**Methods**

**Selecting the CCHS Cohort**

Containing 799 frontal photos over 144 patients, the CCHS photos used in this study comprise of those taken at the Center for Autonomic Medicine in Pediatrics at Lurie Children’s Hospital, both national and international patient referrals, and at the Orlando Convention in 2006. We identified each subject by sex at birth, ethnicity, location by country, date of birth, and CCHS genotype. The ethnicity for each patient is self-reported; those without a reported ethnicity were marked as having one that is unknown. Each patient’s CCHS diagnosis has been genetically confirmed. The photos picture the subjects ranging from 0 to 35 years old. In addition to the information included for each patient, each photo was affiliated with an age determined by calculating the age of the patient given their date of birth and date of the photo. See Table 1 for the ethnicity, sex, and age distribution of the CCHS cohort.

We made the assumption that all patients included in the dataset were age-advancing. There was an average of 6 photos per CCHS subject.

**Selecting the Control Cohort**

***Pre-Existing Databases***

**UTKFace Dataset.**

The UTKFace database originally consists of over 20,000 face images, annotated by sex, ethnicity, and age in photos, ranging from 0 to 116 years old. We filtered the photos to keep the ones relevant to our project, working with 15,431 face images ranging from 0 to 35 years old.

**FGNET Dataset.**

The FGNET aging database originally consists of 1,002 frontal images of 82 Caucasian male and female subjects ranging from 0 to 69 averaging eight photos per subject. Each image is annotated by patient number and age of subject in photo. We similarly filtered the photos to keep the relevant ones for our project, working with 911 frontal images ranging from 0 to 35 years old. This corresponds to roughly 11 images per subject.

***Incorporating Age-Advanced Photos to Construct the Control Cohort***

When constructing the control dataset, we considered multiple factors in order to assemble the control cohort proportionally to the CCHS cohort both by number of photos and number of subjects across each ethnic group.

**Caucasian.**

To establish proportionality across total photos and subjects for each ethnicity, it was necessary to first examine the age-advanced sub-cohort for each ethnic group. For the Caucasian group, we used the FGNET dataset. All 911 FGNET frontal images were populated and organized by age group. We picked a roughly 1:1 ratio of CCHS to age-advanced control subjects, as well as a 1:1 ratio of CCHS to age-advanced control photos. This meant that from the 82 FGNET subjects, 515 photos were randomly selected using a Python script and the Random module to match the CCHS Caucasian photo total. To ensure all 82 subjects were still included in the age-advancing cohort, the script made six passes through each of the 82 subject folders, randomly selecting a photo to add the resulting 515 photos. If a subject had fewer than six photos, a count was incremented, which was then reviewed at the end of the script to randomly select that number of photos from the other patients. The script was written such that duplicate photos could not be included in the resulting cohort. Not only did this allow us to keep a proportion of six photos per subject as seen in the CCHS Caucasian cohort, but it also allowed us to keep a random variability of number of photos per patient (where some patients have fewer than six photos and some having more than six) that is also seen in the corresponding CCHS cohort.

After selecting the FGNET cohort to use, we organized all 5,333 Caucasian UTKFace by age group. To establish proportionality across the age groups to the CCHS Caucasian cohort, we examined the proportion of the sum of selected FGNET photos and UTKFace images to the corresponding CCHS photos for each group. We then used the lowest proportion of photos (selected because it corresponded to our limitation in usable photos) and multiplied it by the CCHS photos for each age group. The sum of the photos for each age group corresponded to the total photos in the control cohort for the Caucasian group. To calculate the number of UTKFace photos to then use, we subtracted the total controls number by the corresponding selected FGNET number for each age group. The total selected UTKFace Caucasian photos also represents the distinct number of patients (minus 82 FGNET patients) in this ethnic group. The new sum of photos across each age group consists of the selected FGNET photos and the selected UTKFace photos.

**African American and Other.**

We could not find publicly available aging databases for other ethnic groups. Thus, we artificially produced aging images for these subjects using a published and well-tested method by Or-El et. al. that simulates the aging process given an image. To ensure we maintained the 1:1 ratio of CCHS to age-advanced subjects, we selected 10 for the African American group and 17 for the Other group. These original 27 images also appear in the selected UTKFace cohort so as not to double count patients. The Lifespan Age Transformation Synthesis algorithm was trained on six anchor age groups: 0-2, 3-6, 7-9, 15-19, 30-39, and 50-69. We generated a full progression of images between the anchor groups, interpolating 20 images between each. Thus, for every subject image we inputted, 101 images would be outputted.

Because we could not modify the groups to suit our needs, we developed an algorithm to match the generated photos to our own. The script determined the age for the interpolated ages by

|  |  |
| --- | --- |
|  | (1) |

where *t*, *t* + 1 represent adjacent age classes and *α* the interpolation step (in this case, 0.05). This allowed us to represent all relevant images in five age ranges: 0-2, 3-6, 7-14, and 15-29. Since each of the images were numbered from 0 to 100 in increasing order of age, we could associate each of the calculated age ranges to image numbers less than or equal to 19, 39, 59, and 79, respectively. Thus, outputted photos numbered 80 or greater were discarded. For aged subjects from the 2-3-year range, images numbered 19 or less were also discarded.

As was done with the FGNET cohort, we maintained a 1:1 ratio of CCHS to age-advanced photos for each race group. Thus, for the African American cohort, we had 9 CCHS patients to 10 age-advanced subjects and 34 CCHS photos to roughly 31 age-advanced photos. For the Other cohort, we had 13 CCHS patients to 17 age-advanced subjects and 120 CCHS photos to 142 age-advanced photos. This means we have an average of 3 photos per subject in the African American cohort and 8 photos per subject in the Other cohort. We modified the FGNET Python script to randomly select the desired number of age-advanced photos for each ethnic group while ensuring the chosen number of age-advanced subjects was kept.

To match the age groups generated by the algorithm from Or-El et. al. to ours, we corresponded our age groups that were encompassed in their predefined ones. As a result, two to three of our age groups were matched with each of theirs. To fairly divide the selected age-advanced photos amongst our age groups, we divided the number of photos in each of their age groups by the number of our age groups it encompassed. See Table 2 for an illustrative depiction of this technique.

The same method used to select UTKFace photos after age-advanced photo selection was implemented. Image numbers were slightly adjusted to ensure we could have proportionality for total photos and patients across all ethnic groups, as discussed in the next section. See Table 1 for the demographics of our control cohort.

**Table 1**

*CCHS and Control Cohort Distribution of Patients and Photos.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | CCHS | | Control | | | | |
| UTKFace | FGNET | | Artificial | |
| Subjects | Photos | Photos | Subjects | Photos | Subjects | Photos |
| Ethnicity | | | | | | | |
| African American | 9 | 34 | 139 | 0 | 0 | 10 | 31 |
| Caucasian | 91 | 587 | 1254 | 82 | 515 | 0 | 0 |
| Other | 13 | 120 | 216 | 0 | 0 | 17 | 142 |
| UNKNOWN | 31 | 58 | 0 | 0 | 0 | 0 | 0 |
| Sex | | | | | | | |
| Female | 74 | 455 | 873 | 34 | 216 | 19 | 117 |
| Male | 70 | 344 | 736 | 48 | 299 | 8 | 56 |
| Age groups (years) | | | | | | | |
| 0-1 | -- | 112 | 262 | -- | 37 | 22 | 17 |
| 2-3 | -- | 132 | 285 | -- | 53 | 5 | 30 |
| 4-5 | -- | 90 | 207 | -- | 45 | -- | 17 |
| 6-8 | -- | 125 | 278 | -- | 61 | -- | 35 |
| 9-11 | -- | 111 | 248 | -- | 54 | -- | 19 |
| 12-14 | -- | 84 | 172 | -- | 57 | -- | 18 |
| 15-17 | -- | 58 | 89 | -- | 60 | -- | 25 |
| 18-25 | -- | 70 | 60 | -- | 106 | -- | 12 |
| 26-35 | -- | 17 | 8 | -- | 42 | -- | 0 |
| Total | 144 | 799 | 1609 | 82 | 515 | 27 | 173 |

*Note.* For the UTKFace cohort, the number of subjects is equal to the number of photos.

**Table 2**

*Model for Matching Between Interpolated Image Age Groups and Our Defined Age Groups*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Our Defined Age Groups | | | | | | | | |
| Or-El et. al. groups | 0-1 | 2-3 | 4-5 | 6-8 | 9-11 | 12-14 | 15-17 | 18-25 | 26-35 |
| 0-2 | X | X |  |  |  |  |  |  |  |
| 3-6 |  | X | X | X |  |  |  |  |  |
| 7-14 |  |  |  | X | X | X |  |  |  |
| 15-29 |  |  |  |  |  |  | X | X | X |

*Note.* The number of X’s in each row is the number we divided the number of images generated in each Or-El et. al. age group by.

***Total Proportionality***

**Photos.**

Subdividing the CCHS photos by ethnicity and correspondingly by age group, we sought a roughly 1:3 ratio of CCHS photos to total control photos across the age groups for each ethnicity, as well as a roughly 1:3 ratio across the total number of photos for each ethnicity. Table 3 depicts the photo distribution by ethnic group and Figures 1, 2, and 3 depict the individual distribution for each ethnicity by age group. Unknown ethnicities were not included in the control cohort and therefore are excluded from the proportion.

**Table 3**

*Proportionality of Total Photos by Ethnic Group*

|  |  |  |  |
| --- | --- | --- | --- |
| Ethnic Group | CCHS | Control | CCHS to Control Proportion |
| African American | 34 | 170 | 0.20 |
| Caucasian | 587 | 1769 | 0.33 |
| Other | 120 | 361 | 0.33 |
| Unknown | 58 | -- | -- |
| Total | 799 | 2300 | 0.35 |

**Figure 1**

*African American Photo Distribution of CCHS and Total Control by Age Group*

**Figure 2**

*Caucasian Photo Distribution of CCHS and Total Control by Age Group*

**Figure 3**

*Other Photo Distribution of CCHS and Total Control by Age Group*

**Patients.**

We also looked to maintain proportionality between the CCHS and control patient cohorts. There is roughly a 1:15 ratio of CCHS patients to control subjects, including the age-advanced patients from the FGNET group and singly counting the UTKFace subjects who were artificially age-advanced. Unknown ethnicities were not included in the control cohort and therefore are excluded from the proportion. Table 4 depicts the patient distribution by ethnic group.

**Table 3**

*Proportionality of Subjects by Ethnic Group*

|  |  |  |  |
| --- | --- | --- | --- |
| Ethnic Group | CCHS | Control | CCHS to Control Proportion |
| African American | 9 | 139 | 0.065 |
| Caucasian | 91 | 1336 | 0.068 |
| Other | 13 | 216 | 0.060 |
| Unknown | 31 | -- | -- |
| Total | 144 | 1691 | 0.085 |

*Note.*The number of subjects for the control group is equivalent to the number of photos selected from the UTKFace dataset. In the case of the Caucasian group, the reported number of patients also includes 82 distinct FGNET subjects.

**Building and Training the Models**

***Multilayer Perceptron***

***Ridge Regression***

***Logistic Regression***

***K-Nearest Neighbors***

***Decision Tree***

***Support Vector Machine***

***Naive Bayes***

**Optimizing the Models**

***Pruning the Dataset***

***Hyperparameter Tuning***

***Landmark Reduction***

**Phase 1.**

TKTKTKTK

**Phase 2.**

TKTKTKTK