Race-Specific Convolutional Neural Networks for Accurate and Fair Juvenile Face Recognition

ECE324: Machine Intelligence, Software, and Neural Networks

April 13th, 2023

Thomas Nguyen
India Tory
Nida Copty

Table of Contents

1. Abstract

Recognising the ages of juveniles is important in many situations, however existing age estimation algorithms have difficulty achieving accurate results. Although methods for detecting age such as x-ray and dental analysis are precise, the task becomes more difficult when only an image of the face is available [1]. In this report, we propose an innovative approach that uses five race-specific convolutional neural networks (CNNs) to improve the accuracy and fairness of juvenile facial recognition algorithms. We begin by outlining the purpose of the study through a literature review of existing research in age estimation models. We then describe the proposed solution, followed by a presentation of model architecture used in the study and breakdown of the dataset employed.

The results of this research show the effectiveness of incorporating race-specific architecture when developing facial age estimation models. The Combined Race Model, which does not individually examine each race, performs worse on all metrics (including accuracy, loss, and error) when compared to the average of the five race-specific models. In addition to having favourable results, the standard deviation of testing accuracies considering each race is lower for the race-specific models showing that there is better fairness in performance. Despite this improved overall performance, the model for the Black category showed the poorest performance amongst the five race-specific models. Nevertheless, the race-specific model performed significantly better than the Combined Race Model for Black facial images.

Further research can be pursued to improve the architecture, examine the performance of models based on alternative demographics, and investigate factors that contribute to performance differences by race.

Overall, this research exposes the importance of considering race-specific architecture in the development of facial age estimation systems. The proposed approach has the potential to not only improve juvenile facial age estimation accuracy, but also contribute to the advancement of more inclusive facial recognition technology.

2. Introduction

There are various situations in which juvenile age recognition is essential to the safety, wellbeing, and livelihood of society. Some of these important situations include preventing underage social media use, enforcing alcohol and substance age restrictions, investigating child sexual abuse (CSA) cases, etc.. Although there are accurate methods of detecting age involving x-ray and dental analysis, the task becomes far more challenging when only an image of the face is available [1].

Despite having various cues to help people predict someone's age, including wrinkling, hair colour, and facial structure, humans are not able to accurately estimate age from face images. For example, people tend to overestimate the age of a person who is smiling versus a person with a neutral facial expression [2]. In a study by Wilkinson and Ferguson, adults were only able to guess the age of caucasian youth between ages 0-16 with 33% accuracy. This study also found that neither working with nor living with juveniles improved the adult's ability to predict their age[1].

Considering the important scenarios that require accurate juvenile age estimation from photographs, coupled with how inaccurate people are at guessing juvenile ages, there is a need for machine learning models that correctly estimate juvenile ages based on face images. Although there are existing models already, they do not perform well enough considering the importance of the use cases [3].

The authors of this report are proposing an innovative model that could revolutionise the subject of age classification and pave the way for more reliable and accurate outcomes. This improved model involves utilising multiple CNNs that are individually trained to consider a single racial group. As a result, the model will be able to consider the unique ageing patterns of different race groups and therefore produce more accurate results..

The goal of this report is to offer an overview of the process taken to develop this innovative age-detection model. First, the report will outline a literature review that was conducted to identify the gap that this model fills and offer evidence for why the model should have improved performance. Next, the report will provide a basic overview and description of the proposed model. Then, the report will describe the implementation and training process used as well as an analysis of the results. Finally, the report will outline the impacts and implications of this model.

3. Literature Review

3.1 Work Considering Impact of Gender Information on Age Estimation

3.1.1 A Study on Automatic Age Estimation using a Large Database

The goal of this paper was to use a large database to study problems relating to human age estimation from face images. Within the study, the authors considered the impact of gender on age estimation. The authors ran a test in which they preceded the age estimation model with gender classification models. The authors saw an improvement in age estimation accuracy when the model was also considering gender information. Additionally, the age estimation error was lowest with the most accurate gender classification model [3]. This shows evidence that age estimation is improved when the model is also given accurate gender information.

3.1.2 Age Estimation Using Gender Information

Understanding the unique ageing patterns of men and women, the authors recognized the opportunity to improve age range categorization using gender information. Neural networks were used to create two separate age estimators, one for male and one for female images. By using two separate age estimators, each model was able to learn the unique ageing pattern of each gender. As seen in the table below, there was a notable increase in the accuracy of age estimation when it followed accurate gender classification [4].

Methods	**Set A** (%)	**Set B** (%)
Gender classification	92.523	87.850
Age estimation	72.274	72.274
Gender classification followed by age estimation	77.259	74.455

Table 1: Results of Study Considering Age Estimation Using Gender Information

3.1.3 A Multifeature Learning and Fusion Network for Facial Age Estimation

Deng et al. recognized that most existing research on facial age estimation has not considered how demographic features impact human ageing patterns. Two such features are gender and race. For example, females tend to look younger than males in the younger age range due to different changes during puberty. The authors used three different subnetworks to predict age, race and gender. The results of these three classifications are later used by a regression and ranking estimator to estimate the age. To evaluate the model's performance, the authors compared their results with other single-feature and multi-feature-based state-of-the-art models. The mean absolute error of the proposed model was better than single-feature-based models and comparable to other well-performing multifeature-based models, suggesting that the overall performance of multifeature-based methods were better than single-feature-based methods [5].

3.1.4 Conclusion

Considering the papers above, there is evidence for a relationship between gender information and age estimation. In both studies, the authors found that knowing gender can improve age estimation accuracy. This shows that models trained to consider unique ageing patterns, rather than trying to generalise multiple ageing patterns, perform better.

3.2 Work Considering Unique Aging Patterns by Ethnicity

3.2.1 Racial and Ethnic Differences in Self-Assessed Facial Aging in Women: Results From a Multinational Study

This paper examines racial and ethnic differences in self-assessed facial ageing among women in different countries. The study analysed data from a multinational survey of over 3,000 women aged 18 to 75 from the United States, Canada, the United Kingdom, and Australia. The results showed that the impact of race/ethnicity was significant in differences in self-assessed facial ageing, with Black women reporting the least amount of facial ageing and Caucasian women reporting the most. Over 30% of Black women did not indicate the presence of moderate to severe facial ageing until they were between the ages of 60 to 79. In contrast, for most Hispanics and Asians, this did not occur until they were between 50 to 69 years old. For Caucasians, it typically did not occur until they were between 40 to 59 years old [6].

3.2.2 Ageing Differences in Ethnic Skin

This paper explores the differences in the ageing process of ethnic skin, focusing on four major ethnic groups: African American, Asian, Hispanic, and Native American. The study highlights that ethnic skin ages differently from Caucasian skin due to variations in skin structure, physiology, and melanin content. Ethnic skin is characterised by a thicker dermis, more abundant collagen, and greater melanin protection from UV radiation, which leads to delayed wrinkle formation and less severe photoaging (sun damage). However, ethnic skin is also prone to hyperpigmentation, hypopigmentation, and keloid formation [7].

3.2.3 Conclusion

Both papers in this section determine that humans have different facial ageing patterns based on their race and ethnicity. Considering the conclusion from Section 3.1, stating that age estimation models that consider the unique ageing patterns of men and women performed better, this suggests that age estimation models that take into account different races have the potential to be more accurate compared to age estimation models that are generalised for every race.

3.3 Work Considering Juvenile Age Estimation

3.3.1 Apparent Age Estimation from Face Images Combining General and Children- Specialized Deep Learning Models

This paper uses transfer learning to develop a model that more accurately accomplishes juvenile age detection. The authors first created a general CNN that was trained on images of people ages 0-99 using the IMDB-Wiki dataset. The authors noted that the dataset had very few images of ages 12 and younger, so manually collected 5723 images of children between ages 0-12. They then used the information learned by the general model as a starting point for training a child-specialised model using the 5723 juvenile images. The results show that the combination of general and children-specialised deep learning models improves age estimation accuracy, particularly for younger age groups [8].

3.3.2 Age Estimation in Juveniles using Convolution Neural Network

This paper emphasises that most of the existing estimation models have performed significantly better on adults compared to juveniles, so their research focuses on building a CNN-based architecture for the age estimation of juveniles. This has various applications, such as finding lost children, surveillance monitoring, face classification, and managing access to unwanted content for children. Sharma et al. used 90,000 face images of people with ages in the range of 0-20 to train their model. The architecture consisted of four convolutional layers and ReLU as the activation function. The model was especially successful for ages 0-8 as the predicted ages differed at most two years from the actual age. The model performed weaker for ages 8-20 as it tended to underestimate the age [9].

3.3.3 Conclusion

Considering these two papers, there are two key takeaways that will be applied to the model. Firstly, since both papers had success using CNNs, the model's architecture will involve a CNN. Secondly, despite focusing on juvenile age estimation, the model will be trained on a larger age range. Considering the results seen in Section 2.3.1, training the dataset on a broad range of ages but tuning it for juveniles should increase the accuracy of juvenile age estimation.

4. Implementation Process

4.1 Proposed Model Description

Considering the literature review in Section 3, a gap was identified in the current research. Prior work in Section 3.1 has shown that multi-featured models that take into account the gender information of the subject, rather than generalising for both females and males, perform better in age estimation than single-featured models. This is because males and females have different facial ageing patterns, such as the growth of facial hair and the formation of wrinkles. Similarly, studies in Section 3.2 show that facial ageing patterns also vary by race and ethnicity. Given that

gender-based multi-feature age estimation has improved performance due to the unique ageing patterns of men and women, coupled with the fact that ageing is also race dependent, the following key gap was identified: there is an opportunity to design a method that involves multiple unique models to consider age estimation for each race. Consequently, this paper is proposing a race-based multi-feature age estimation model. The goal of this report is to find an improvement in age estimation accuracy when comparing the race-based multi-feature model to a single-CNN-based architecture.

To accomplish this, five different CNNs will be trained corresponding to one of the four categories of race. The five categories of race (Black, East Asian, Indian, Caucasian, and Other) were chosen based on the dataset labelling. By having five different CNNs that are each specifically trained for estimating age of one of the racial groups, there is potential to see an even more significant increase in accuracy than prior models that considered gender.

Our solution involves three main steps. First, the input image will be classified into one of the five categories of race (Black, East Asian, Indian, Caucasian, or Other). This image will then be fed into the corresponding race-specific CNN which will output an age estimate. Finally, the age will be placed into one of the corresponding 17 age bins. The bin distribution is shown in Figure 1. For ages 0-10 and 25-80, the bin size is five years. From age 11-25, the bins are either three or four years in size. This distribution of bins, with smaller bins in the age range of focus, was designed to require increased accuracy from ages 11 to 25.

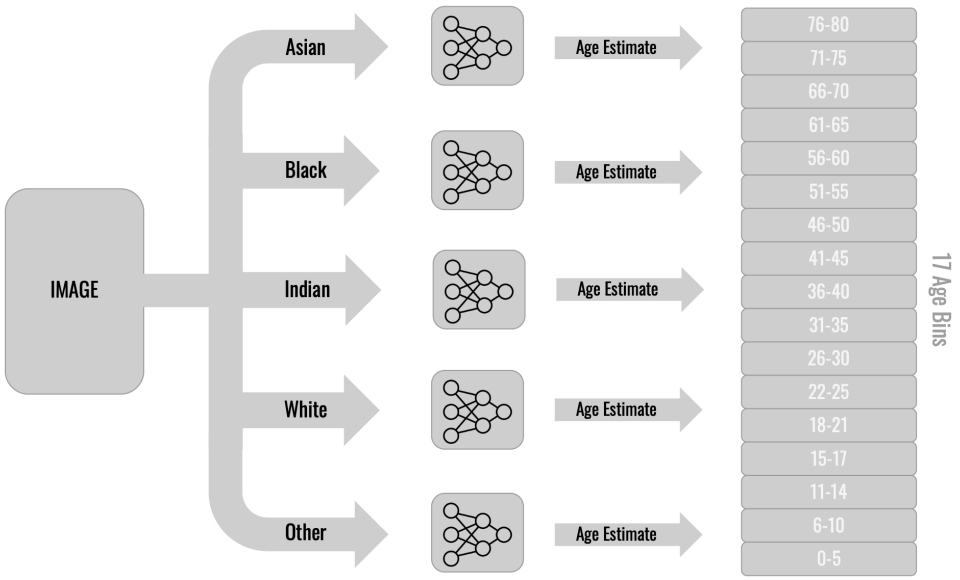


Figure 1: Schematic of Proposed Multi-Feature Model

CNNs are commonly used in machine learning applications involving image recognition because they can automatically learn and extract the relevant features without requiring explicit feature engineering. The CNN will use convolutional layers to scan and filter input data, allowing it to learn spatial and temporal relationships between facial features. The model will also employ

pooling layers to downsample the output and reduce the number of parameters in the network, allowing the CNN to process thousands of data points.

Further, since the focus is on age estimation for juveniles, a dataset that consists mostly of ages 10-25 will be used. Prior work has shown that combining a youth-specific model with a more general model that considers a wider age range has improved accuracy, especially for the youth [8]. Consequently, the training dataset will consist of face images of subjects from ages 1-80 while having a higher proportion of the images falling in the 10-25 age range.

Finally, this model's performance will be tested against a Combined Race Model. The Combined Race Model will have the same architecture as the five unique models, but will be trained on a model including images from all race groups. The combined dataset used includes 5000 images total, equally representing all five race categories.

4.2 Data Collection Process

A two-stage data collection process was implemented for this project.

The first stage involved gathering data using photos of the authors, their friends, and their family. A total of 450 face images were collected. Gathering this data involved finding photographs of people, cropping the faces, and labelling the image with the person's age, gender, and race. The age was determined based on the date the photograph was taken relative to the person's date of birth. Race was determined based on the knowledge of the person, their family, and their ethnic background.

The second stage involved using images from the internet. Since the dataset requires accurate age, gender, and race labels, large numbers of images could not be scraped from platforms such as Google Images or Flickr. Consequently, the authors were unable to generate a dataset large enough to accurately train a model using only images they had gathered. To resolve this issue, the dataset was supplemented with images from public datasets online.

Considering online datasets, very few are labelled with the information required to train the new model. Most of the images were collected from UTKFace. UTKFace is a database with 20,000 images of people from ages 0 to 116. Each image is labelled with age, gender, race, and time that the photograph was collected for UTKFace. In addition to UTKFace, the data was supplemented with images from the FaceARG dataset. FaceARG is a dataset with 175,000 images that are labelled with age, race, and gender.

The overall dataset was collected in such a way that it accomplishes two goals. First, the dataset has improved consistency with respect to the number of datapoints in each race category. When initially only using the UTKFace dataset, there was a significantly disproportionate number of images in the Caucasian category. After parsing through the images in FaceARG, the dataset became more balanced by adding additional images to the Black, Asian, and Indian categories. By combining images from FaceARG and UTKFace, a final dataset was created with between 5000-8000 images for each of the White, Black, Asian, and Indian categories. Unfortunately,

FaceARG did not have an ‘others’ category for groups such as Hispanic, Latino, and Middle Eastern. As a result, the number of data points in this category was significantly lower than the others. The final breakdown of the dataset by number of images and categories can be seen in Figure 2 below.

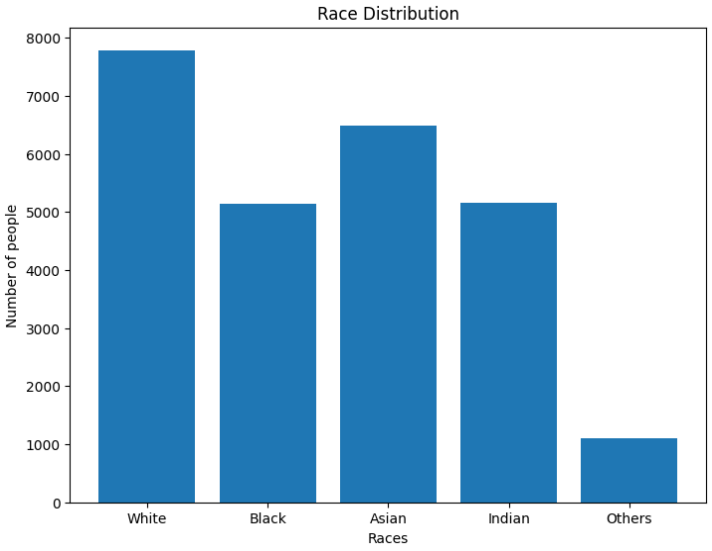


Figure 2: Dataset Breakdown

The second goal accomplished by the dataset was to have a youth-focused age distribution. Since the model is being specifically designed for juvenile age estimation, it will be specifically trained to perform best on images of people aged 10 to 25. Based on findings from Section 3.3.1, the model will be trained on a dataset containing images from age 0 to 80. Within this range, however, there are more images in the range of focus. The data was parsed to maximise the number of images in the range of 10-25, including less images as the age grew older. Graphical depictions of the age distributions for each unique race category can be found in Figures 3-7 below.

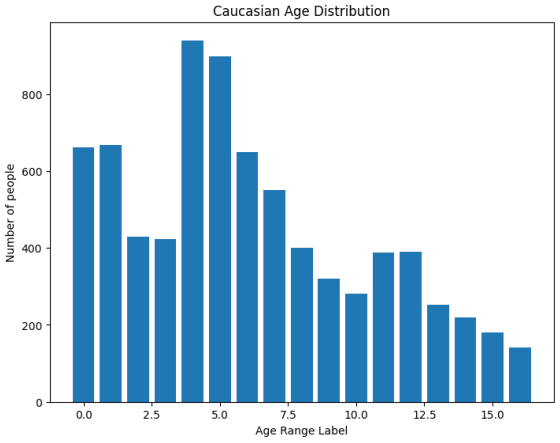


Figure 3: Caucasian Age Distribution

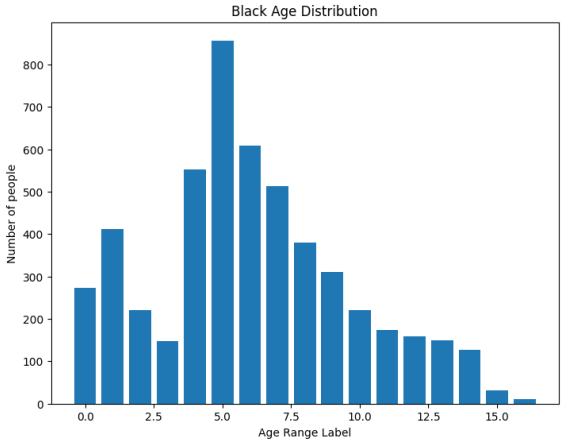


Figure 4: Black Age Distribution

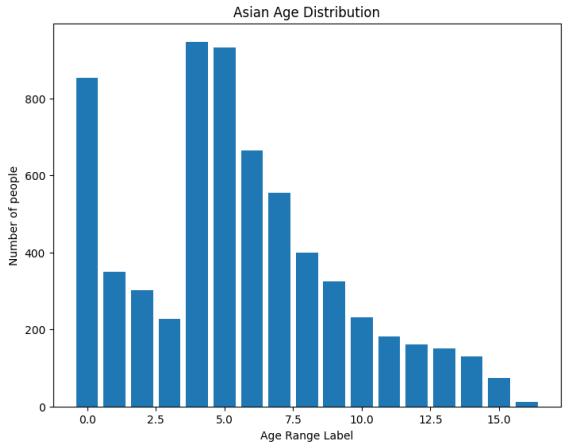


Figure 5: Asian Age Distribution

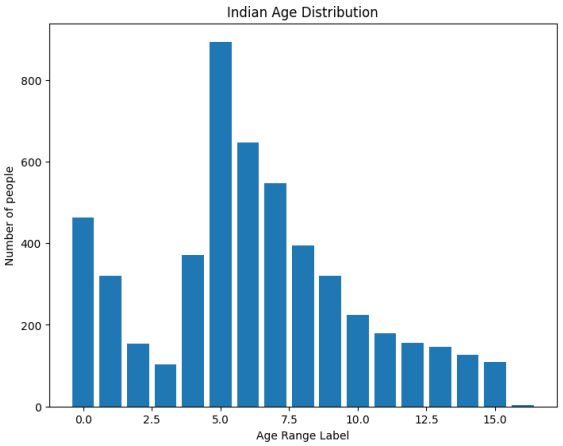


Figure 6: Indian Age Distribution

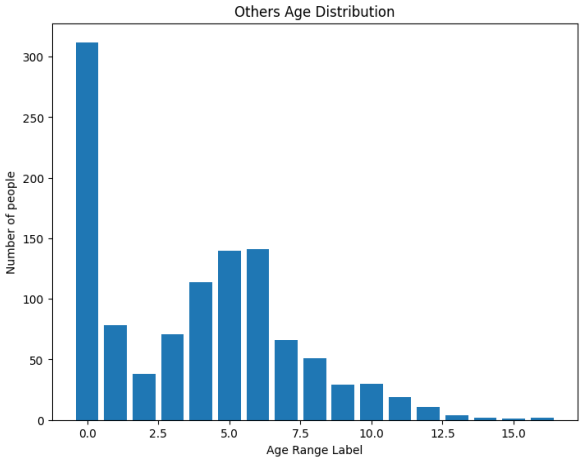


Figure 7: Others Age Distribution

4.3 Model Implementation and Training

Given an image size of 3 x 32 x 32, considering the three RGB colour channels, the architecture of the CNN is as follows:

- First convolutional layer of input size 3, output size 6, and filter size 3
- First pooling layer of size 2 by 2
- Second convolutional layer of input size 6, output size 16, and filter size 3
- Add a padding of 1 dimension
- Second pooling layer of size 2 by 2
- Third convolutional layer of input size 16, output size 24, and filter size 3.
- Add a padding of 1 dimension
- Second pooling layer of size 2 by 2
- Flatten the tensor by reshaping it to a 1-dimensional vector
- First fully-connected layer of input size 24 * 3 * 3 and output size 120
- Second fully-connected layer of input size 120 and output size 84

- Third fully-connected layer of input size 84 and output size 36
- Fourth fully-connected layer of input size 36 and output size 17

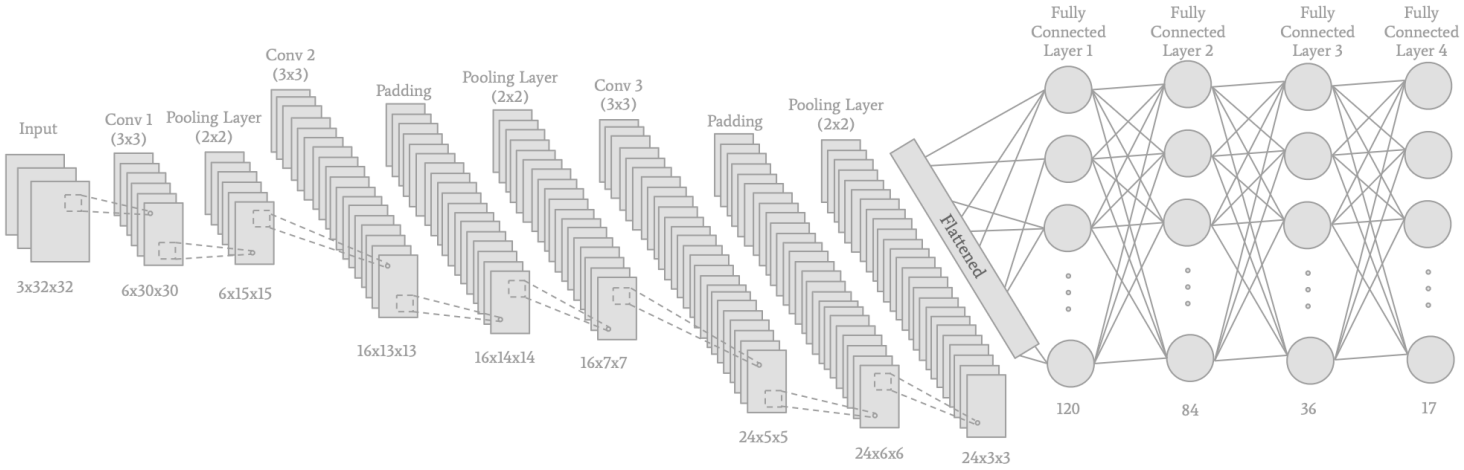


Figure 8: Diagram of Convolutional Neural Network

There are various aspects of the model architecture to consider. First, the architecture involves three convolutional layers, each with increasing numbers of channels. This allows the network to learn complex hierarchical features of the facial images, capturing both low- and high-level features. Effectively capturing the features is very important for our model, since it needs to predict age based on slight variations on facial features. The max pooling layers are used to reduce the spatial dimensions and therefore make the feature extraction more efficient. Max pooling was chosen, rather than average pooling, because selecting the largest value in the pooling windows helps it preserve sharp features from the images which can lead to better feature discrimination. The model also has four fully connected layers. Initially, the architecture was designed with three fully connected layers however a fourth was added based on the complexity of the task at hand. Adding depth to the network allows it to capture more abstract representations of the data which is important due to the complexity of the inputted face images, however further depth was not implemented in order to prevent overfitting and ensure enough information was preserved. Finally, the architecture's output size is 17, which matches the 17 age bins involved in the classification.

When training the architecture described above, there were several hyperparameters to tune to achieve the desired performance. The hyperparameters were tuned to be the following:

Number of Epochs	=	40
Batch Size	=	120
Learning Rate	=	0.001
Loss	=	Cross Entropy Loss
Activation Function	=	Rectified Linear Unit (ReLU)

Tuning the number of epochs involved monitoring the training and validation losses during training. This was conducted both manually, by intentionally changing the number of epochs based on the result, as well as by implementing early stopping. The batch size was tuned manually by experimenting with different batch sizes and observing the resulting performance. The learning rate was also tuned manually, following the same method used for batch size tuning. Step decay and exponential decay learning rate schedulers were tested in the tuning process, however both were found to lead to overfitting. Cross entropy loss was chosen due to the nature of the classification task at hand, involving each input only belonging to one distinct, predefined class. Cross entropy loss informs optimisation by evaluating the disparity between the predicted class probabilities and true class labels. Finally, ReLU was chosen for the activation function. ReLU was chosen due to its ability to capture non-linear patterns. This is relevant for image classification because images typically contain non-linear patterns, such as textures.

5. Results

The results of the model training are shown in Figures 9 to 11. Additional results can be found in Appendix A.

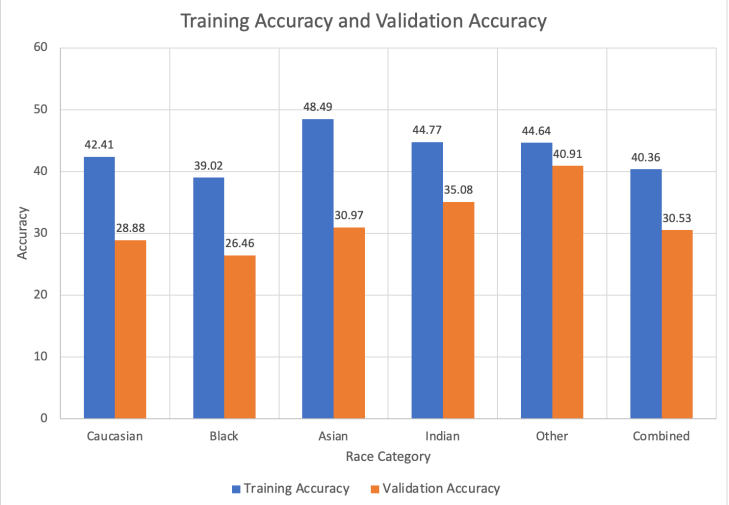


Figure 9: Graph of Training Accuracy and Validation Accuracy

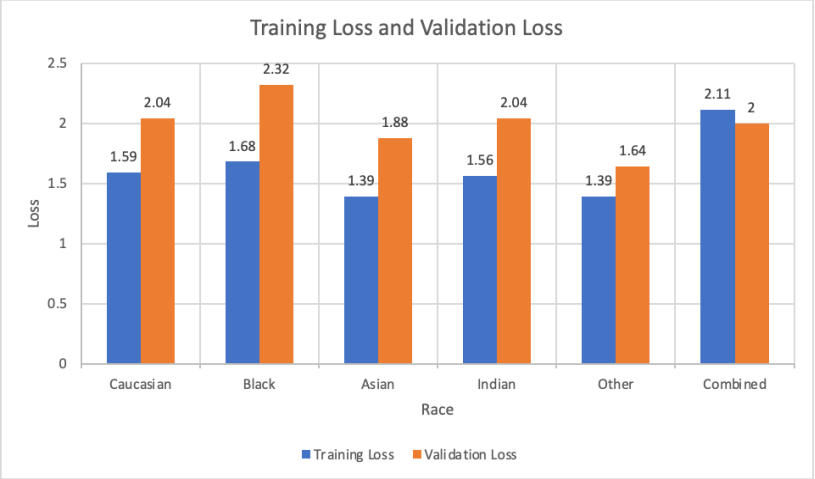


Figure 10: Graph of Training Loss and Validation Loss

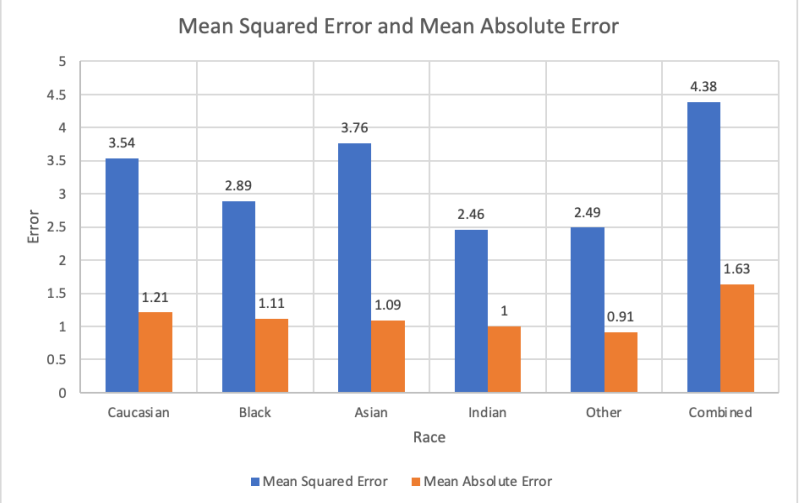


Figure 11: Graph of Mean Squared Error and Mean Absolute Error

5.1 Discussion

The first results to consider are for the Combined Race Model, the model that did not consider each race individually. This model performed worse on all metrics when compared against the average of the five race specific models. As seen in Table 2, it had lower training and validation accuracy, larger training and validation loss, and larger mean squared error and mean average error. These results offer empirical evidence that the race-specific model was able to achieve improved accuracy, reduced loss, and reduced error when compared to a model with the same CNN architecture that is not trained to specifically consider the subject's race. This supports the hypothesis that age recognition models that consider the unique ageing patterns of different racial groups perform better.

Table 2: Comparison of Combined Race Model vs Average of 5 Race-Specific Models		
Performance Metric	**Average of 5 Race-Specific Models**	**Combined Race Model**
Training Accuracy	43.87	40.36
Validation Accuracy	32.46	30.53
Training Loss	1.52	2.11
Validation Loss	1.98	2.00
Mean Squared Error	3.03	4.38
Mean Absolute Error	1.06	1.63

Race-specific testing accuracies were computed for each race for the Combined Race Model. These testing accuracies showed the Combined Race Model performed worse in every single race category compared to the race-specific models, as seen in Table 6.21. Additionally, the standard deviation of the testing accuracies of each race tested on the Combined Race Model (3.57) and the standard deviation of the testing accuracies of each race tested on their corresponding race-specific model (3.38) were computed. The fact that the standard deviation of the testing accuracies of each race being tested on their race-specific model is lower than the standard deviation when they are tested on the Combined Race Model shows that the testing accuracies are closer to each other when the models are trained for a specific race. This suggests that a model with race-specific CNNs is not only more accurate, but also more fair in the sense that a particular race does not perform a lot better or worse than the others.

Although the proposed model was more fair, it is important to note that certain race-specific models still performed worse than others. When looking at training and validation accuracy, the model for the Black category had the poorest performance. The model for the Black category also had the highest training and validation loss. Although this performance was worse compared to the four other race-specific models, there was improvement in the performance of the black-specific model for Black individuals compared to the Combined Race Model. The combined race model had a testing accuracy of 20.79 on Black face images while the black-specific model had a testing accuracy of 28.68. Considering these results, the proposed model was able to achieve a significant improvement of about 28.7%. Although the performance of the black-specific model is not desirable, it is aligned with existing inequity in face recognition algorithms. According to an article by Harvard, a study found that 189 face recognition algorithms performed least accurately on women of colour [11].

5.2 Evaluation of Performance

It is not straightforward to compare the performance of different machine learning-based facial age estimation models. There are several features, such as type of the model (regression or classification) or the number of classes, that heavily affect the performance metric that is used.

For example, the comparison of exact accuracies for classification tasks is not a good indicator of model performance given that the number of bins and their corresponding sizes may vary notably, having a significant impact on accuracy values. For this reason, mean absolute error (MAE) was chosen as the metric to compare the proposed race-specific model with state-of-the-art studies in the area of facial age estimation.

As shown in Table 2, the MAE of the proposed model was found to be 1.06. Note that this value indicates that, on average, the model predicts the age to be 1.06 bins away from the actual bin. Considering the differences in bin size, the weighted average of this model's bin size is calculated to be 4.60. This means that the model, on average, predicts the age to be $1.06 * 4.60 = 4.90$ years away from the actual age. For the age range 10-25, which is the study's age range of focus, the weighted average of the bin size is 3.55, which means that the model predicts $3.55 * 1.06 = 3.78$ ages away from the actual age. By converting the MAE of the number of bins to MAE of the ages, a comparison can be made between the proposed race-specific model and state-of-the-art regression models.

Table 3: MAE values for different age estimation regression models and our model

Authors	**Dataset**	**MAE**
Wang et al. [12]	FG-NET	4.26
	MORPH-II	4.77
Taheri et al. [13]	FG-NET	3.08
	MORPH	2.81
Niu et al. [14]	MORPH-II	3.42
Rothe et al. [15]	LAP	5.007
Proposed model for all ages	UTKFace + FaceARG + author's data	4.90
Proposed model for ages 10-25	UTKFace + FaceARG + author's data	3.78

The values shown in Table 3 demonstrate that the proposed model is comparable to other state-of-the-art models. When considering the range from ages 10-25, the model performs especially well. When comparing the models in Table 3, it should also be noted that the dataset used for the proposed model was usually much smaller in size than the datasets used in other works. This suggests that the use of a larger dataset may further improve the performance of the proposed model significantly.

6. Impact and Implications

This age recognition technology can be applied to a wide variety of fields and scenarios. Examples of applications, as given in the introduction, include preventing underage social media use, enforcing alcohol and substance age restrictions, and investigating child sexual abuse (CSA) cases. Each of these examples offers scenarios in which this technology can be applied for the benefit of society, ensuring the safety and wellbeing of juveniles. Further examining the example of preventing underage social media use, social media services such as Instagram or Snapchat could utilise age recognition technology in their registration process to ensure that the person behind the screen is of the correct age. Considering the example of CSA, accurate age estimation is essential to the forensic investigation of the crime. Determining the age of juvenile subjects in photographs can be not only used to help identify the victim but also determine the sentence given to the offender [1].

In addition to these benefits, there are also many risks that must be considered. Firstly, the technology could be used to perpetrate ageism. Ageism is a form of discrimination based on someone's age. In Canada, people are legally protected for ageism from age 18 and above [16]. An example of how ageism could be used would be if applicants for a job were rejected based on their predicted age being too young or too old. A second example is the potential use in surveillance and mass monitoring. Age recognition technology could be used to improve monitoring processes by governments, organisations, or even individuals, which could infringe on people's rights of privacy. A final risk is the manipulation of the technology. Since the model is trained based on the ageing patterns of different racial groups, people could identify specific features that signify different ages and change their appearance to deceive the age recognition software.

It is also important to note the ethical implications of the model's decreased performance on Black facial images. Although the race-specific model did achieve improved performance on Black facial images, it still perpetuates systemic bias in machine learning. As previously stated, a study found that 189 facial recognition algorithms performed worst on dark skinned females. Currently, there are several methods being explored to build a more equitable face recognition landscape. Several reasons for this inequity have been identified, including that default camera settings not being optimised to capture Black skin tones and therefore producing lower-quality images of Black people [11]. Further research into the causes of and solutions for this inequity could improve societal outcomes of this model.

It is essential to be aware of the potential risks of the technology and harmful implications of malignant applications. In order to prevent misuse, research should be conducted to investigate current ways in which age recognition technology is being misused in society. With a better understanding of this, future models could be specifically designed to prevent these uses. Additionally, an understanding of current misuses might help researchers identify potential future areas of misuse and preemptively determine prevention methods.

5. Conclusion

In conclusion, the research presented by this report demonstrates that race-specific convolutional neural networks (CNNs) have improved accuracy for juvenile facial age estimation. The results show the race-specific model was able to achieve improved accuracy, reduced loss, and reduced error when compared to a model with the same CNN architecture that is not trained to specifically consider the subject's race. Additionally, the race-specific model was shown to have comparable mean average error compared to existing state-of-the-art models. These findings contribute to the current literature on facial age estimation models, emphasising the importance of considering race-specific facial ageing patterns while developing accurate models. This supports the idea that a one-size-fits-all approach is not good enough when considering something as unique as facial ageing. As a result, a single-feature-based model may not be effective in facial age recognition algorithms. These findings also show that considering racial diversity in both the training data and model architecture can lead to more fair machine learning systems.

There are various avenues for further research based on the findings of this study. Firstly, further research can be conducted on improving the architecture proposed by this model. Additional modifications to the architecture might be able to achieve improved performance. Second, the performance of a model that is trained uniquely on both gender and race could be investigated. Considering that model performance is improved when examining each demographic individually, it could be proposed that a model considering the two demographics together could achieve higher accuracy. Finally, research should be conducted into why there are differences in performance among different racial categories. By identifying the cause of these differences, future research might be able to design different model architectures or collect different datasets that can overcome these challenges and further improve fairness.

Overall, the race-specific CNNs approach proposed by this report is a promising next step towards advancing facial recognition technology to be more accurate and inclusive. Additionally, this approach brings society closer to achieving exceptionally accurate juvenile age estimation that can be used in important applications, such as investigating child sexual abuse cases. By designing this race-specific model, the authors strive to develop technology that can accurately protect juveniles of every race.

6. Appendix

Table 6.1: Model Evaluation Data

Race	**Training Accuracy**	**Training Loss**	**Mean Squared Error**	**Mean Absolute Error**	**Validation Accuracy**	**Validation Loss**
Caucasian	42.41	1.59	3.54	1.21	28.88	2.04
Black	39.02	1.68	2.89	1.11	26.46	2.32
Asian	48.49	1.39	3.76	1.09	30.97	1.88
Indian	44.77	1.56	2.46	1.0	35.08	2.04
Other	44.64	1.39	2.49	0.91	40.91	1.64
Combined	40.36	2.11	4.38	1.63	30.53	2.00

Table 6.2: Testing Accuracies for Each Race Tested on the Combined Model and the Race-Specific Model

Race	**Testing Accuracy (%) on the Combined Model**	**Testing Accuracy (%) on the Race-Specific Model**
Caucasian	24.98	30.18
Black	20.79	28.68
Asian	30.85	36.91
Indian	26.12	34.57
Other	29.63	36.69
Average	26.47	33.41
Standard Deviation	3.57	3.38

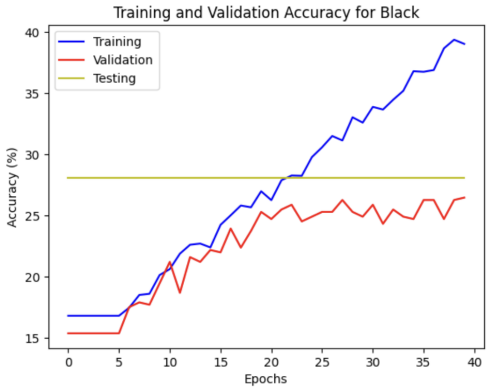


Figure 6.1: Training and Validation Accuracy for Black Category

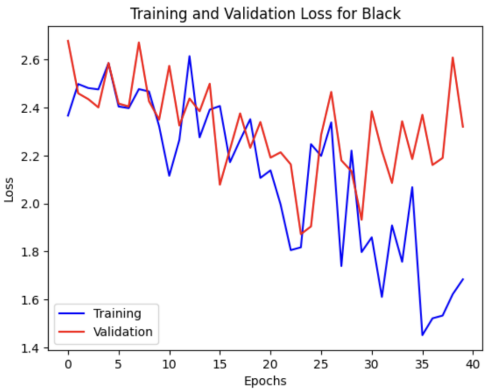


Figure 6.2: Training and Validation Loss for Black Category

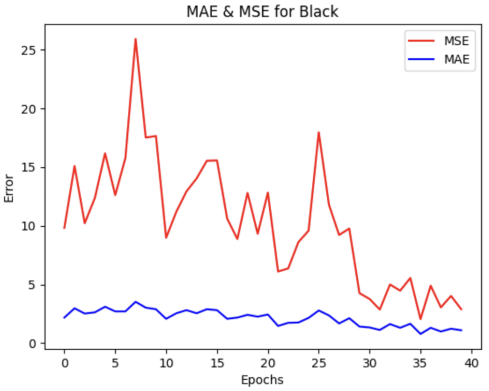


Figure 6.3: Mean Average Error and Mean Squared Error for Black Category

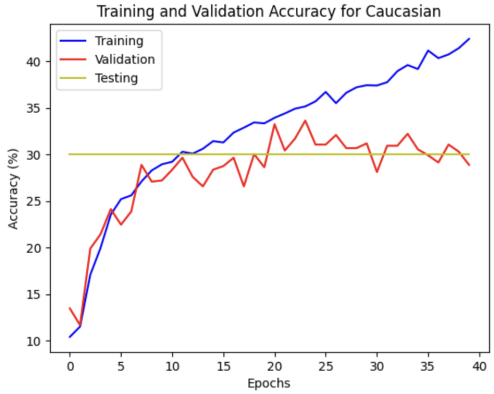


Figure 6.4: Training and Validation Accuracy for Caucasian Category

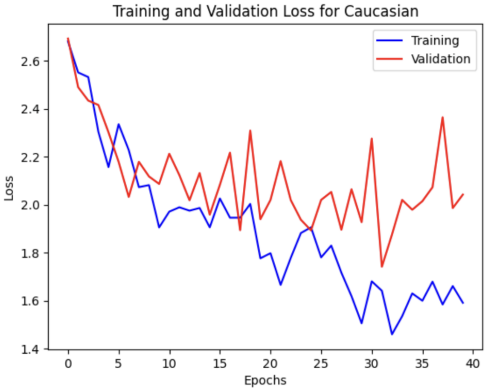


Figure 6.5: Training and Validation Loss for Black Category

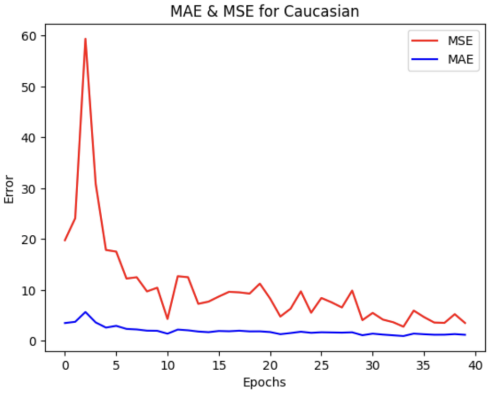


Figure 6.6: Mean Average Error and Mean Squared Error for Caucasian Category

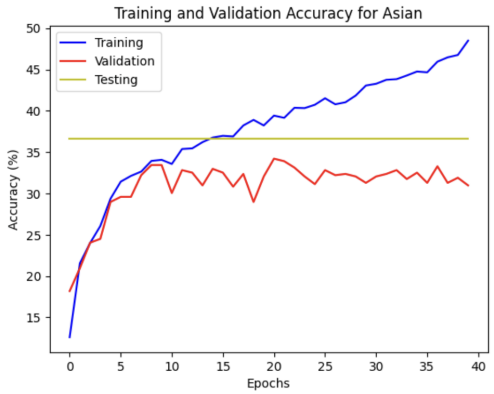


Figure 6.6: Training and Validation Accuracy for Asian Category

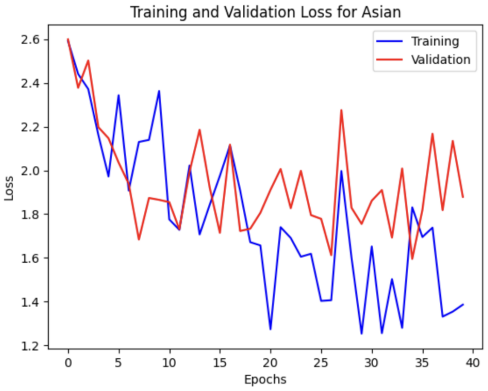


Figure 6.7: Training and Validation Loss for Asian Category

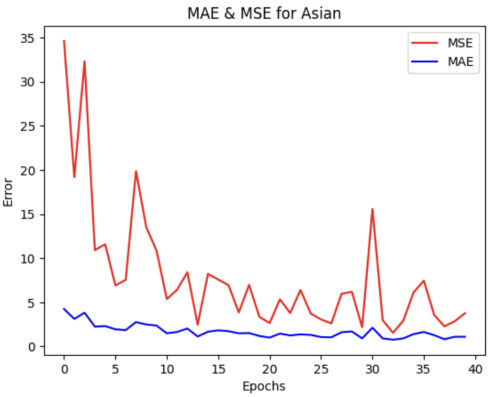


Figure 6.8: Mean Average Error and Mean Squared Error for Asian Category

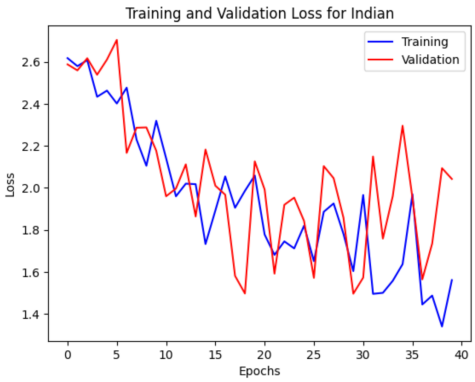


Figure 6.9: Training and Validation Accuracy for Indian Category

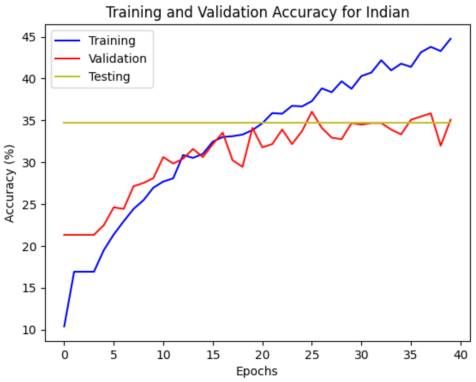


Figure 6.10: Training and Validation Loss for Indian Category

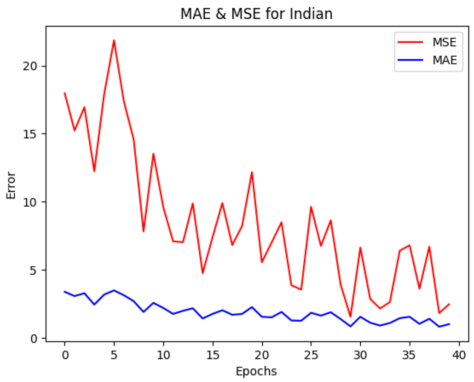


Figure 6.11: Mean Average Error and Mean Squared Error for Indian Category

Figure 6.12: Training and Validation Accuracy for Others Category

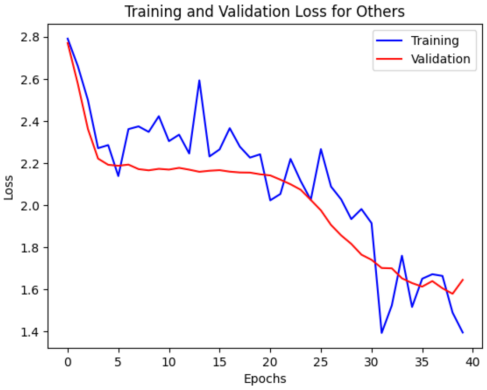


Figure 6.13: Training and Validation Loss for Others Category

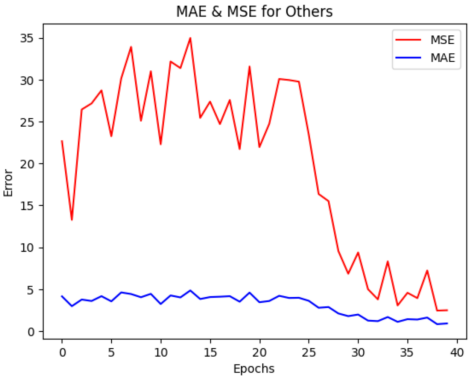
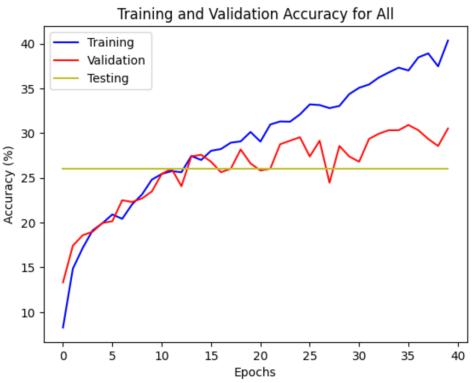
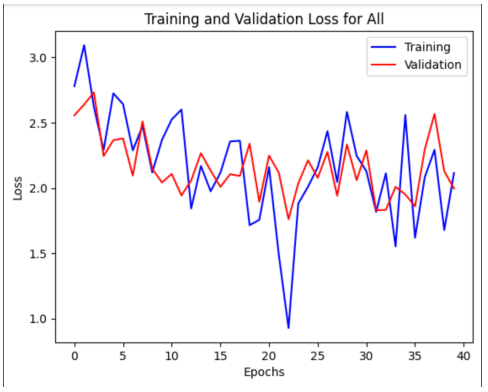


Figure 6.14: Mean Average Error and Mean Squared Error for Others Category





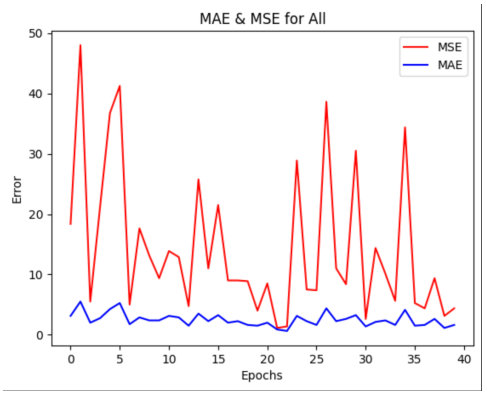
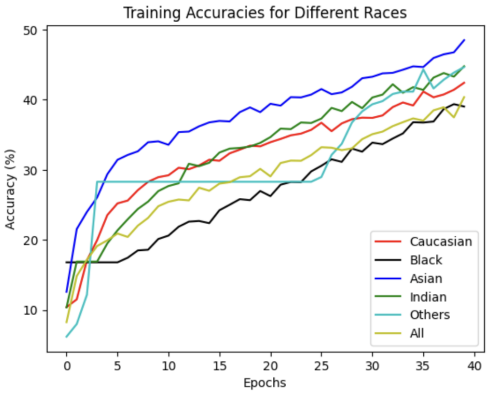
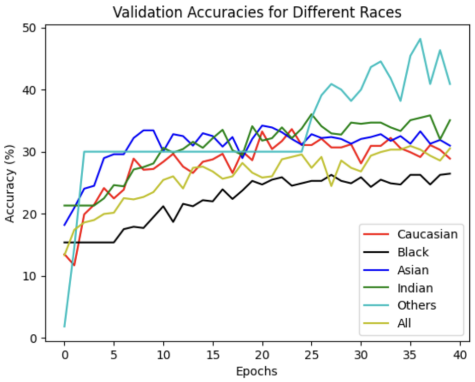


Figure 6.15: Training and Validation Accuracy for Combined Race Model

Figure 6.16: Training and Validation Loss for Combined Race Model

Figure 6.17: Mean Average Error and Mean Squared Error for Combined Race Model





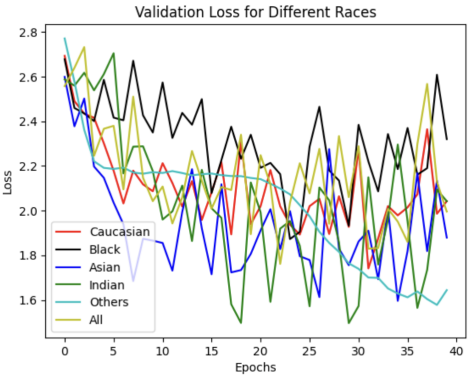
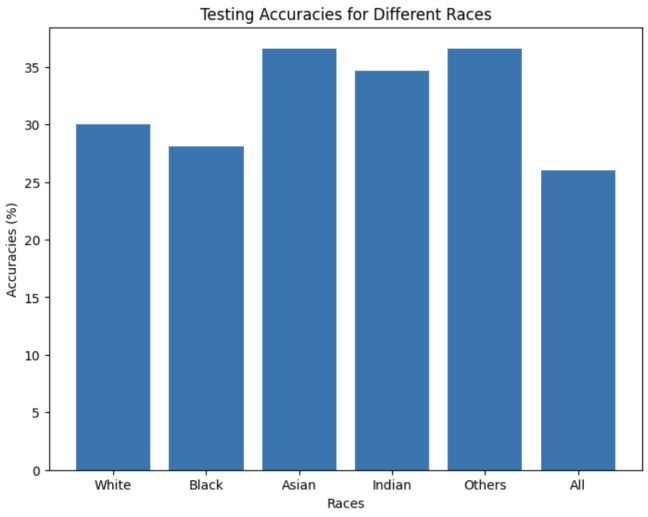


Figure 6.17: Training Accuracy for All Models

Figure 6.18: Validation Accuracy for All Models

Figure 6.19: Validation Loss for All Models



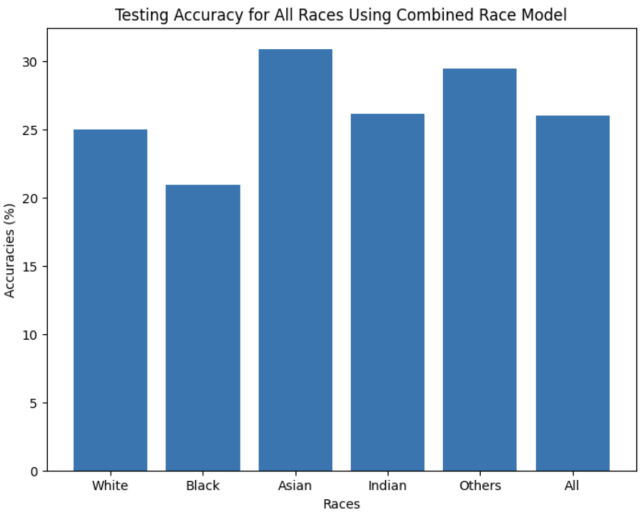


Figure 6.20: Testing Accuracies for Each Race Considering All Models (All representing accuracy of the Combined Race Model)

Figure 6.21: Testing Accuracies for Each Race When Applied to the Combined Race Model

7. Works Cited

[1] E. Ferguson and C. Wilkinson, “Juvenile age estimation from facial images,” Science & Justice, vol. 57, no. 1, pp. 58–62, 2017.

[2] T. Ganel, C. Sofer, and M. A. Goodale, “Biases in human perception of facial age are present and more exaggerated in current AI technology,” Scientific Reports, vol. 12, no. 1, 2022.

[3] G. Guo et al., “A Study on Automatic Age Estimation using a Large Database,” 2009 IEEE 12th International Conference on Computer Vision (ICCV), pp. 1986–1991, 2009.

[4] N.S., L., J., B., S., M. (2011). “Age Estimation Using Gender Information”. In: Venugopal, K.R., Patnaik, L.M. (eds) Computer Networks and Intelligent Computing. ICIP 2011. Communications in Computer and Information Science, vol 157. Springer, Berlin, Heidelberg. https://doi-org.myaccess.library.utoronto.ca/10.1007/978-3-642-22786-8_26

[5] Y. Deng, S. Teng, L. Fei, W. Zhang, and I. Rida, “A Multifeature Learning and Fusion Network for Facial Age Estimation,” Sensors, vol. 21, no. 13, p. 4597, Jul. 2021, doi: 10.3390/s21134597

[6] A. F. Alexis, P. Grimes, C. Boyd, J. Downie, A. Drinkwater, J. K. Garcia, and C. J. Gallagher, “Racial and ethnic differences in self-assessed facial aging in women,” Dermatologic Surgery, vol. 45, no. 12, pp. 1635–1648, 2019.

[7] Vashi, N. A., de Castro Maymone, M. B., & Kundu, R. V. “Aging Differences in Ethnic Skin”. The Journal of clinical and aesthetic dermatology, 9(1), pp 31–38, 2016.

[8] G. Antipov, “Apparent Age Estimation from Face Images Combining General and Children-Specialized Deep Learning Models,” CVF, pp. 96–104.

[9] R. Sharma, N. Pandey, Y. S. Thakur, A. Gangwar and S. Suman, "Age Estimation in Juveniles using Convolution Neural Network," 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 2021, pp. 1-4, doi: 10.1109/CONIT51480.2021.9498483.

[10] V. R. P. T. Khaled ELKarazle, "Facial Age Estimation Using Machine Learning Techniques: An Overview," Big Data and Cognitive Computing, vol. 6, no. 128, 2022.

[11] A. Najibi, “Racial Discrimination in Face Recognition Technology,” *Science in the News*, 26-Oct-2020. [Online]. Available: https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/. [Accessed: 11-Apr-2023].

[12] Wang X, Guo R, Kambhamettu C (2015) Deeply-learned feature for age estimation. In: 2015 IEEE winter conference on applications of computer vision. IEEE, pp 534–541

[13] Taheri S, Toygar O. (2019) On the use of dag-cnn architecture for age estimation with multi-stage features fusion. Neurocomputing.;329:300–310. doi: 10.1016/j.neucom.2018.10.071.

[14] Niu Z, Zhou M, Wang L, Gao X, Hua G (2016) Ordinal regression with multiple output cnn for age estimation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4920–4928

[15] Rothe R, Timofte R, Van Gool L (2015) Dex: Deep expectation of apparent age from a single image. In: Proceedings of the IEEE international conference on computer vision workshops, pp 10–15

[16] E. and S. D. Canada, “Government of Canada,” Ageism in Canada: Summary of the Discussion Guide - Canada.ca, 06-Feb-2023. [Online]. Available: https://www.canada.ca/en/employment-social-development/corporate/seniors/forum/consultation-ageism/summary.html. [Accessed: 11-Apr-2023].