FACIAL RECOGNITION TO DETERMINE RACE AND ETHNICITY

Introduction:
Facial recognition software was first attempted in 1964 by a research team led by Woodrow Bledsoe. While it didn't work at the time, it has become an integral part of our society today. Some examples of its use include facial recognition software and surveillance footage. Because it is often used for security purposes, ensuring accuracy is extremely important, as a mistake can lead to a loss of privacy for a person. In our project, we intend to explore the potential racial biases in facial recognition systems. We want to see if the facial recognition system has issues identifying certain ethnic groups, and if so, are there certain features that the system struggles to recognize. With our results, we hope to learn which races facial recognition struggles with, so programmers can focus on fine tuning the system for these races. We plan to use a random forests model, convolution neural networks, and transfer learning as our model and evaluate it using accuracy, recall, and F1 Score.

Problem Statement:
Our project aims to explore various facial recognition models that can accurately predict ethnicities of people while ensuring equitable performance across a diverse range of ethnic groups. Current models often demonstrate variable accuracy and performance disparities across various ethnic groups. We hope to highlight the possible risks and biases that these inconsistencies can lead to and in doing so minimize them in our own models.

Task Definition:
The input of the model is an image of a person's face. The output is a categorical label representing the person's predicted ethnicity. Our predefined labels include White, Black, Asian, Indian, and Others (like Hispanic, Latino, Middle Eastern) which are labels of the existing dataset we are using. Our predictive facial recognition is both interesting and important because it will allow us to explore the potential biases and risks that are associated when using various models. The inconsistencies that these models produce can lead to systemic biases which can negatively affect both individuals and communities.

Algorithm(s) Definition:
Random Forests: In this model we create a "forest" of diverse decision trees, each trained on random subsets of the data and features. We can then use these trees collectively to make more accurate and stable predictions.

Convolutional Neural Nets (CNNs) are a class of deep neural networks that are particularly effective for analyzing visual imagery. CNNs leverage the spatial hierarchies in data, processing data through multiple layers of arrays. The primary purpose of these convolutional layers is to extract features from the input image, for example they detect shapes or edges in images. The features extracted can then be reasoned to make predictive classifications.

Transfer learning / tuning with pre-trained models is a technique where a model developed for a specific task is reused as the starting point for a model on a second task.

Objectives:
The main objective of the project is that, given a picture of someone's face, the model is able to predict the person's ethnicity. In addition to this, we wish to conduct further analysis into our model's results. Specifically, how does the model perform by race? Are there certain facial features that the model struggles with? In addition to reporting an accuracy metric, we wish to dive into why the model is performing in such a way.

Data Description:
For this project, we will be using the UTKFace dataset. This dataset consists of over 20,000 images, along with the individual's respective age, gender and ethnicity. All images were collected from the internet and all attributes were estimated through the Deep Expectation Algorithm and later checked by a human annotator. These photos range in pose, facial expression, illumination, and resolution. Age is measured as an integer and ranges from 0 to 116. Gender is a binary variable with values 0 (male) and 1 (female). Race is recorded as an integer as well, ranging from 0 to 4 where each number denotes a specific race (0: White; 1: Black; 2: Asian; 3: Indian; 4: Other). The last variable is a date/time variable which displays when the image was collected by UTKFace.

Methodology:
Data Mining
Initial Exploration of Photos:
Clustering
Data Summary Visualizations
Bias analysis

Algorithms
Random Forests: Ensemble method that would provide crude baseline performance to compare to more complex models

Convolutional Neural Nets (CNNs): CNNs have a long standing, successful history in computer vision and would provide a good mid-tier bench mark in terms of model complexity

Transfer Learning / Tuning with Pre-Trained Models (VGG/ResNet):
These large scale models provide an opportunity for us to tune these models rather than starting from scratch. It will be interesting to see how the "base" untuned model performs, and its underlying biases, compared to post tuning. Would also be a good idea to get a general idea of how these models were initially trained to get an idea if any biases would be present.

Criteria
Multiclass accuracy: Since 4 racial groups are present in the dataset, we would have to weight or modify the calculation of the F1 or accuracy score.

Per Class F1, Accuracy, Recall: In order to tangibly test racial bias, we can individually calculate common metrics to measure how well the model performs within each racial group.

Confusion Matrix: Confusion Matrix expanded to include all 4 groups

ROC Curve and AUC (One-vs-All): Modify ROC curve for multiclass classification rather than binary classification.

Proposed Workflow
Our first step will be to collect data through the UTKFace dataset. Next, we will preprocess the image data to use it for model training. We can do this using normalization, resizing, and some other techniques. We also want to split the data into testing and training sets with a 80%, 20% split. Then, we will extract relevant features from the facial images such as facial landmarks, texture descriptors, color histograms that can be used for ethnicity prediction. After extracting features and preprocessing the data, we then want to start training the machine learning models by using the random forest, CNN, and Transfer Learning algorithms. After tuning our hyperparameters, we can begin evaluating each of the three machine learning models using metrics such as accuracy, precision, recall, and F1-score. Using this we can determine which machine learning model fits our problem the best. We also want to analyze the performance of the models across different racial groups to identify potential biases. We can compute fairness metrics to quantity biases in the model and investigate the sources of biases and their potential impact on different demographic groups. Finally, we can interpret the results and findings from the bias analysis and document the entire project, including the data collection process, methodology, model performance, bias analysis, and ethical considerations.

Expected Outcomes:
The expected outcomes are 3 models and their respective metrics in how well they predicted a person's ethnicity. We expect that at least one model will do a significantly better job in determining a person's ethnicity, which will provide insights into what class of models are best for the facial recognition field. We can also see which model has the highest/lowest racial bias and investigate what those reasons are.

Team Members:
Andrew Martinez, Allen Choi, Jasmine Cabrera, Joshua Lee, Dylan Le, Brandon Eng

References:
Datasets:
https://susanqq.github.io/UTKFace/
https://vis-www.cs.umass.edu/lfw
(CNN's)
https://medium.com/geekculture/multiclass-image-classification-dcf9585f2ff9#:~:text=Multiclass%20image%20classification%20is%20a,discrete%20and%20continuous%20data%20respectively.
Transfer Learning with ResNet
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8933872/
(Multiclass classification transfer learning example)
https://github.com/ovh/ai-training-examples/blob/main/notebooks/computer-vision/image-classification/tensorflow/resnet50/notebook-resnet-transfer-learning-image-classification.ipynb