not breaking the law, using outdated mobile technology (without built-in cameras), covering webcams unless necessary, and shopping either online or in places without cameras, like farmstands, farmer's markets, and smaller stores.

Pandora's box has been opened. As the world comes to terms with computer vision, there will doubtless be legislation proposed to both limit its scope and to protect its adherents. Computer vision is both preventing crimes and supporting autocratic states. The question is no longer whether facial recognition will become pervasive but, rather, how humanity will deal with the inevitable consequences.

References

https://ibug.doc.ic.ac.uk/resources/facial-point-annotations/

https://www.nec.co.nz/market-leadership/publications-media/a-brief-history-of-facial-recognition/

https://www.nytimes.com/wirecutter/blog/how-facial-recognition-works/

https://www.facefirst.com/blog/brief-history-of-face-recognition-software/