Analyzing Lower Torso Curves for Computational Apparel Design and Modeling

Wenxuan Hong, Aristotle Kolefas, Diane Xu, Zongjie Yin, and Fanjun Zeng

I. Abstract

We employ techniques such as regression, tree-based models, clustering, and dimension reduction to address women's challenges with ready-to-wear clothing. The goal is to create a blueprint for fitting and grouping crotch curves across different races. Our analysis is based on a dataset containing 551 observations collected in 2020, including 404 White individuals, 69 African Americans, 57 Asian Americans, 6 Hispanics, and 15 individuals from other races. To achieve our objective, we fit the crotch curve for each individual using spline models with different numbers of knots and varying polynomial degrees and then determine the best fit by comparing the R^2 value and the model complexity. Using the model coefficients of the fitted curves and basic body measurements, we classify each individual based on race. Finally, with the aid of basic body measurements and race, we use ridge regression to predict the model coefficients and generate the crotch curve for all individuals.

II. Introduction

The current fashion system is flawed, especially in the fit of ready-to-wear clothing for women. Around half of women have trouble finding clothes that fit them well, particularly those with non-flat and curvy body types. The lower torso curve is integral to determining how pants fit on the body: the crotch curve in pants patterns is often drawn carelessly, leading to issues with fit. Our research aims to address this issue. Throughout our analysis, we focused on five main questions:

1. Do body shapes vary based on factors such as ethnicity, height, or BMI?
2. Is there any correlation between basic measurements (hip circumference, height, weight, etc.) and front and back curve coefficients?
3. Is there any distinct pattern that differentiates the races?
4. Is it possible to predict the race of the subject based on his/her body measurements and front and back curve coefficients?
5. Is it possible to predict curve shapes based on measurements such as maximum hip circumference and crotch depth?

In answering these questions, we aim to find ways to improve the fit of ready-to-wear clothing for women and help make the design process of pants patterns more accurate.

III. Literature/Background

The measurement of torso curves and other anatomical features is an essential aspect of creating clothing that fits the human body. Researchers have developed various techniques and methods to generate customized patterns for different body types.

One study that explores this is *Individualized Generation of Young Women's Crotch Curve Based on Body Images*. The researchers use image measurements and parameter-point-curve fitting rules to generate crotch curves of the human body, resulting in errors mainly between -5 and 5 mm. The study suggests modifying pattern crotch curves based on body crotch curves for improved clothing fit.

Another study, *An Image-Based Shape Analysis Approach and its Application to Young Women's Waist-Hip-Leg Position*, proposes an image-based method for classifying women's waist-hip-leg position based on body images, using factor analysis, K-means clustering analysis, and specific classification rules. The study finds good consistency between image-based and 3D measurements, with automatic identification accuracy reaching 93.3% and 90% for waist-hip and leg, respectively. Further research could be conducted on pattern differences of pants for various body types to improve clothing fit.

The paper *Categorization of Lower Body Shapes for Adult Females Based on Multiple-View Analysis* proposes a method for identifying lower body shapes using principal component analysis and K-means clustering analysis. The study develops two discriminant functions to provide a simple and intuitive application method for defining a new person's body shape group. This method offers a more reliable and objective approach to identifying lower body shapes compared to traditional methods. Additional research could involve expanding this method to other demographic groups or body parts and exploring its potential applications in clothing design and fit.

Pants Alteration by Graphic Somatometry Techniques focuses on developing a pant alteration technique that results in a superior fit for pants. The study classifies five body variations: round figure, pear-shaped-hip, weight-in-front, weight-in-back, and average figure variation. Future studies could focus on expanding the study's scope to include other types of clothing, such as jackets or blouses.

Parametric 3D Modeling of Young Women’s Lower Bodies Based on Shape Classification presents a method for creating 3D models of the lower body based on body shape classification using 3D human body scanning. The study uses a clustering method to classify the 24 body feature measurements of 333 young women into three categories: transverse, height, and crotch factors. The study shows that the proposed method can accurately create 3D models of young women's lower bodies. Subsequent studies could focus on expanding the study's scope to include men’s upper bodies.

The study *Affect the Comfort of Pants Crotch Curve Analysis of the Factors* explores the factors that affect the comfort of pants, with a specific focus on the shape of the crotch curve. The study finds that the shape of the crotch curve is determined by human pygal morphology, and the crotch width should meet the body thickness and movement function. The design of the crotch bend structure plays a crucial role in determining the fitness of trousers. Elastic fabric and hip-huggers were found to improve wearing comfort, while smooth fabric used as pants lining effectively reduces friction between the pants and the human body. The study further analyzes the structure of pants, the scalability of trousers fabric, style shape, technology, and other aspects to propose effective improvements in pants crotch bend comfort measures. The paper also explores the importance of the design of the back jut-out and the two main factors affecting the values of body crotch movement: the physical structure of the human waist and hips and the motion of the human waist and hips. Finally, the paper divides trousers into four categories: body-fitted trousers, less-fitted trousers, loose trousers, and slacks, and analyzes the back seat of each category.

Individual Pattern Making of Skirts and Pants Using Computerized Draping discusses the development of a computerized draping method for creating individualized patterns for skirts and pants. The authors modified a traditional draping method to be used in conjunction with modern 3D modeling techniques, which accounts for the characteristics that allow fabric to conform to a 3D shape. The method was tested on making patterns for tight skirts and pants, and the entire process was done on a computer. The authors believe that this method has the potential to be more efficient and simpler for creating custom-made clothes. However, while the method works well for fitting dummies, there are difficulties and problems in fitting human body shapes, particularly in setting ease allowance for patterns on 3D human body shapes. The article provides a detailed explanation of the theory behind the computerized draping method, the measurement of 3D data, and the reconstruction method of 3D body shapes for tight skirts and pants.

The paper *Converting Lower-body Features from Three-dimensional Body images into Rules for Individualized Pant Patterns* introduces a method to fit a human's lower body and use statistical and visualization tools to predict a curve and produce a pant pattern by using measurements of young females' lower-body features. Firstly, the paper introduces two methods to generate individualized patterns from body data. The first method involves scanning the lower part of the human body using 3D scanning technology and then flattening the scanned 3D data using a reasonable algorithm used in the industry. The key benefit of this method is that it creates pattern pieces directly from a 3D virtual model and transfers the form aspects of a customer's figure into the pattern. However, incorporating this virtual design technique into the conventional pattern-making system is difficult. This technique appears to lack the practicality and precision that the industry requires. The second method is based on an individual's experience. The data of AI, ANNs, and FL are modified and applied through human experience. This method avoids the impracticality of the first method, but the requirement for experience and the accuracy of the experience itself still make this method unsuitable for wide application. This thesis introduces another concept to solve this problem. The fitting problem of the pant pattern is transformed into a computational problem of six parameters that determine the pattern and shape of pants: waist, abdomen, hip, thigh, knee, and ankle measurements. By using height, girth, and the crotch of the human body measured by 3D scanning technology, this paper uses linear regression and logistic regression to find the relationship between these three variables and the aforementioned key features that affect the pants' curve and pattern. The regression model restricted the error tolerance of the curve to within 1-2 cm, making the production of pants more precise.

These studies indicate that the measurement of torso curves and other anatomical features is critical to creating clothing that fits the human body. The techniques and methods developed by researchers provide insights into how to generate customized patterns for different body types and improve clothing fit and comfort. More work can be done to expand the scope of these techniques and methods to include other types of clothing, such as jackets or blouses, or to other demographic groups or body parts.

IV. Data

Description of Datasets

The data we employed was collected in 2020 and falls into two categories: measurements involving each individual's waist, and coordinates of the lower torso curves of each individual. We analyzed 551 individuals, of whom 404 were White, 69 were African American, 57 were Asian American, 6 were Hispanic, and 15 were of another race.

The datasets containing waist measurements included nine different quantitative variables, such as subject number, body mass index (BMI), height, maximum hip circumference, anterior-posterior length, depth, crotch curve length at back waist, front crotch length, and back crotch length. All the quantitative variables are in millimeters except BMI, which is measured in kg/m^2. The anterior-posterior length is the distance between the front and back curves at waist level, and depth is the distance between the waist level to the deepest point of the torso curve. The datasets also included plots of each individual's torso curve along with comments about aberrations in the shape of the curve.

Each file of coordinates contained a column of x and a column of y-values.

Data Cleaning

To clean the data, we merged all the waist measurement datasets into one Excel file, which we imported into R. Then we removed all individuals with at least one missing value for any of the recorded variables and all individuals with curves described as abnormal; 544 individuals remained, including 400 Whites, 66 African Americans, 57 Asian Americans, 6 Hispanics, and 15 individuals of another race. Since the plots of the torso curves and associated comments could not be rendered in R, we analyzed them separately using the provided coordinates.

Summary Statistics

	Mean	**Median**	**Std. Dev**	**Minimum**	**Maximum**
BMI	23.18	22.37	3.89	15.25	41.45
Weight (kg)	63.91	62.02	11.36	39.23	125.62
Height (mm)	1660.41	1656.50	71.59	1248.00	1879.00
Max Hip (mm)	1019.94	1006.00	80.86	800.00	1303.00

Ant.-Poster. Length (mm)	271.30	265.69	33.44	197.84	412.10
Depth (mm)	264.63	264.66	24.09	198.42	329.57

Table 1: Summary Statistics for Whites (n = 400)

	Mean	**Median**	**Std. Dev**	**Minimum**	**Maximum**
BMI	25.86	24.80	4.88	18.10	41.00
Weight (kg)	69.87	66.74	14.15	44.31	103.00
Height (mm)	1642.83	1644.00	79.74	1382.00	1849.00
Max Hip (mm)	1060.02	1046.50	103.50	856.00	1337.00
Ant.-Poster. Length (mm)	294.55	287.20	40.72	205.63	402.84
Depth (mm)	259.56	258.65	23.39	217.42	313.67

Table 2: Summary Statistics for African Americans (n = 66)

	Mean	**Median**	**Std. Dev**	**Minimum**	**Maximum**
BMI	22.44	21.89	2.67	17.43	31.75
Weight (kg)	56.14	55.78	7.03	39.91	80.27
Height (mm)	1581.84	1584.00	48.15	1488.00	1704.00
Max Hip (mm)	955.39	954.00	55.74	817.00	1114.00
Ant.-Poster. Length (mm)	254.70	253.73	21.21	208.58	331.73
Depth (mm)	250.85	251.55	23.38	194.40	304.08

Table 3: Summary Statistics for Asians (n = 57)

	Mean	**Median**	**Std. Dev**	**Minimum**	**Maximum**
BMI	22.29	21.07	3.73	19.19	29.02

Weight (kg)	58.43	57.48	11.16	46.71	77.78
Height (mm)	1617.50	1611.50	71.18	1541.00	1736.00
Max Hip (mm)	971.67	949.50	85.09	885.00	1085.00
Ant.-Poster. Length (mm)	263.81	254.06	37.94	231.03	338.38
Depth (mm)	252.97	250.38	21.76	232.54	293.18

Table 4: Summary Statistics for Hispanics (n = 6)

	Mean	**Median**	**Std. Dev**	**Minimum**	**Maximum**
BMI	24.78	24.44	3.35	19.66	31.03
Weight (kg)	65.59	66.89	11.07	48.07	86.17
Height (mm)	1624.60	1625.00	78.40	1453.00	1831.00
Max Hip (mm)	1030.33	1038.00	80.17	916.00	1142.00
Ant.-Poster. Length (mm)	283.03	286.95	34.33	233.44	336.15
Depth (mm)	267.42	268.70	28.14	219.03	327.32

Table 5: Summary Statistics for Others (n = 15)

The data suggests that African Americans typically have the highest mean BMI of 25.86 kg/m^2, maximum hip circumference of 1060.02mm, and anterior-posterior length of 294.55mm. White individuals tend to be the tallest, while Asians are the shortest, with means of 1660.41 and 1617.50 mm, respectively (see Tables 1 and 3). Interestingly, individuals of other races have the largest depth at 267.42 mm, while Asians have the smallest at 250.85 mm. BMI and weight are most variable in African Americans, with standard deviations of 4.88 kg/m^2 and 14.15 kg, respectively (see Table 2). On the other hand, Asians generally show the least variability in both BMI and weight. Height is most variable among African Americans and Others, while it is most consistent among Asians. Additionally, the maximum hip circumference has the largest standard deviation among African Americans, while it has the smallest standard deviation in the Asian cohort. Finally, anterior-posterior length is usually most variable among African Americans and least variable among Asians, while Whites and those of the "Other" race show the most variation in Depth.

Correlations

We conducted pairwise correlation analysis on the numerical explanatory variables and found that BMI, weight, maximum hip circumference, and anterior-posterior length are highly positively correlated, with pairwise coefficients greater than 0.85. These variables are also positively correlated (around 0.7) with crotch curve length. Depth and crotch curve length are significantly correlated (ρ = 0.88), and the correlation coefficient between depth and front crotch length and between depth and back crotch length are 0.74 and 0.84, respectively.

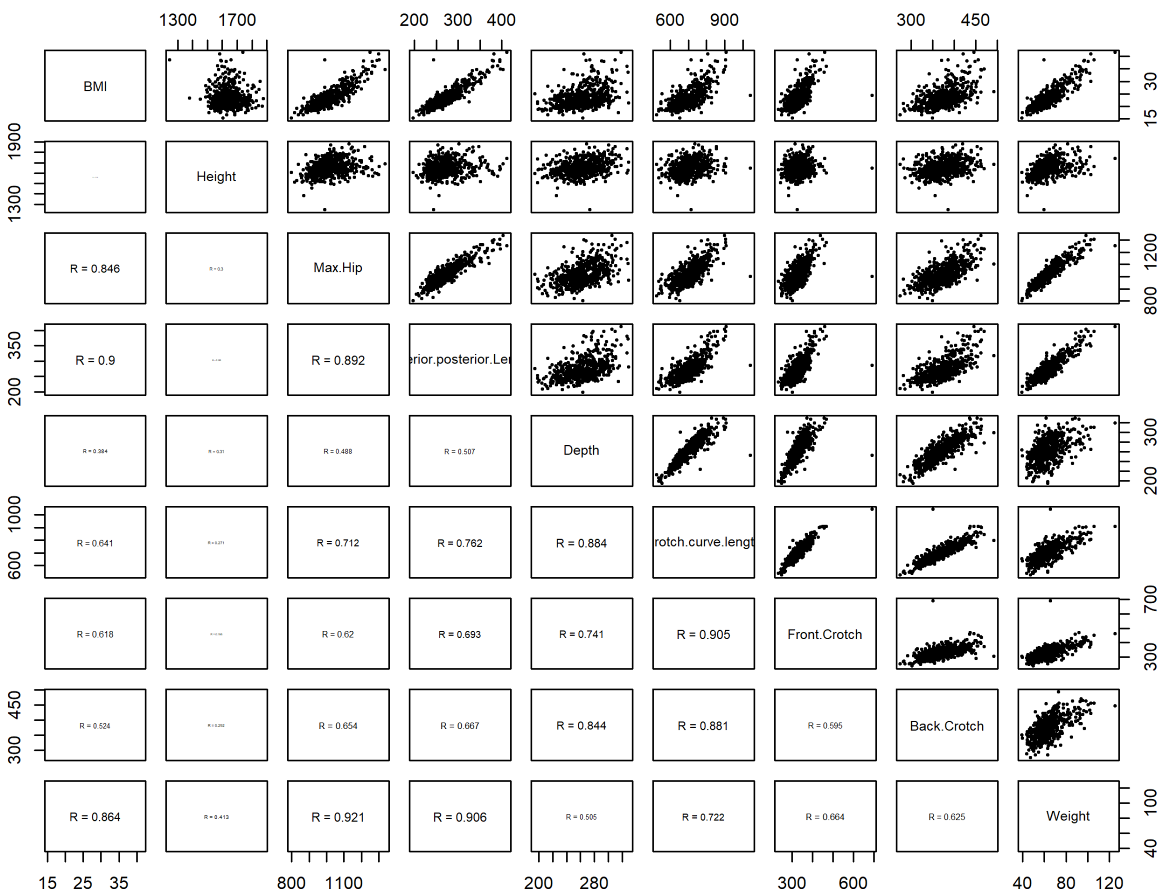

Figure 1: Correlation Scatterplot Matrix

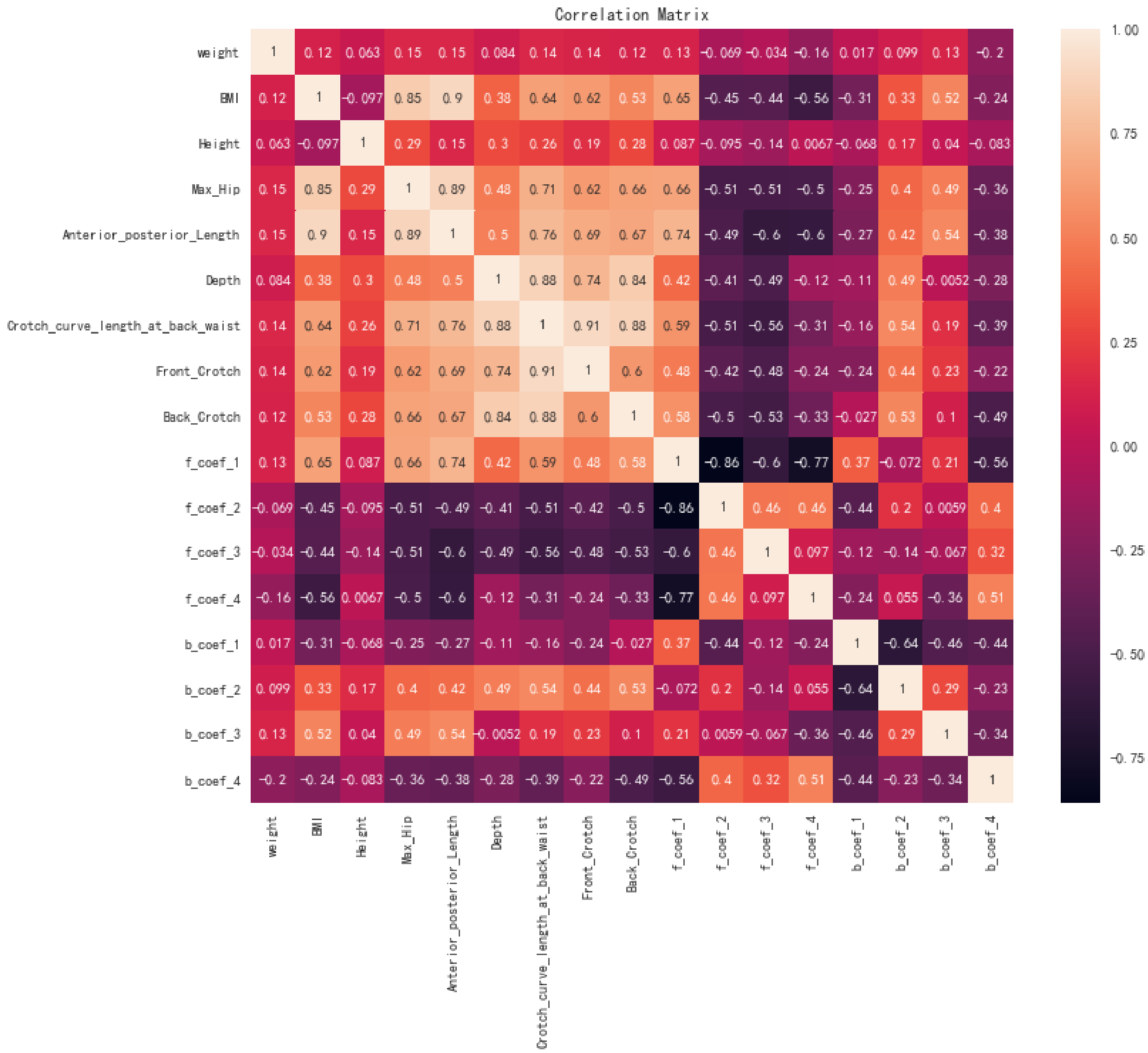

Figure 2: Correlation Heatmap

V. Analysis/Methodology

5.1 Linear Regression

We started our analysis by performing several linear regression models on the data. In each model, we used factored race, height, weight, and maximum hip circumference as predictors, and one of crotch curve length at the back waist, depth, and anterior-posterior length as the response variable. Our findings indicate that weight and maximum hip circumference are the most effective predictors of crotch curve length at the back waist (with p-values of 0.000 for the Wald test in both cases). The White race indicator variable, weight, and height are the most useful predictors of depth (with p-values of 0.000, 0.007, and 0.007, respectively). Weight, height, and maximum hip

circumference are the most useful predictors of anterior-posterior length. The model with anterior-posterior length as the response variable has the best fit, with an R^2 value of 0.888 and an adjusted R^2 value of 0.887, while the other models have R^2 values below 0.550. The residuals of all models are normally distributed, which suggests that a linear model is appropriate.

5.2 Curve Fitting

5.2.1 Purpose and Strategy

In this section, we aim to investigate whether body shapes (crotch curves) vary based on factors such as ethnicity, height, or BMI. However, analyzing each crotch curve, which is represented by approximately 100 coordinates, is a challenging task. Therefore, our strategy is to use a model to fit the crotch curves (coordinates) and extract the model coefficients to characterize and represent the curves. Each crotch curve has much fewer coefficients than coordinates, making it easier to analyze the variation of the model coefficients of the crotch curves, which reveals the variation in the body shapes.

5.2.2 Dividing Curves

Before we began fitting curves, a notable issue was that the crotch curves did not have a unique y-coordinate for every x-coordinate. This meant that they were not functions and could not be properly analyzed in their original form. Therefore, we divided each curve into two pieces, front and back curves, from its deepest point, and then rotated the front and back curves by 90 degrees to turn them into functions.

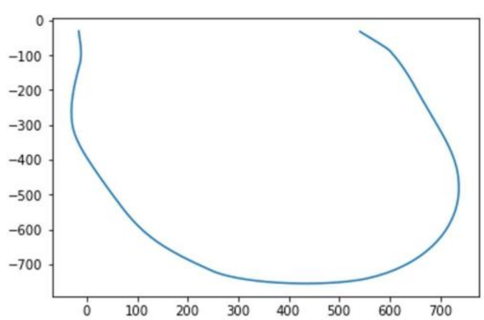

Figure 3. Original Curve

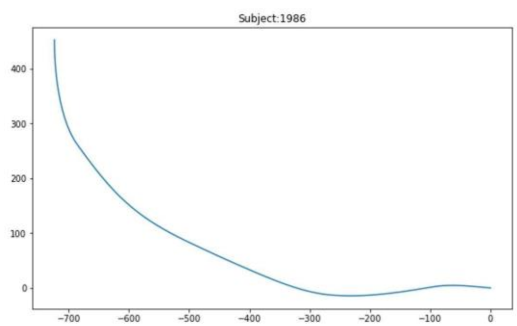

Figure 4. Front curve after rotation

5.2.3 Model Selection

Since our approach uses model coefficients to represent crotch curves, the hypothesis class of the model should not only be able to fit all curves well but also generate a fixed and small number of coefficients for each curve. For this reason, the spline model is a suitable choice for our purposes.

To ensure consistency, we shifted the coordinates of the front and back curves during curve fitting, so that the topmost points for all curves are (0, 0). In order to reduce the complexity of the model (number of coefficients) while maintaining performance, we compared the fitted curves of

spline models with different hyperparameters, including polynomial degrees of 2 and 3, and between 3 and 4 knots, by plotting the original curves and fitted curves.

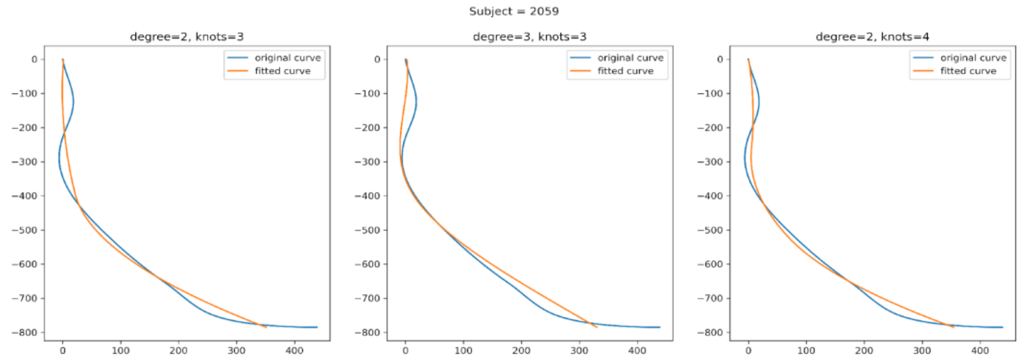

Figure 5. Fitted Curve vs. Original Curve with Different Hyperparameters

After comparing the plots, we found that the spline model with a polynomial degree of 2 and 3 knots has the lowest model complexity (generating only 4 coefficients for each curve), while also fitting the original curve well. Once we determined the hyperparameters of the spline model, we fit all curves and generated 4 coefficients for each curve. To measure the quality of our fitting, we retrieved the coordinates that form the fitted version of each curve based on its coefficients. We then calculated the R^2 values for the coordinates of the fitted curves and original curves, which allowed us to quantify the accuracy of our fitting:

Race	**White**	**African American**	**Asian**	**Hispanic**	**Other**
Median R^2 Value	0.992	0.991	0.992	0.991	0.992

Table 6. Median R^2 Values of Fitted and Original Curves

From the plots and the table above, we can conclude that this spline model can fit the crotch curves well for all race groups.

5.2.4 Analysis of Variations

After we computed the model coefficients for each subject, we can analyze the variations of the body shapes by plotting the relationship between coefficients and features (BMI, Height, Race) using boxplots and scatterplots:

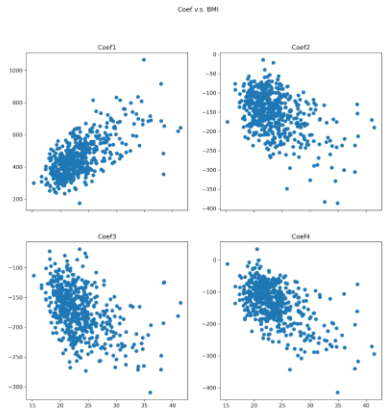

Figure 6. Coefficients vs. BMI

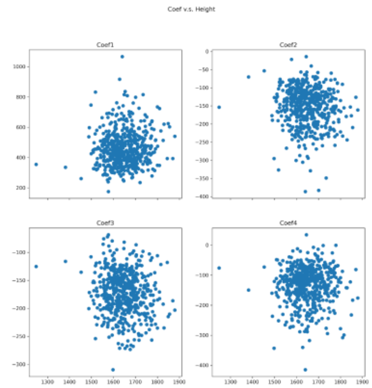

Figure 7. Coefficients vs. Height

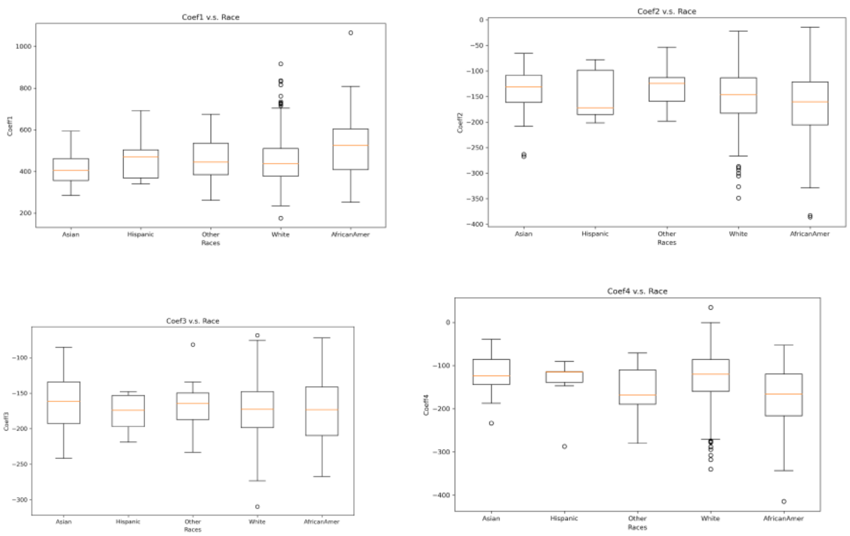

Figure 8. Coefficients vs. Race

From the scatter and box plots, we can conclude that the body shapes vary based on the BMI and races but not height.

5.3 Curve Prediction

In this section, our goal is to predict curve shapes based on measurements such as maximum hip circumference and crotch depth. Since we already generated 4 model coefficients for each curve in the last section, from which we retrieved the coordinates and plotted the curve, our strategy is to use supervised learning to predict the 4 spline model coefficients based on features for each curve. We will then plot the predicted crotch curves based on the predicted coefficients.

5.3.1 Predict Coefficients

We trained a ridge regression model to predict the coefficients, as a linear model can successfully capture the pattern without the risk of overfitting. Specifically, we used BMI, height, max hip, anterior posterior length, depth, and crotch curve length at the back waist as features, and created labels from the 4 spline model coefficients. We then trained separate ridge regression models for the back and front curves, as well as for different races. Once the training was complete, we predicted the coefficients for each curve and compared them with the true coefficients, extracted directly from the spline models, by calculating the R^2 values. However, because the sample sizes are small for all race groups except white, the R^2 values are highly dependent on the way we split the training and testing sets, rendering them meaningless. Therefore, we only present R^2 values for the White race group.

Coefficient	**Training Set for Back Curves**	**Testing Set for Back Curves**	**Training Set for Front Curves**	**Testing Set for Front Curves**
coef1	0.24	0.05	0.58	0.38
coef2	0.41	0.17	0.4	0.22
coef3	0.52	0.31	0.63	0.33
coef4	0.43	0.34	0.41	0.33

Table 7. R^2 Values of Predicted Coefficients for White

5.3.2 Predict and Fit Curves

After predicting the coefficients, we were able to plot the front and back curves separately. The next step was to concatenate these two pieces of crotch curves. To achieve this goal, we rescaled the coordinates of the curves to match the measurements in our dataset and concatenated the front and back curves using the anterior and posterior length instead of simply using the deepest point. This is because the bottom part of the curve is often flat, which can lead to a high loss. The anterior and posterior length equals the distance between the topmost points of the front and back curves. Finally, if the deepest points did not match, we connected them with a line. These steps allowed us to effectively use the features to predict coefficients and create accurate front and back curves for each subject.

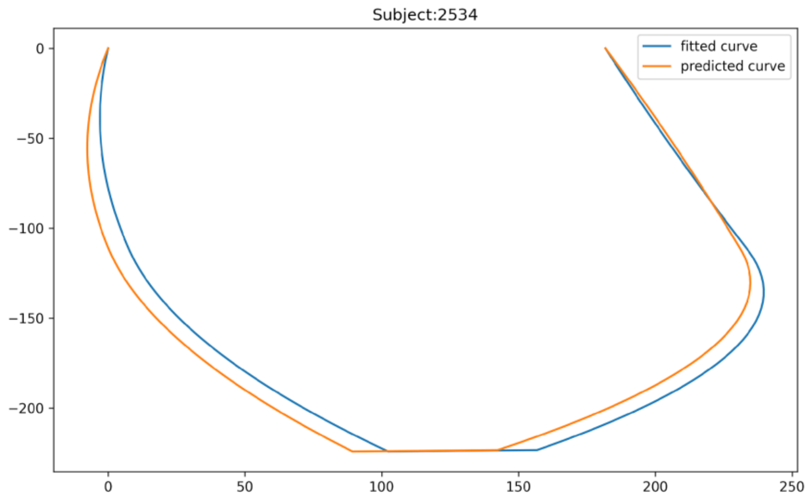

Figure 9: Predicted Curve and Fitted Curve

To measure the goodness of our predictions, we calculated the median R^2 values between the fitted curves and the predicted curves for different race groups.

Race	**White**	**African American**	**Asian**	**Hispanic**	**Other**
Median R^2 Values	0.993	0.981	0.966	0.293	0.8
Sample Size	396	66	55	6	14

Table 8: Median R^2 Value of Fitted and Predicted Curve

The above table indicates that when the size of a racial group is large (i.e., Whites and African Americans), estimating the spline model coefficients yields predicted curves very close to the fitted curves.

5.3.3 Predict Curves and Original Curves

Although the predicted curves are similar to the fitted curves, our goal is to predict the original curves, not the fitted ones. Therefore, it is more reasonable to directly compare the predicted curves with the original curves. We computed the median R^2 values between this pair of curves to assess the accuracy of our predictions.

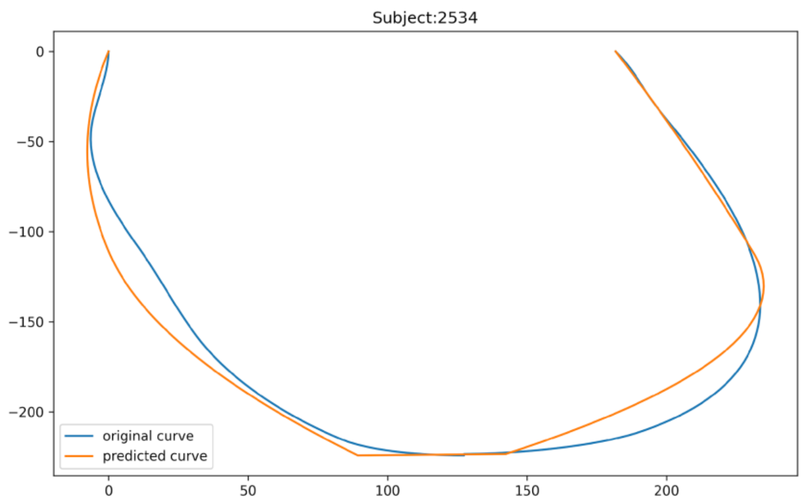

Figure 10. Predicted and original curves

Race	White	African American	Asian	Hispanic	Other
Median R^2 Value	0.985	0.974	0.958	0.292	0.795
Sample Size	396	66	55	6	14

Table 9: Median R^2 Value of Predicted Curve and Original Curve

According to this table, we can observe that although the average R^2 values of the predicted curves compared to the fitted curves are slightly lower than those compared to the original curves, for race groups with large sample sizes, the R^2 values are high enough to indicate that our predictions of the crotch curves are accurate.

5.4 PCA and K-means

In this section, we will primarily focus on the White, African American, and Asian groups, as they constitute most individuals in the dataset. Firstly, we utilized the results from polynomial curve fitting with 3 knots and a degree 2 for both the front and back curves. We then merged it with the original dataset using the subject ID. Following this, we standardized all columns to avoid possible discrepancies from different measurements in different columns while computing the distance measurement in K-means. To explore any possible differences between the different race groups, we utilized principal component analysis (PCA) as a dimension reduction technique, with n = 2, to visualize the data points on the x- and y-axes and attempted to identify any underlying patterns distinguishing different races. After standardization, we employed the Elbow method to determine the optimal number of K-means algorithm clusters. We obtained k = 3 with a distortion score of 1929.158 for White vs. African American (See Figure 11) and k = 3 with a distortion score of 1920.364 for White vs. Asian (See Figure 12).

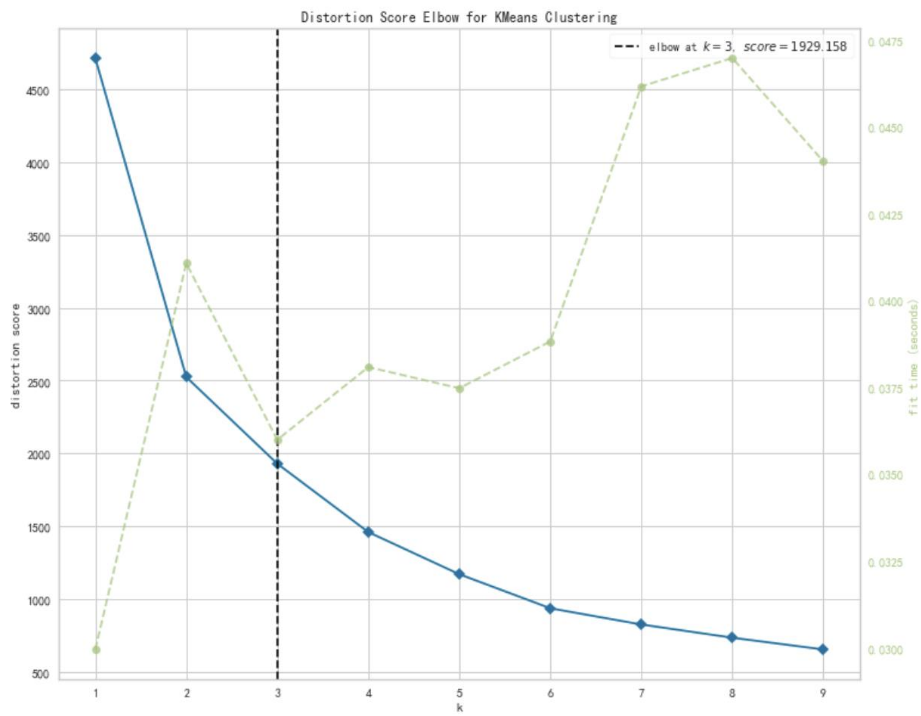

Figure 11: Distortion Score Elbow for K-Means Clustering of African Americans vs. Whites

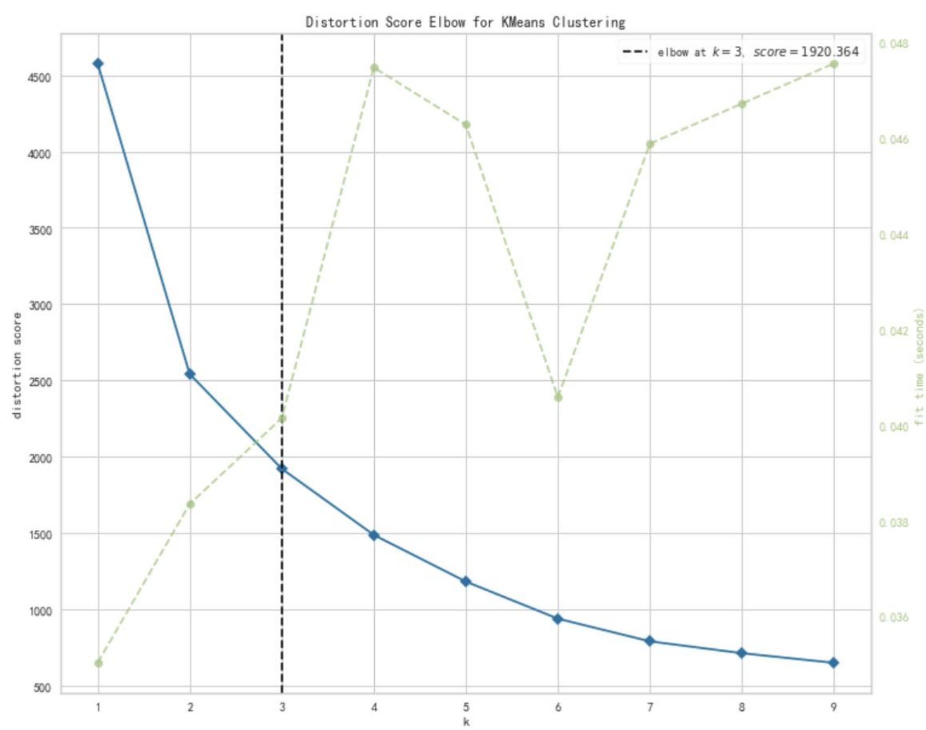

Figure 12: Distortion Score Elbow for K-Means Clustering of Asians vs. Whites

Finally, we generated scatterplots for both the White vs. African American group and the White vs. Asian group, plotting the different K-means resulting labels for each race separately (see Figures 13 and 14).

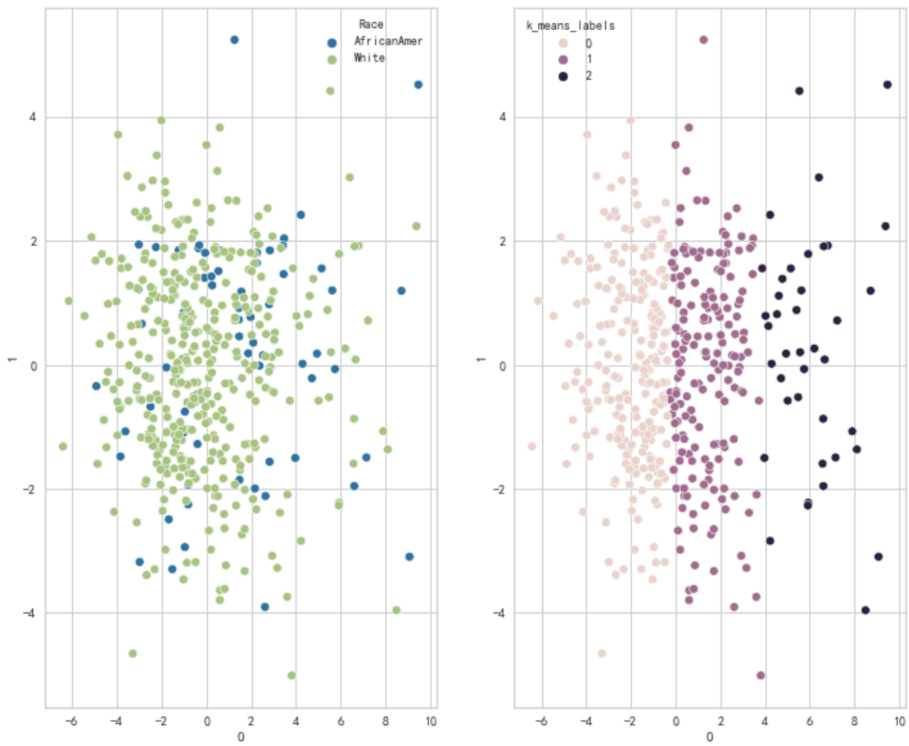

Figure 13: K-Mean clustering for African Americans vs. Whites

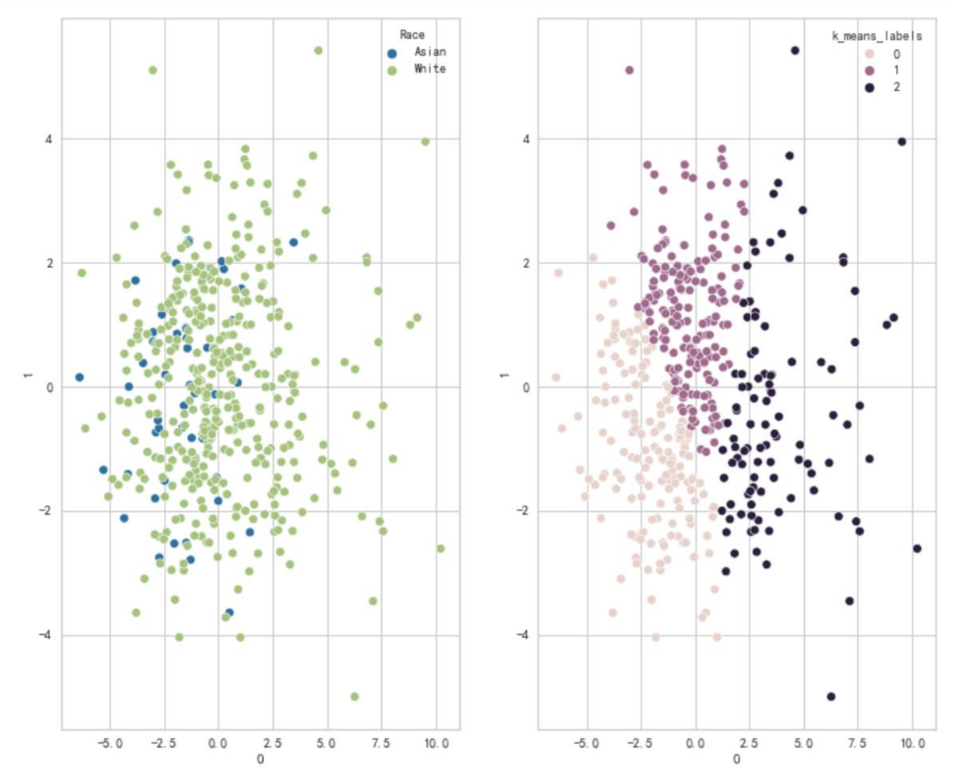

Figure 14: K-Mean clustering for Asians vs. Whites

Regrettably, we did not find any distinct pattern that differentiated the races. Subsequently, we created cluster maps for all numerical covariates to find relationships between different races. For White vs. African American groups (see Figure 15), we discovered that four of the front and back curve coefficients shared traits, as did the remaining eight features. Our findings were similar for White vs. Asian groups (see Figure 16). Overall, the clustering approach did not provide evidence that the features are strong predictors of race.

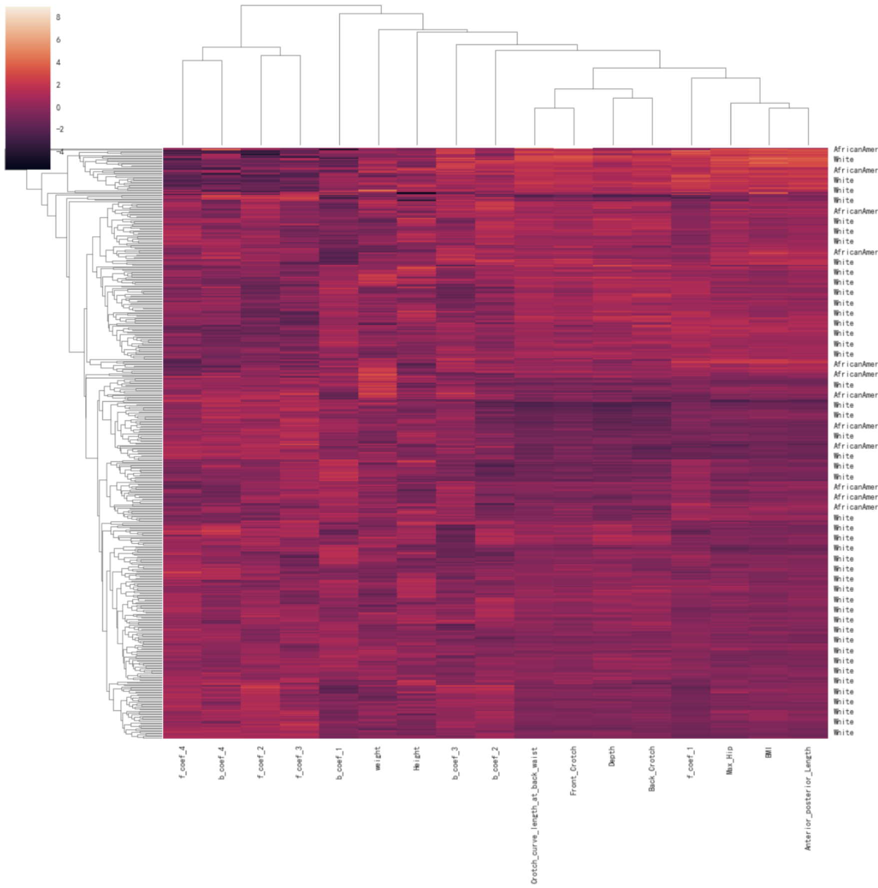

Figure 15: Heatmap for African American VS White

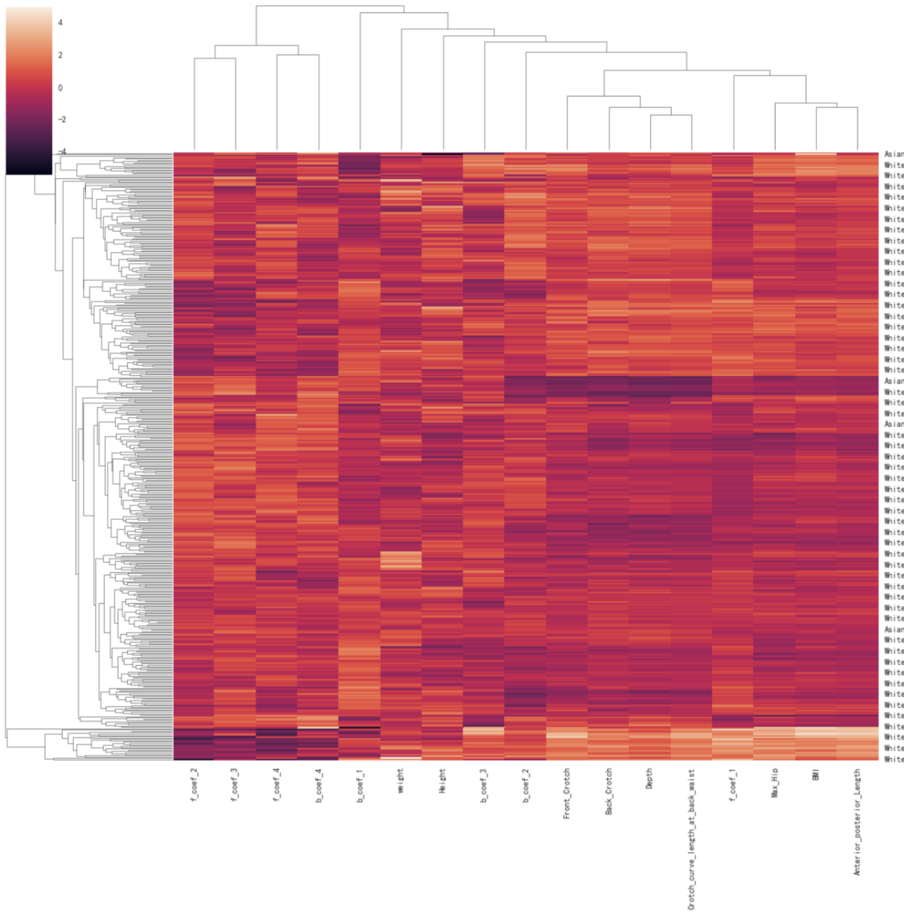

Figure 16: Heatmap for Asian vs. White

Definition of each Variable

Variable Name	Definition
f_coef_1	Front Curve Coefficient 1
f_coef_2	Front Curve Coefficient 2
f_coef_3	Front Curve Coefficient 3
f_coef_4	Front Curve Coefficient 4
b_coef_1	Back Curve Coefficient 1
b_coef_2	Back Curve Coefficient 2
b_coef_3	Back Curve Coefficient 3
b_coef_4	Back Curve Coefficient 4

Table 10: Definition of the Variables

5.5 Logistic Regression

We subsequently employed machine learning techniques to predict the race of each individual based on the covariates in our dataset. Our initial choice for a classification algorithm was logistic regression, but the data was imbalanced, with 397 White individuals, 66 African

Americans, and 56 Asians. To address this issue, we utilized the SMOTE method. Once we had resolved the data imbalance problem, we applied logistic regression to both groups and obtained the following results (graphs are shown in Figures 17 to 20 and Figure 27 to 30):

	White VS African American		White VS Asian	
	Balanced train set			
	Balanced test set	Imbalanced test set	Balanced test set	Imbalanced test set
Accuracy	0.78	0.73	0.68	0.62
Precision	0.77	0.93	0.63	0.88
Recall	0.76	0.73	0.80	0.65
AUC	0.83	0.80	0.81	0.52

Table 11: Results of Logistic Regression

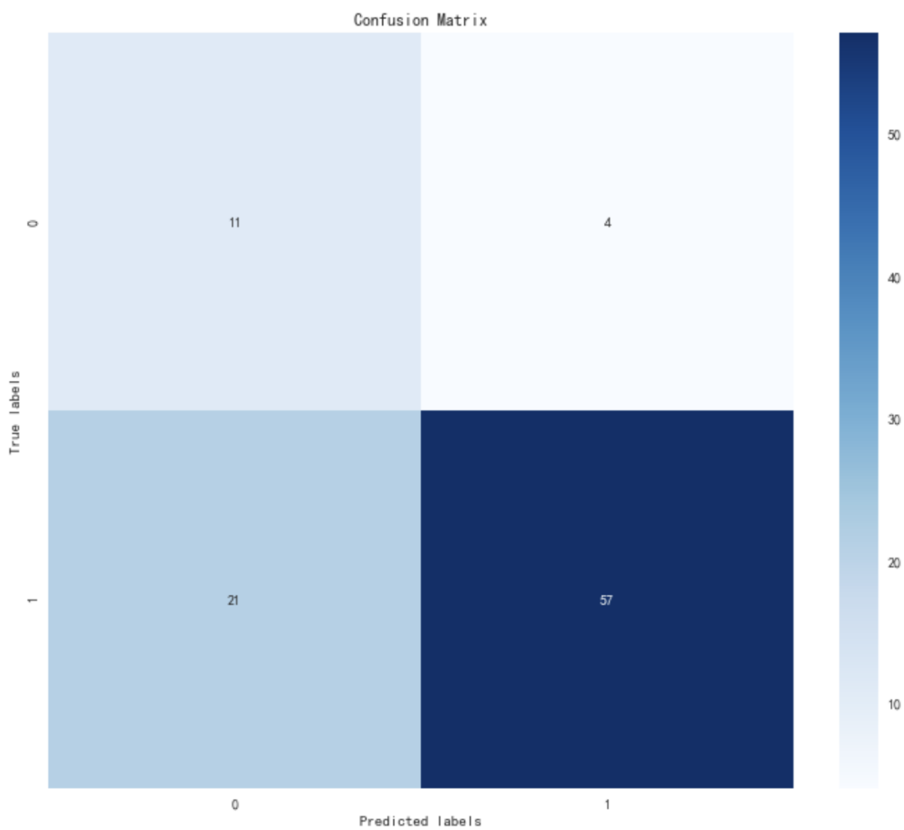

Figure 17: Confusion matrix for logistic regression of African Americans vs. Whites (imbalanced data)

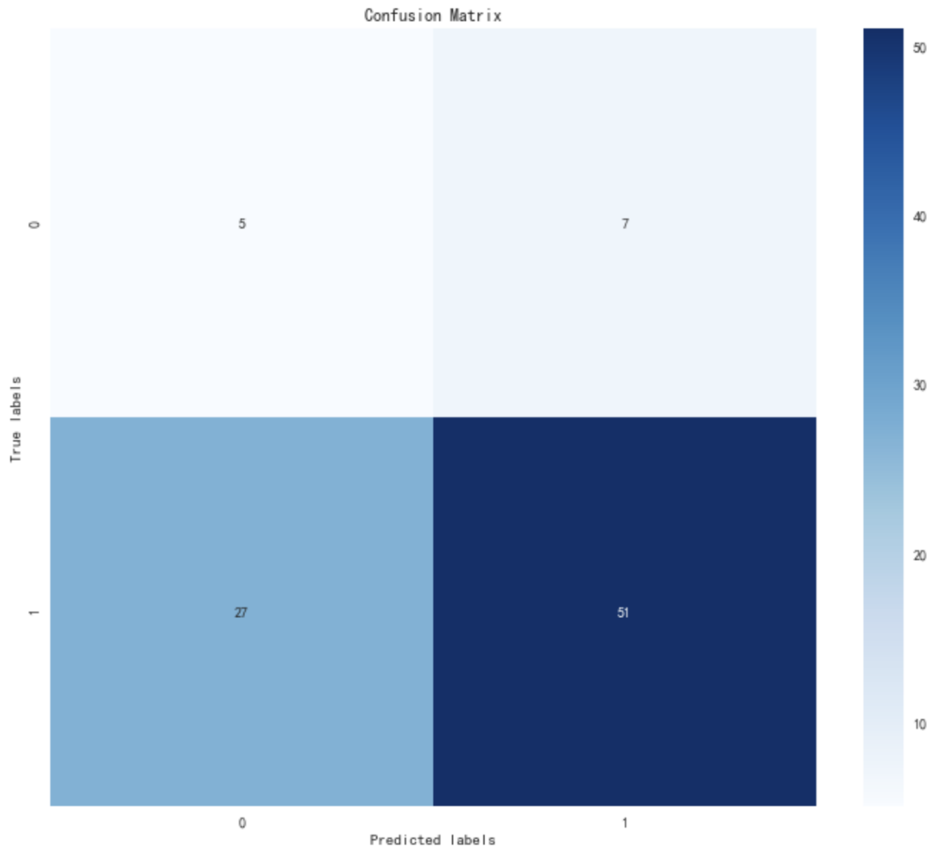

Figure 18: Confusion matrix for logistic regression of Asians VS. Whites (imbalanced data)

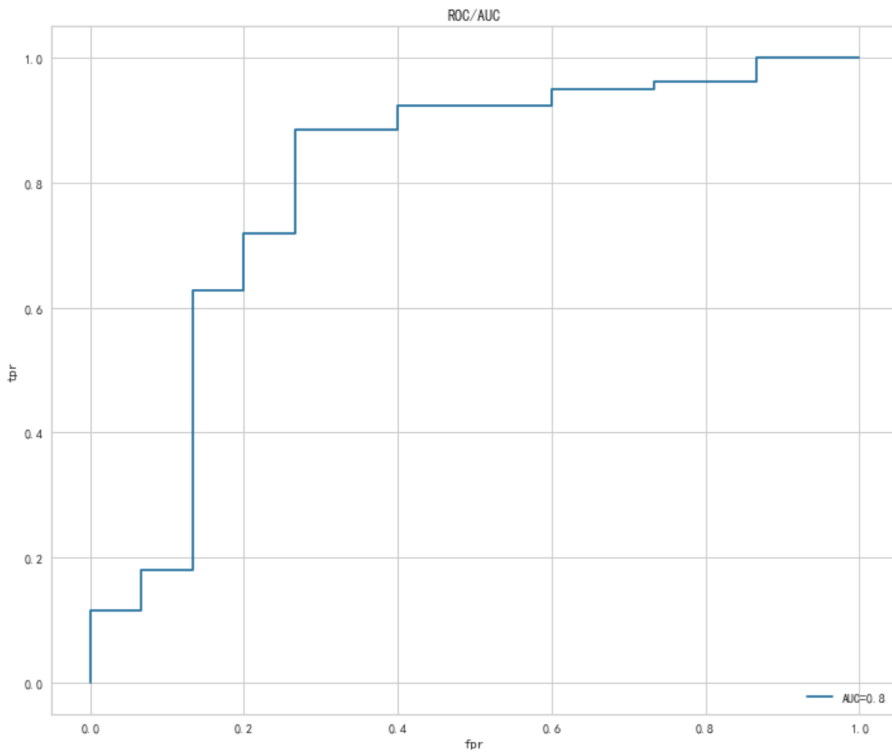

Figure 19: ROC/AUC curve for logistic regression of African Americans vs. Whites (imbalanced data)

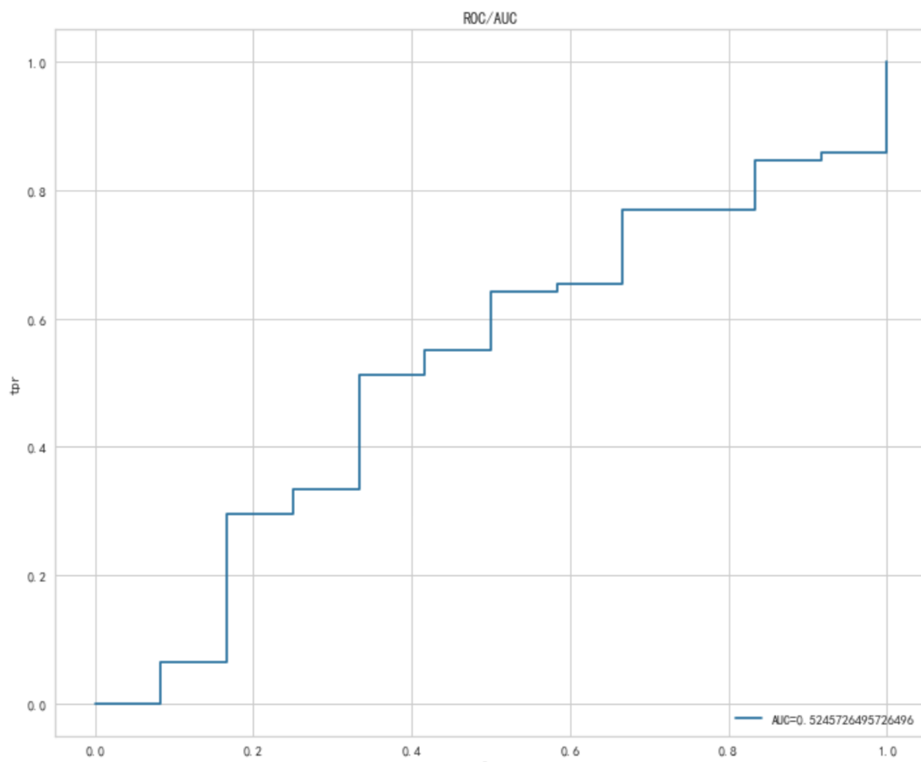

Figure 20: ROC/AUC curve for logistic regression of Asians vs. Whites (imbalanced data)*

We created Precision-Recall Curves for both groups that were useful for analyzing the unbalanced data (see Figures 21-22 and 31-32).

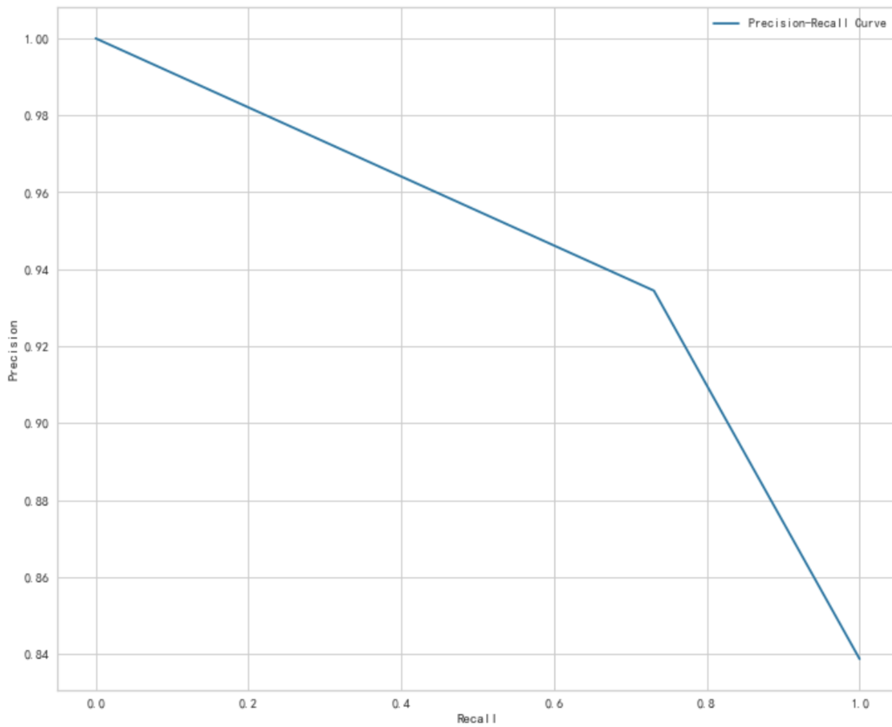

Figure 21: Precision Recall graph for logistic regression of African Americans vs. Whites (imbalanced data)

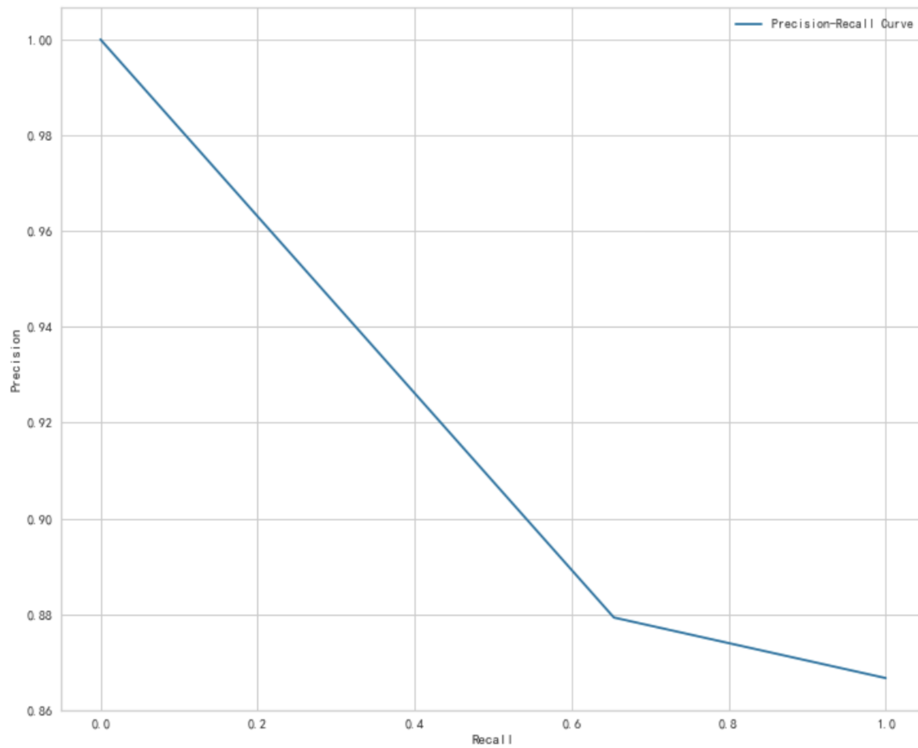

Figure 22: Precision Recall graph for logistic regression of Asians VS. Whites (imbalanced data)

Body measurements and both front and back curve coefficients have predictive power for race. However, data imbalance affects the accuracy of logistic regression.

5.6 Random Forest

We also made use of the Random Forest ensemble learning method, which is a classical tree-based algorithm. Random Forest utilizes decision trees and combines the output of multiple decision trees to produce a single result. We addressed the issue of data imbalance by employing the Synthetic Minority Oversampling Technique (SMOTE). Next, we performed grid search and cross-validation to identify the optimal combination of hyperparameters and enhance the stability and robustness of the model. We considered the following hyperparameter set:

n_estimators	max_features	max_depth	min_samples_split	min_samples_leaf
[100, 300, 500, 1000]	[‘auto’, ’sqrt’]	[10, 30, 50]	[2. 5, 8]	[1, 2, 3]

Table 12: Random Forest Hyperparameter Selection

Definition of each Hyperparameter:

Hyperparameter Name	Definition
n_estimators	Total number of trees

max_features	Maximum number of features selected at every split
max_depth	Maximum depth of each individual tree
min_samples_split	Minimum number of samples required to split an internal node
min_samples_leaf	Minimum number of samples in each leaf

Table 13: Definition of the hyperparameter

The best hyperparameter set yielded the following results (see Figures 23-24 and 33-36):

	White VS. African American		White VS. Asian	
	Balanced train set			
	Balanced test set	Imbalanced test set	Balanced test set	Imbalanced test set
n_estimators	100	100	1000	500
max_features	‘sqrt’	‘sqrt’	‘auto’	‘sqrt’
max_depth	50	30	50	50
min_samples_split	2	2	2	2
min_samples_leaf	1	2	1	1
Accuracy	0.88	0.81	0.92	0.86

Table 14: Results of Random Forest

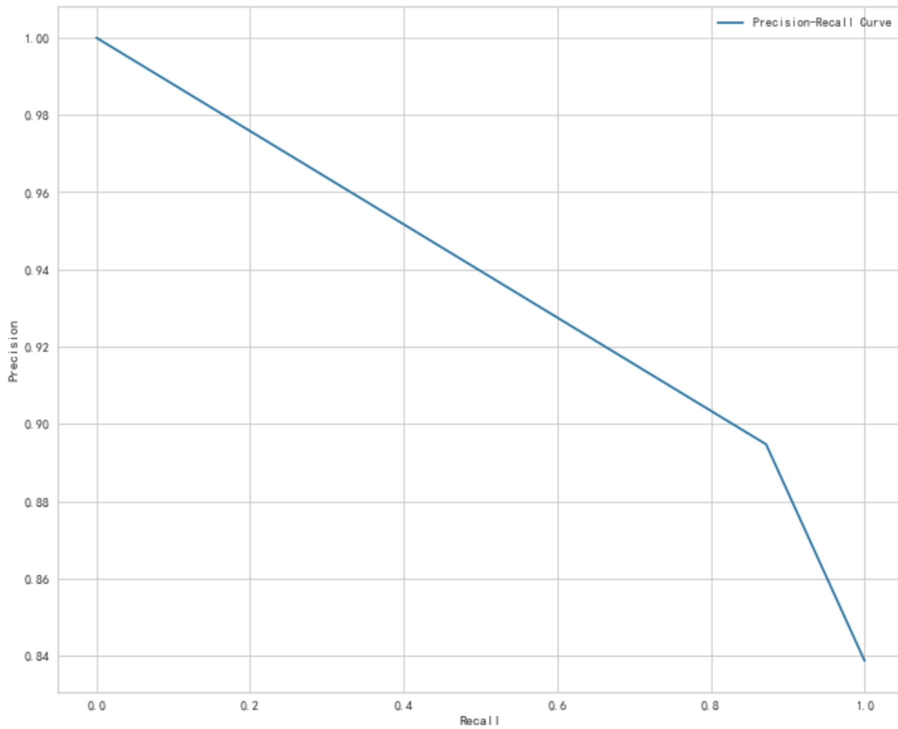

Figure 23: Precision Recall graph for random forest of African Americans vs. Whites (imbalanced data)

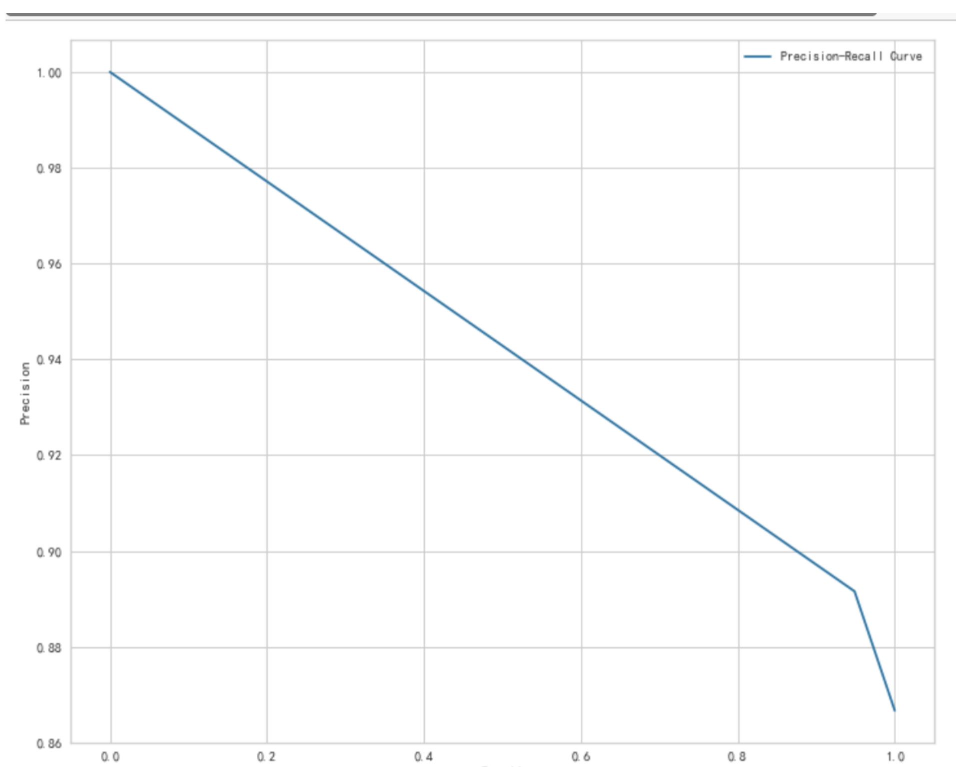

Figure 24: Precision Recall graph for random forest of Asians VS. Whites (imbalanced data)

The top three important features were as follows (See Figure 35 to Figure 38):

	White VS African American				White VS Asian			
	Balanced train set							
	Balanced test set		Imbalanced test set		Balanced test set		Imbalanced test set	
1	b_coef_3	0.199	b_coef_3	0.183	height	0.237	height	0.244
2	b_coef_4	0.105	b_coef_4	0.127	weight	0.114	weight	0.143
3	f_coef_4	0.087	f_coef_4	0.092	Back_crotch	0.095	max_hip	0.115

Table 15: Feature importance of Random Forest

The curve coefficients were the determining factor in predicting the race for the White vs. African American group, whereas height was significant in predicting the race for the White vs. Asian group. Random forest is less sensitive to data imbalance, leading to a smaller reduction in accuracy.

5.7 XGBoost

XGBoost is known for its speed and performance as it optimizes the gradient boosting algorithm to achieve high accuracy and fast training. Again, we used SMOTE to improve the data imbalance problem. Then, we performed grid search and cross-validation to find the best combination of hyperparameters and increase the stability and robustness of the model. We considered the following hyperparameter set:

n_estimators	max_depth	learning_rate
[100, 300, 500, 1000]	[3, 4, 5, 10]	[0.001, 0.01, 0.1]

Table 16: XGBoost Hyperparameter Selection

Definition of each Hyperparameter:

Hyperparameter Name	Definition
n_estimators	Total number of trees
max_depth	Maximum depth of each individual tree
learning_rate	The step size of each gradient descent update

Table 17: Definition of the hyperparameter

The best hyperparameter set yielded the following results (See Figures 25 to 26 and Figure 39 to 40):

	White VS African American	White VS Asian
	Balanced train set	

	Balanced test set	Imbalanced test set	Balanced test set	Imbalanced test set
n_estimators	1000	300	1000	500
max_depth	10	4	4	10
learning_rate	0.1	0.1	0.01	0.1
Accuracy	0.88	0.81	0.92	0.85

Table 18: Results of XGBoost

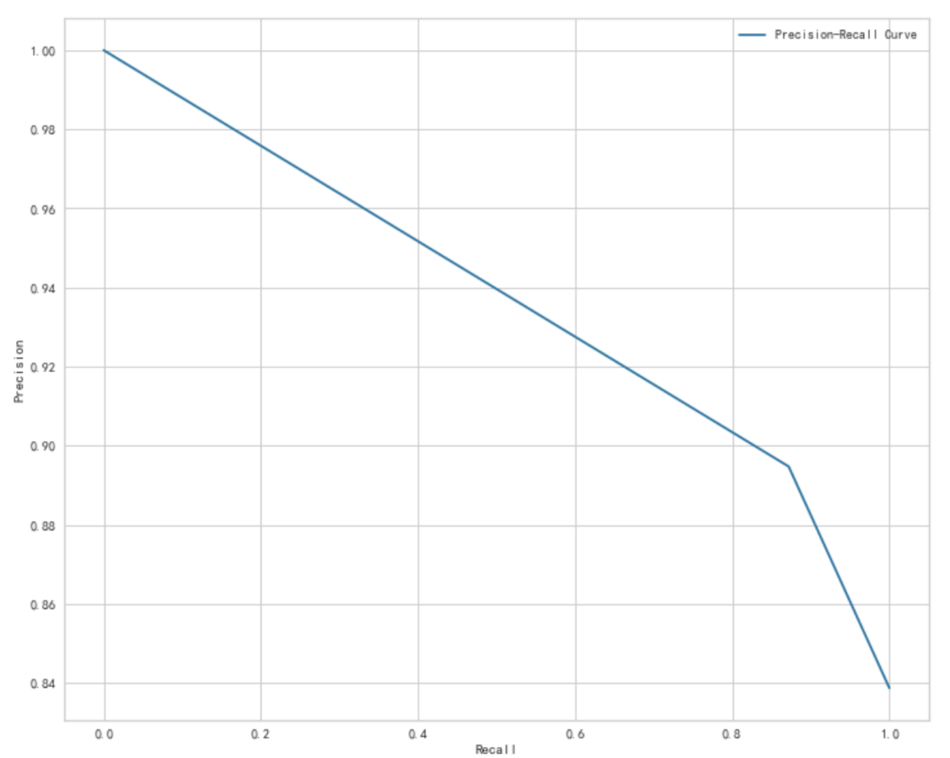

Figure 25: Precision Recall graph for XGBoost of African Americans vs. Whites (imbalanced data)

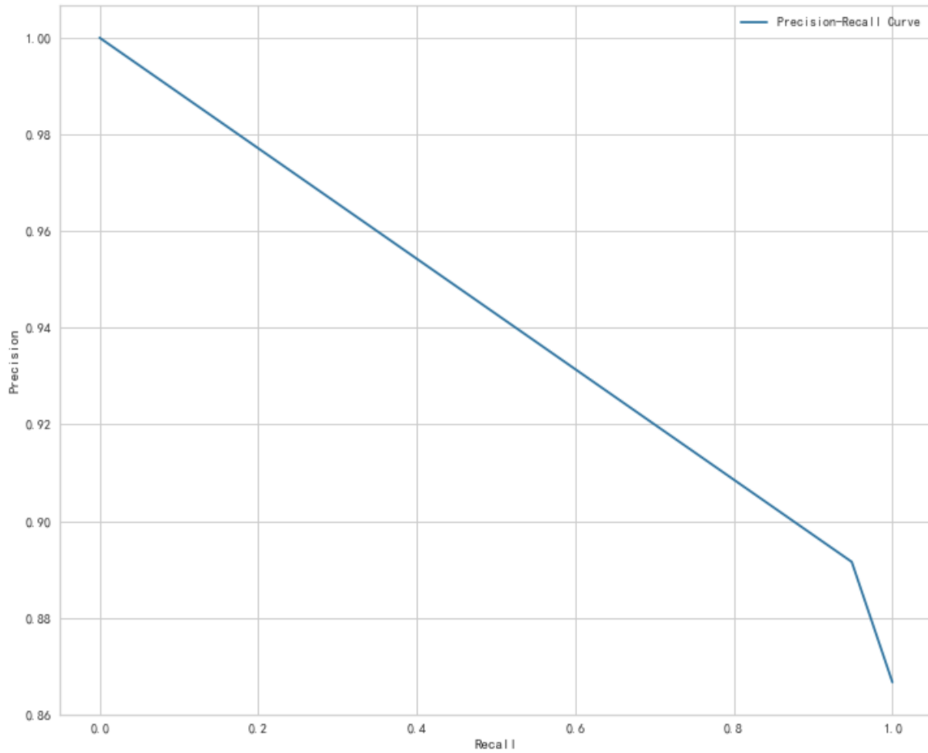

Figure 26: Precision Recall graph for XGBoost of Asians VS. Whites (imbalanced data)

The top three important features we have are as follows (See Figures 41-44):

	White VS African American				White VS Asian			
	Balanced train set							
	Balanced test set		Imbalanced test set		Balanced test set		Imbalanced test set	
1	b_coef_3	0.230	b_coef_3	0.206	Height	0.204	height	0.263
2	b_coef_4	0.085	b_coef_4	0.091	Crotch curve length at back waist	0.124	weight	0.138
3	f_coef_4	0.078	depth	0.084	Weight	0.107	f_coef_1	0.084

Table 19: Feature importance of XGBoost

The above results suggest that the curve coefficients were useful in distinguishing between Whites and African Americans, while height was useful in distinguishing between Whites and Asians. We also found a substantial improvement in the accuracy of race prediction using the Random Forest and XGBoost methods compared to logistic regression. This suggests the presence of interactions between covariates in the individual data.

VI. Conclusion

Weight and maximum hip circumference are effective predictors of crotch curve length at the back waist, while weight, height, and maximum hip circumference are useful predictors of anterior-posterior length. Body shape is influenced by several factors, including race and BMI. However, a combination of body measurements and front and back curve coefficients can predict race with reasonable accuracy. These predictions are affected by interactions between covariates, such as maximum hip circumference and BMI.

We experienced several limitations in our analysis. Due to uneven sample sizes, the reliability of curve predictions varies among different racial cohorts. The sample sizes of Asian, Hispanic, and other individuals are insufficient to make accurate and reliable curve predictions. Other covariates, such as age and body fat ratio, may have a stronger influence on the shape of the curves than the variables present in the data. Moreover, we were unable to remove the non-convexity, which may have been caused by measurement error, from the curves. Addressing this might result in curves that appear more natural.

Future researchers can use techniques other than ridge regression, such as kernel regression, support vector regression, and neural networks for predicting the coefficients and research how each coefficient affects the body shape.

VII. References

Cho, Y., Komatsu, T., Inui, S., Takatera, M., Shimizu, Y., & Park, H. (2006). Individual Pattern Making Using Computerized Draping Method for Clothing. *Textile Research Journal*, *76*(8), 646–654. https://doi.org/10.1177/0040517506066966

Farrell-Beck, J. A., & Pouliot, C. J. (1983). Pants Alteration By Graphic Somatometry Techniques. *Home Economics Research Journal*, *12*(1), 95–105. https://doi.org/10.1177/1077727x8301200112

Gu, B., Lin, W., Su, J., & Xu, B. (2017). Predicting distance ease distributions on crotch curves of customized female pants. *International Journal of Clothing Science and Technology*, *29*(1), 47–59. https://doi.org/10.1108/ijcst-11-2015-0122

Hu, Y., Xia, Y., & Gu, B. (2023). An image-based shape analysis approach and its application to young women's waist-hip-leg position. *Ergonomics*, 1–17. https://doi.org/10.1080/00140139.2023.2184366

McKinney, E., Gill, S., Dorie, A., & Roth, S. (2016). Body-to-Pattern Relationships in Women's Trouser Drafting Methods. *Clothing and Textiles Research Journal*, *35*(1), 16–32. https://doi.org/10.1177/0887302x16664406

Song, H. K., & Ashdown, S. P. (2011). Categorization of lower body shapes for adult females based on multiple view analysis. *Textile Research Journal*, *81*(9), 914–931. https://doi.org/10.1177/0040517510392448

Wang, J., Li, X., Pan, L., & Zhang, C. (2021). Parametric 3D modeling of young women's lower bodies based on shape classification. *International Journal of Industrial Ergonomics*, *84*, 103142. https://doi.org/10.1016/j.ergon.2021.103142

Xia, Y., Shao, J., Zhang, B., Jin, S., Li, Y., & Gu, B. (2022). Individualized generation of young women's crotch curve based on body images. *International Journal of Industrial Ergonomics*, *90*, 103296. https://doi.org/10.1016/j.ergon.2022.103296

Zhang, Z., & Qiao, L. P. (2013). Affect the Comfort of Pants Crotch Curve Analysis of the Factors. *Applied Mechanics and Materials*, *477-478*, 359–362. https://doi.org/10.4028/www.scientific.net/amm.477-478.359

VIII. Appendix

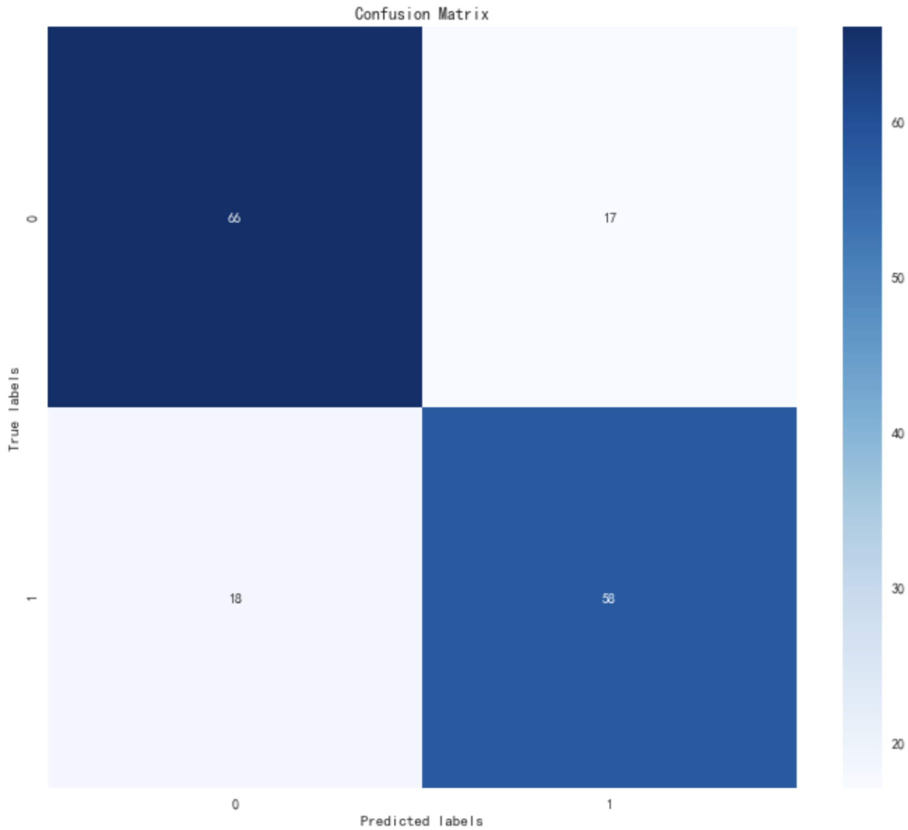

Figure 27: Confusion matrix for logistic regression of African Americans vs. Whites (balanced data)

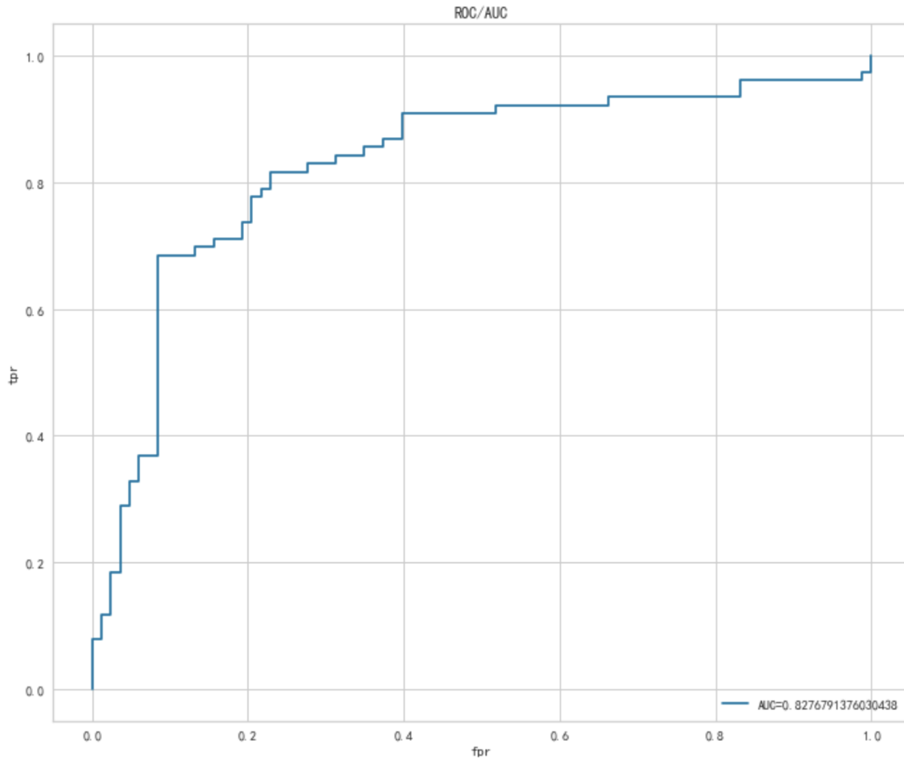

Figure 28: ROC/AUC curve for logistic regression of African Americans vs. Whites (balanced data)

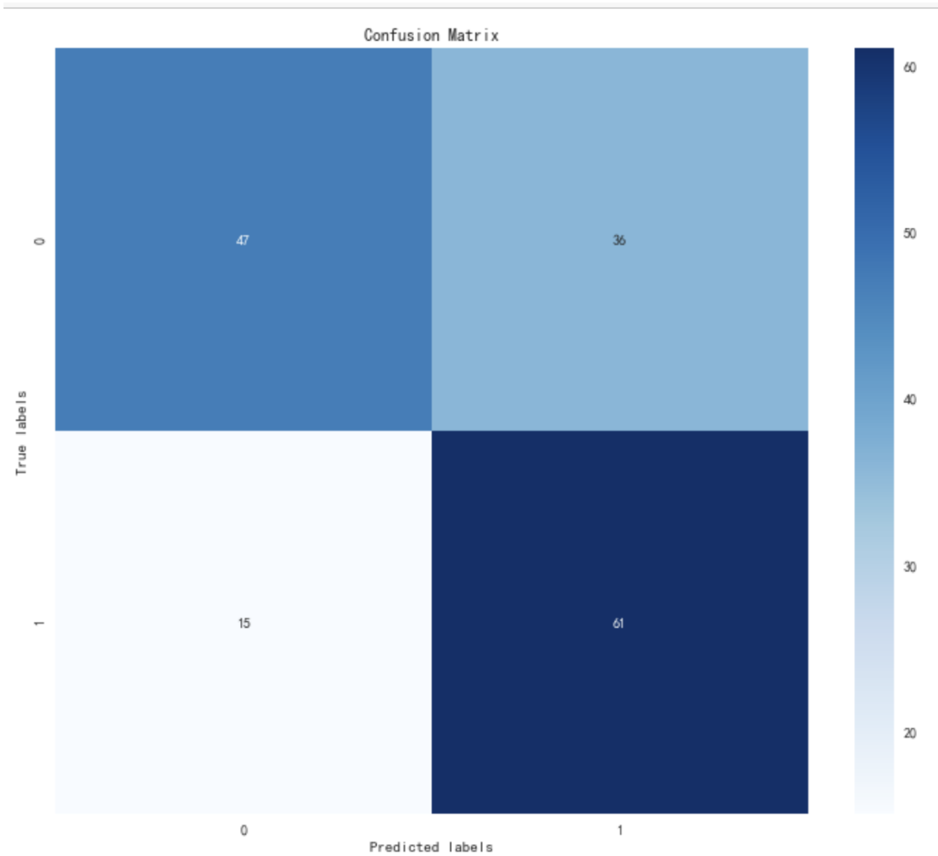

Figure 29: Confusion matrix for logistic regression of Asians VS. Whites (balanced data)

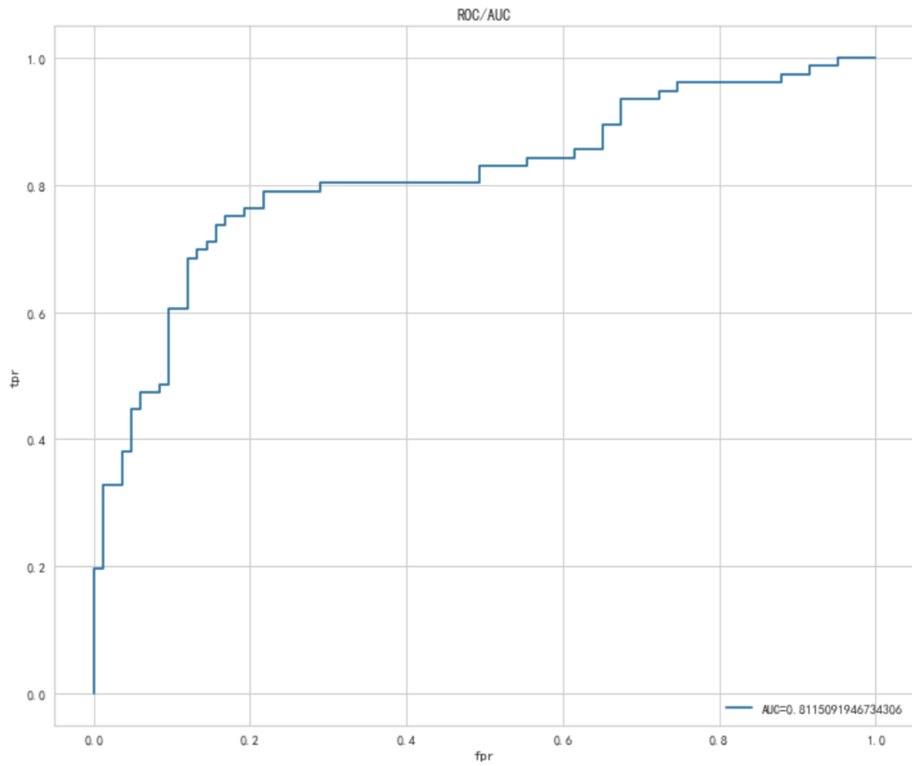

Figure 30: ROC/AUC curve for logistic regression of African Americans vs. Whites (balanced data)

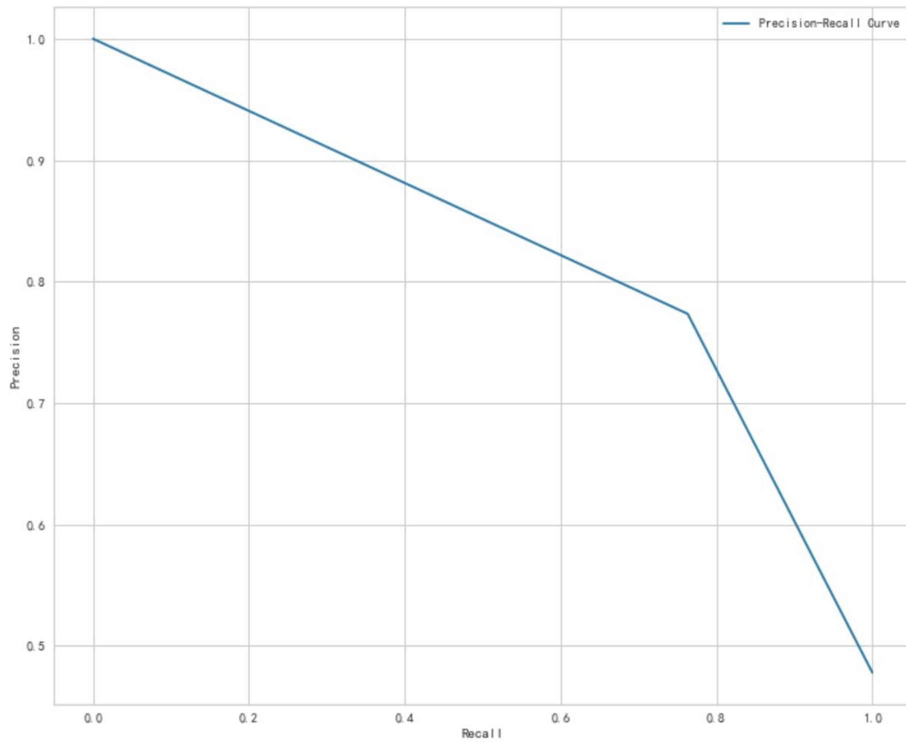

Figure 31: Precision Recall graph for logistic regression of African Americans vs. Whites (balanced data)

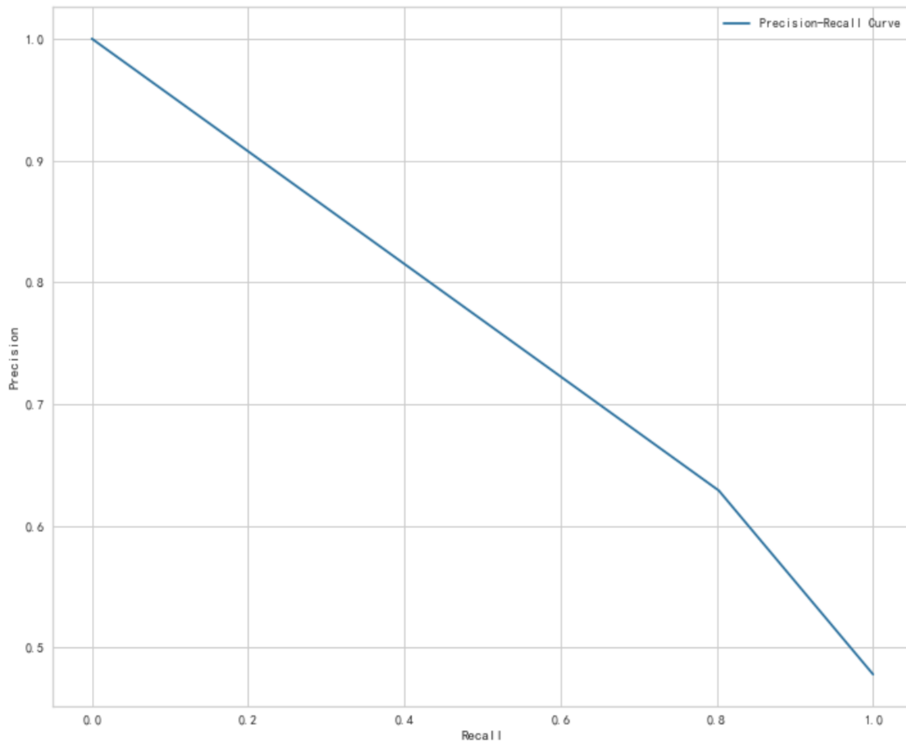

Figure 32: Precision Recall graph for logistic regression of Asians vs. Whites (balanced data)

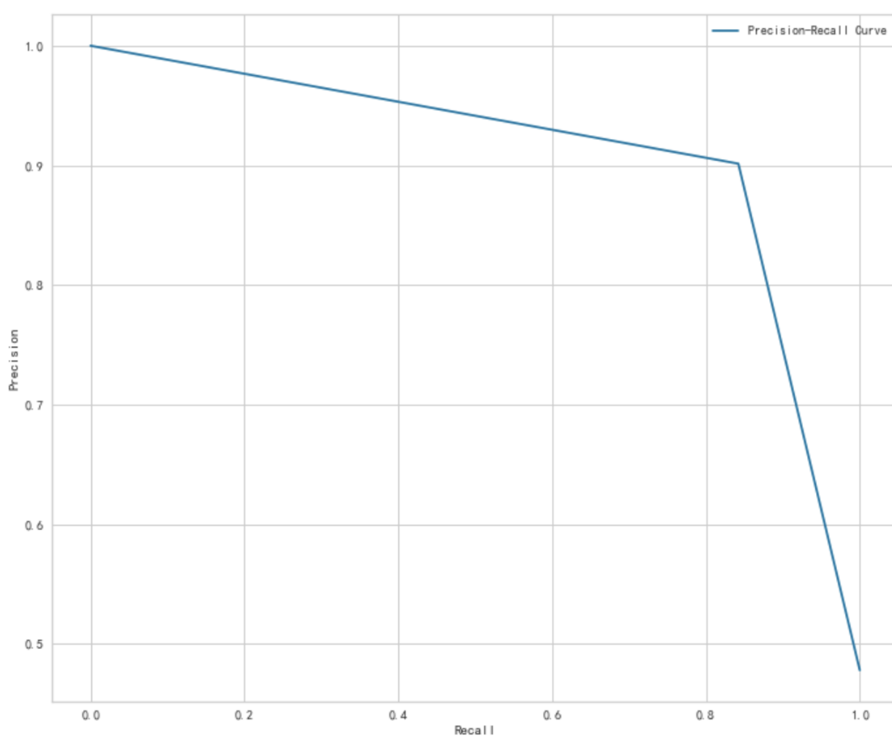

Figure 33: Precision Recall graph for random forest of African Americans vs. Whites (balanced data)

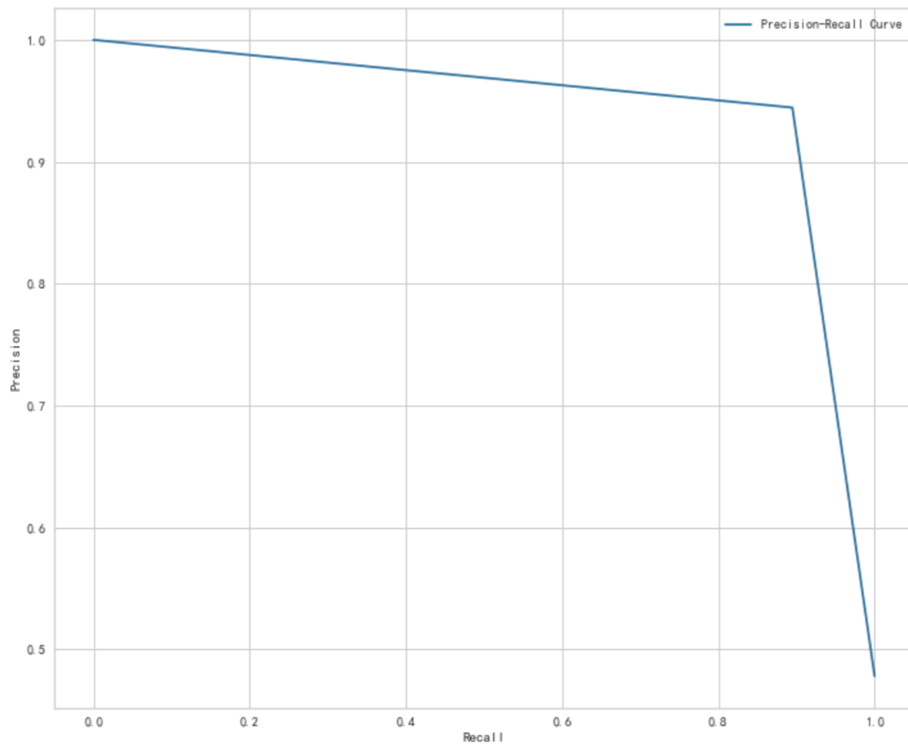

Figure 34: Precision Recall graph for random forest of Asians vs. Whites (balanced data)

	feature_importances
b_coef_3	0.199013
b_coef_4	0.105003
f_coef_4	0.086917
Depth	0.072752
Anterior_posterior_Length	0.058878
BMI	0.051467
Height	0.049448
weight	0.045973
Crotch_curve_length_at_back_waist	0.045769
b_coef_2	0.044281
f_coef_1	0.042012
Back_Crotch	0.040262
b_coef_1	0.038592
f_coef_3	0.032383
Max_Hip	0.029450
Front_Crotch	0.029337
f_coef_2	0.028462

Figure 35: Feature importance of random forest of African Americans vs. Whites (balanced data)

	feature_importances
b_coef_3	0.182649
b_coef_4	0.126609
f_coef_4	0.091740
Depth	0.062353
Anterior_posterior_Length	0.059638
weight	0.054556
b_coef_2	0.046573
BMI	0.044585
Back_Crotch	0.044245
f_coef_1	0.043114
b_coef_1	0.042175
Crotch_curve_length_at_back_waist	0.037581
Height	0.037529
f_coef_3	0.036816
Front_Crotch	0.032857
Max_Hip	0.029766
f_coef_2	0.027214

Figure 36: Feature importance of random forest of African Americans vs. Whites (imbalanced data)

	feature_importances
Height	0.236871
weight	0.113559
Back_Crotch	0.095086
Max_Hip	0.088442
b_coef_4	0.050609
Crotch_curve_length_at_back_waist	0.045086
Anterior_posterior_Length	0.044442
Depth	0.044111
BMI	0.039959
f_coef_1	0.034731
Front_Crotch	0.034530
f_coef_2	0.033403
b_coef_2	0.032548
b_coef_3	0.032388
f_coef_4	0.028899
f_coef_3	0.024618
b_coef_1	0.020717

Figure 37: Feature importance of random forest of Asians VS. Whites (balanced data)

	feature_importances
Height	0.244045
weight	0.142684
Max_Hip	0.114930
Back_Crotch	0.066185
b_coef_4	0.061560
f_coef_2	0.038607
Crotch_curve_length_at_back_waist	0.038080
BMI	0.036382
Anterior_posterior_Length	0.032881
Front_Crotch	0.032766
f_coef_4	0.031394
b_coef_3	0.030120
f_coef_1	0.030082
b_coef_2	0.030075
Depth	0.026151
f_coef_3	0.024979
b_coef_1	0.019078

Figure 38: Feature importance of random forest of Asians VS. Whites (imbalanced data)

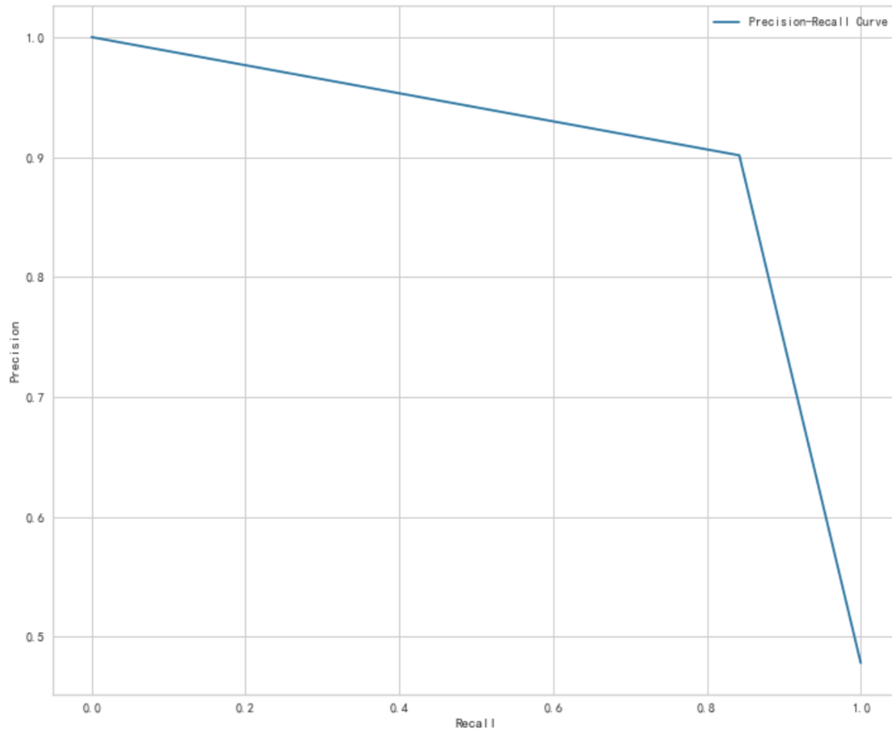

Figure 39: Precision Recall graph for XGBoost of African Americans vs. Whites (balanced data)

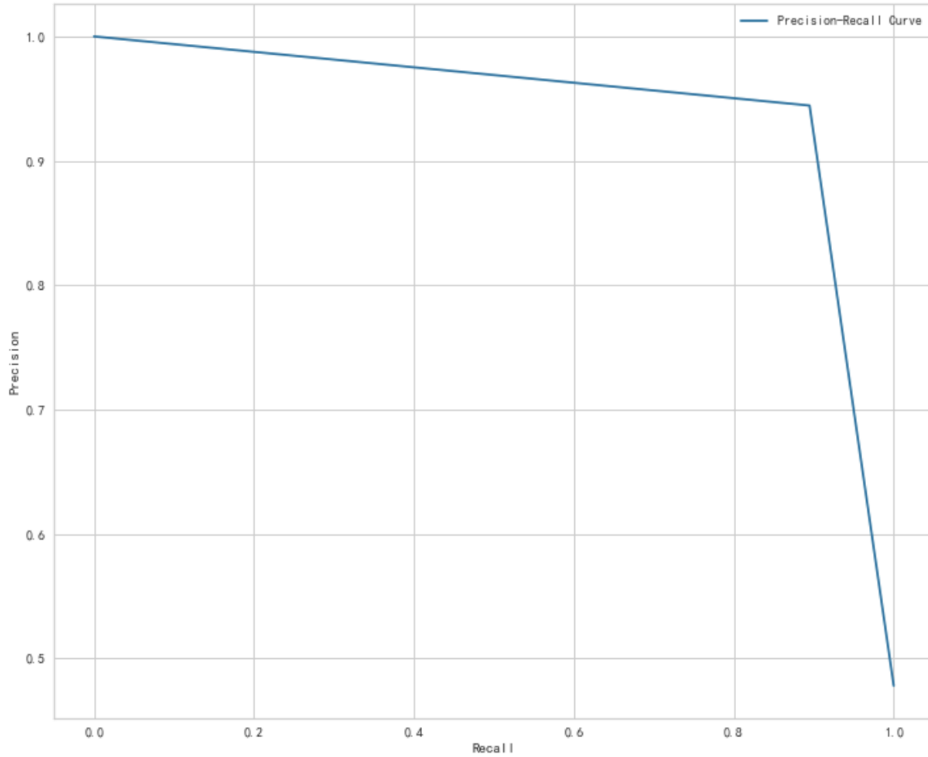

Figure 40: Precision Recall graph for XGBoost of Asians VS. Whites (balanced data)

	feature_importances
b_coef_3	0.230318
b_coef_4	0.085182
f_coef_4	0.078491
Depth	0.072858
b_coef_2	0.066037
f_coef_1	0.058607
Height	0.056326
BMI	0.049110
f_coef_3	0.043781
Crotch_curve_length_at_back_waist	0.042259
Back_Crotch	0.038142
Anterior_posterior_Length	0.037886
weight	0.034895
f_coef_2	0.034208
Front_Crotch	0.031730
b_coef_1	0.020436
Max_Hip	0.019734

Figure 41: Feature importance of XGBoost of African Americans vs. Whites (balanced data)

	feature_importances
b_coef_3	0.205559
b_coef_4	0.090995
Depth	0.084274
f_coef_4	0.072702
b_coef_2	0.063989
Back_Crotch	0.059975
f_coef_3	0.052599
weight	0.048849
f_coef_1	0.046140
b_coef_1	0.045140
Anterior_posterior_Length	0.041066
Front_Crotch	0.040013
BMI	0.034043
f_coef_2	0.032132
Crotch_curve_length_at_back_waist	0.029833
Height	0.029188
Max_Hip	0.023501

Figure 42: Feature importance of XGBoost of African Americans vs. Whites (imbalanced data)

	feature_importances
Height	0.204005
Crotch_curve_length_at_back_waist	0.124223
weight	0.107133
Max_Hip	0.074195
Back_Crotch	0.061916
f_coef_1	0.060594
b_coef_4	0.057069
BMI	0.050234
Depth	0.046393
Anterior_posterior_Length	0.038538
b_coef_3	0.032286
Front_Crotch	0.029537
b_coef_2	0.026666
f_coef_3	0.024937
f_coef_4	0.023988
f_coef_2	0.021283
b_coef_1	0.017002

Figure 43: Feature importance of XGBoost of Asians VS. Whites (balanced data)

	feature_importances
Height	0.263086
weight	0.138149
f_coef_1	0.084276
Back_Crotch	0.064503
Max_Hip	0.062201
b_coef_4	0.052268
b_coef_3	0.044046
Depth	0.037384
f_coef_2	0.034485
Crotch_curve_length_at_back_waist	0.032932
b_coef_1	0.031730
f_coef_4	0.031588
Front_Crotch	0.028609
BMI	0.026440
Anterior_posterior_Length	0.025322
b_coef_2	0.024331
f_coef_3	0.018649

Figure 44: Feature importance of XGBoost of Asians VS. Whites (imbalanced data)