Facial Recognition:
Accuracy by Age, Gender, and Race

Group 11: Nishita Cheekatamarla (leader), Enmanuel Baez, Keivon Chamanara

DS 4002 – April 23, 2024

Project Overview

01 Project Details

02 Data Acquisition and Explanation

03 Analysis Plan and Justification

04 Tricky Analysis Decision

05 Bias and Uncertainty Validation

06 Results and Conclusions

07 Next Steps

Project Details

Motivation

Facial recognition tech has a wide range of applications, but it comes with some drawbacks.
Law professor at the University of Calgary, Gideon Christian, notes that some facial recognition systems have an accuracy rate of around **99% for White men**, but this rate drops to **35% for Black women**.

Research Question

How accurately are different ages, races, and genders identified and which age, race, and gender has the highest percentage of inaccurate classifications?

Hypothesis

Faces of people who are older and of black women will be the most inaccurately classified in this sample.

Modeling Approach

DeepFace on Python on a selected sample of images to see how accurate it is at identifying age, gender, and race on people of different ages, genders, and races, then calculate the percentage of accurate classifications for each characteristic to see which one is most accurately identified in the sample.

Goal

To find out which ages, races, and/or genders are most inaccurately classified in order to determine if biases towards certain groups are present in AI tools and understand which area of AI we should focus on improving.

Data Acquisition and Explanation

Final Dataset

Data Dictionary

Label	Description	Possible Values	Example
age	Age of person photographed	Integer from 0 to 116	46
gender	Gender of person photographed	0 (male) or 1 (female)	1
race	Race of person photographed	Integer from 0 to 4, denoting White, Black, Asian, Indian, and Others (like Hispanic, Latino, Middle Eastern).	3
date&time	Date and time an image was collected to UTKFace	Format of yyyymmddHHMMSSFFF	20170104222158160

Acquisition

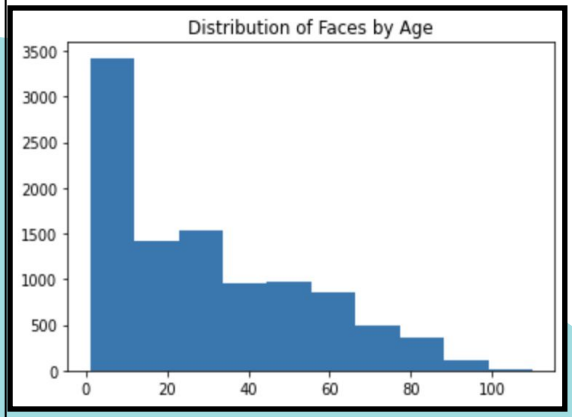
1. Gathered image data from University of Tennessee at Knoxville
2. Downloaded 10,137 images

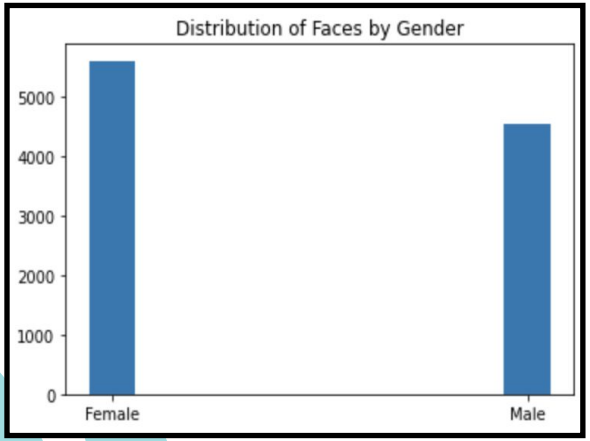
Cleaning

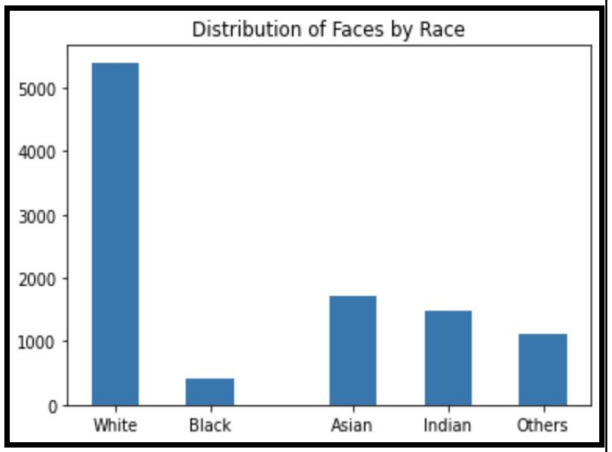
1. Removed 2 images that did not contain a label for race
2. Chose a sample of 1,000 images to work with

Analysis Plan + Justification

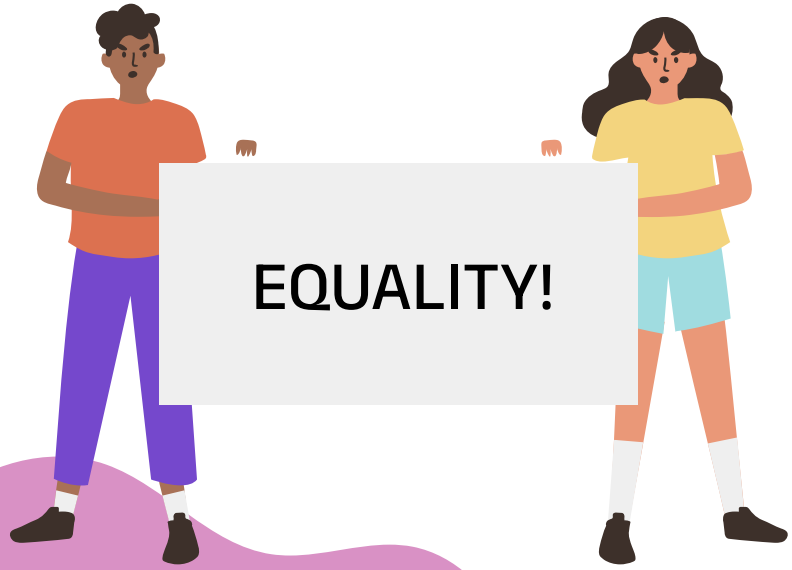
Exploratory Data Analysis







Analysis Plan



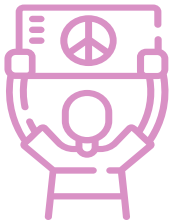
1. After doing our EDA, we were able to see the distribution of the different characteristics in our dataset
2. We ran DeepFace on a sample of 1,000 images
3. We calculated the percentage of correctly classified faces for each characteristic: age, race, & gender
    a. Age: considered accurate if it's within 5 years of the actual age

Tricky Analysis Decision

How much of the data to use

- Analyzing all 10,000+ images would have taken more than 3 hours
- Decided to take a random sample of 1,000 images with a similar data distribution as the full dataset
- We were able to run the code in around 20 minutes

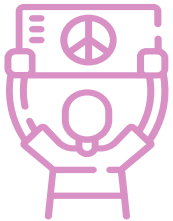
Bias and Uncertainty Validation



Limitations of the chosen sample

Our selected sample may not contain an equal amount of each level of each characteristic →

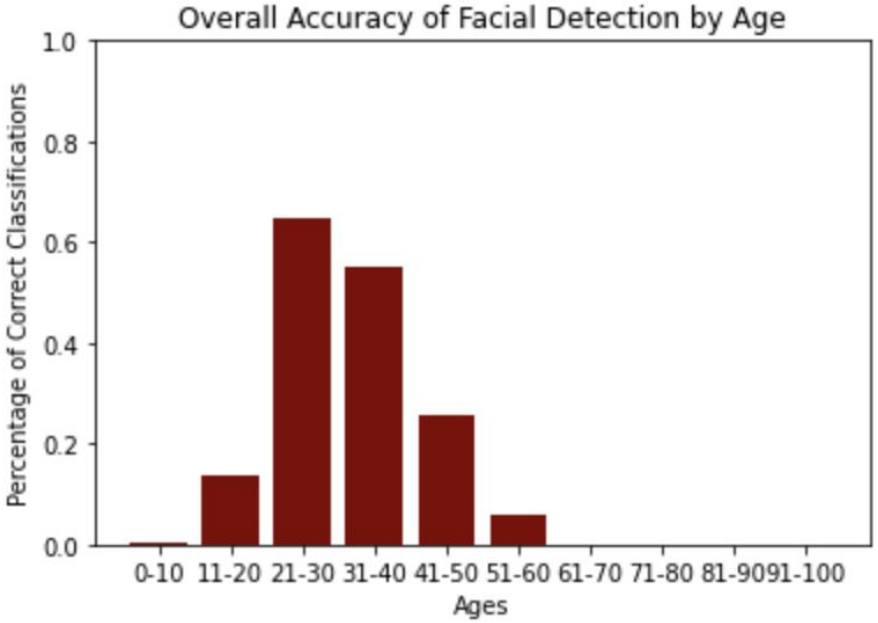
Ex: there weren't as many chances to classify a male's face compared to a female's

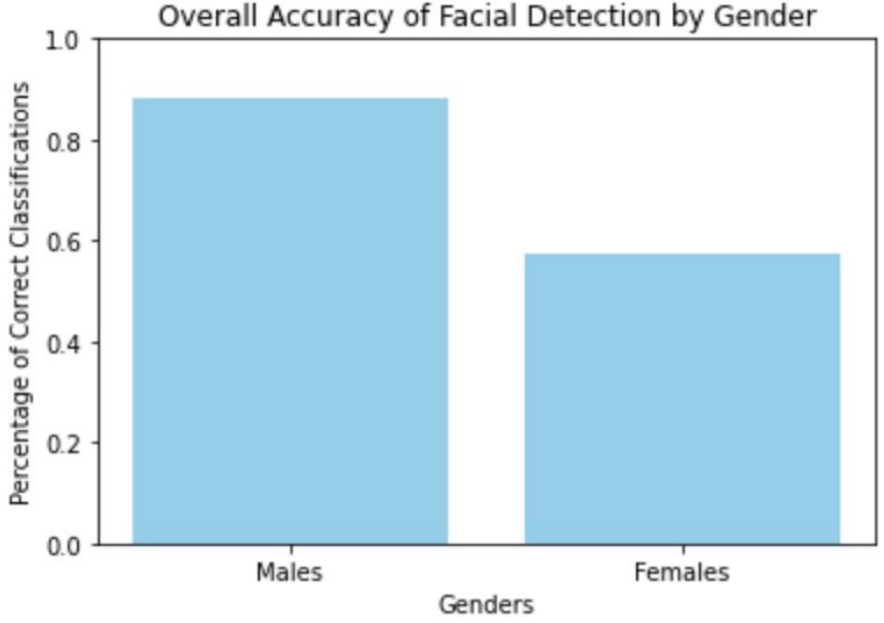


Subjectivity of accuracy of age

Our definition for age being considered "accurate" is being classified as within 5 years of the actual age → This definition can affect the percentage of accurate classifications and can alter our results/conclusions

Results and Conclusions

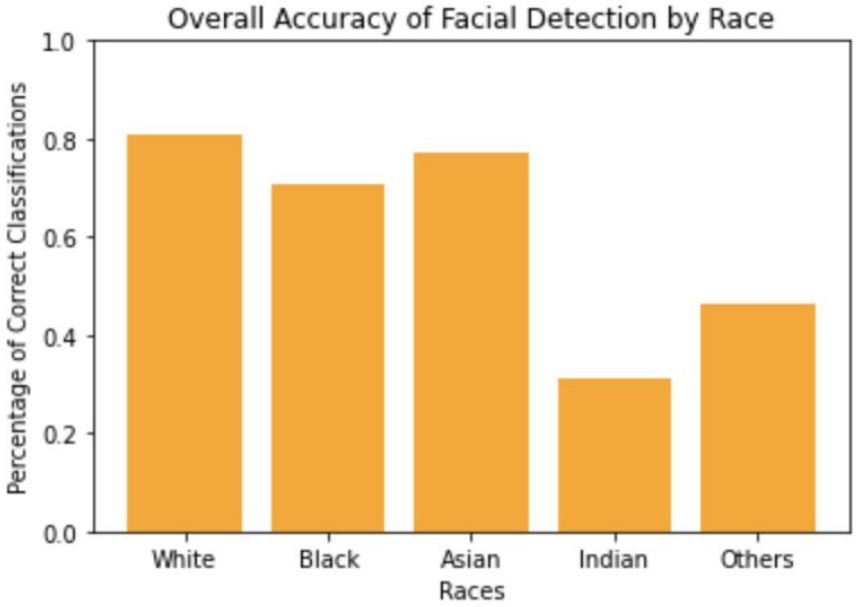


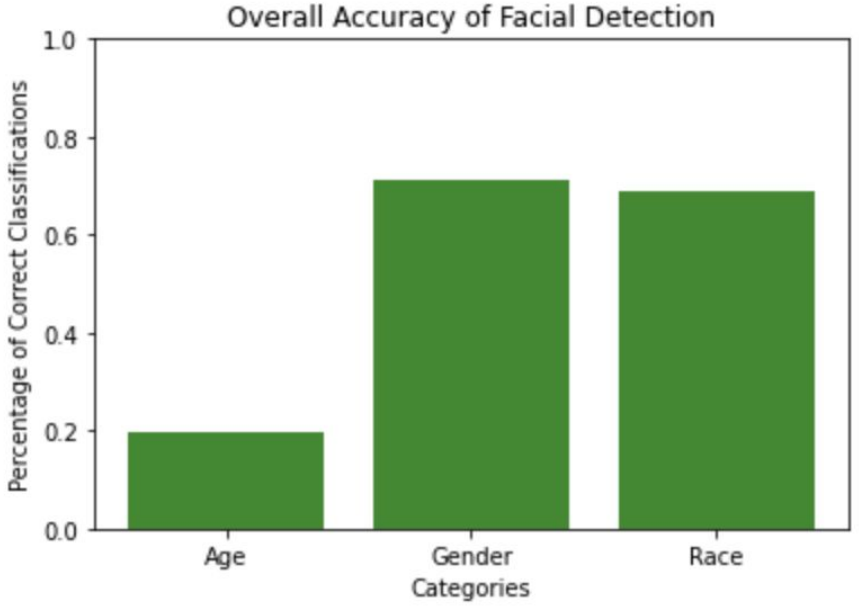


Hypothesis was **correct**: classification **accuracy decreases as age increases**

Hypothesis was **correct**: classification accuracy is **lower for females** than males

Results and Conclusions (continued)





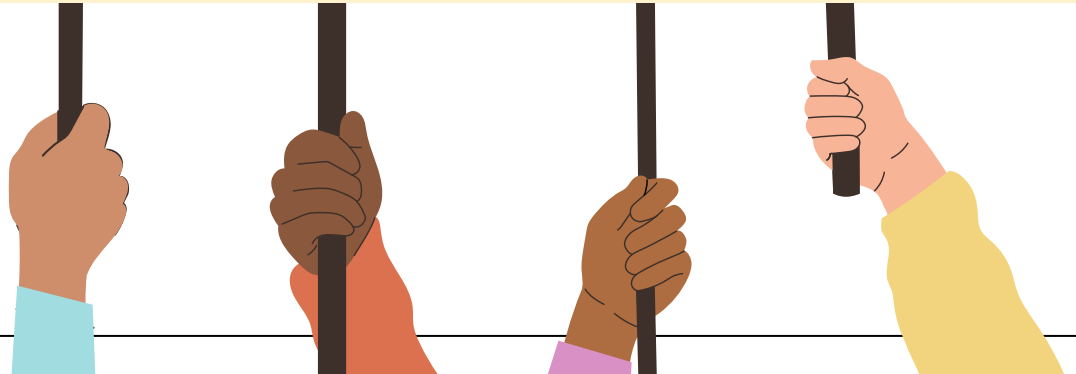
Hypothesis was **incorrect**: classification accuracy was the **lowest for Indian race**

Age was the characteristic that was most inaccurately classified of the three tested

Next Steps

1. Use different sample of images from this dataset to further test the accuracy of DeepFace classification
2. Utilize different facial recognition classification tools to compare accuracies across datasets
3. Use a sample from a dataset that contains more categories of race as opposed to the 5 that were included in this dataset
4. Conduct research regarding why it is harder to classify people who identify as female and people of older age

Thank You!



References

[1] Nada Hassanin, "Law professor explores racial bias implications in facial recognition technology", University of Calgary, Aug. 23, 2023. [Online]. Available: https://ucalgary.ca/news/law-professor-explores-racial-bias-implications-facial-recognition-technology#:~:text=In%20some%20facial%20recognition%20technology,is%20about%2035%20per%20cent.%E2%80%9D. [Accessed Apr. 4, 2024].

[2] Y. Song, Z. Zhang, “UTKFace: Large Scale Face Dataset”, *University of Tennessee, Knoxville*, n.d. [Online]. Available: https://susanqq.github.io/UTKFace/. [Accessed Apr. 5, 2024].