Title
Predictive modeling of whole body dual energy X-ray absorptiometry from 3D optical scans using shape and appearance modeling

Author
Piel, Michaela

Publication Date
2017

Peer reviewed|Thesis/dissertation

Predictive modeling of whole body dual energy X-ray absorptiometry from 3D optical scans using shape and appearance modeling

by

Michaela Piel

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Biomedical Imaging

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgments

Many thanks go out to Dr. John Shepherd, my committee chair and advisor, for his continual and vital guidance, support, and patience. Thank you to Dr. Duan Xu, Dr. Janine Lupo, and Dr. Suneil Koliwad, my committee members, for their invaluable help and support throughout this research and thesis review. Without their guidance, this research would not be possible. Additional thanks go out to Dr. John Shepherd and Dr. Steven Heymsfield and the participants of the Shape Up! Adults Study for providing the data for my thesis. Many thanks to the John Shepherd lab team for their insight and assistance on various tasks throughout the research, especially Bennett Ng, Michelle Nie, Dr. Bo Fan, and Markus Sommer. You all have been a pleasure to work with, and I have gained much from knowing each of you.

Predictive modeling of whole body dual energy X-ray absorptiometry from 3D optical scans using shape and appearance modeling

Michaela Piel

Introduction: Malnutrition and lack of exercise have led to a steep increase in metabolic disorders worldwide. Even though diseases caused by malnutrition have become common, we still lack an accurate, inexpensive, and easily accessible method to assess a person's risk of developing metabolic diseases. In this work, I test a novel method called 3D optical body composition that I hypothesized would be relatively accessible, accurate, and inexpensive.

Methods: The Shape Up! Adults Study is recruiting 720 adults for measures that include whole-body DXA and 3D optical scans. Like image types were spatially registered using 105 and 75 fiducial points for DXA and 3D optical scans respectively. Statistical appearance and shape modeling were then performed on each image type. The sex-specific population variances for shape (3D Optical) and bone, fat, and lean appearance (DXA) were captured as Principal Components (PCs) resulting in 8 PC models 4 for females and 4 for males. Stepwise linear regression was used to predict DXA PCs from 3D optical PCs and other anthropometric and demographic measurements. The predicted DXA PC coefficients of each participant were then inverted to create a pseudo-DXA image. Additionally, k-means cluster analysis was performed on the participants' predicted DXA fat PC coefficients to determine different body phenotypes of males and females and corresponding health risks.

Results: A total of 72 men and 104 women were available at the time of the analysis. To describe 95% of the population variance in men, it required the following number of PC modes: 10 (optical), 32 (fat), 35 (lean), 35 (bone), with women having similar results. The pixel-to-pixel differences in mass between actual and predicted DXA values had no mean bias for all models. The difference in the pixel values had root mean square errors (RSMEs) of 0.015 g of fat, 0.023 g of lean, and 0.012 g of bone for the female

data, and 0.013 g, 0.024 g, and 0.018 g for the male data respectively. These RMSE values were less than 5% of the maximum pixel value within the population. Lastly, I found 9 female and 5 male phenotypes of body fat that were related to unique metabolic characteristics and risk factors.

Conclusion: Whole body and regional distributions of fat, lean, and bone can be accurately predicted from 3D optical scans. With this accessible and accurate method, body composition and metabolic risk phenotype can now be defined in individuals. Our hope is that this will increase awareness of metabolic risks and motivate those at high risk to seek medical advice for risk-reducing strategies.

Table of Contents

List of Tables

List of Figures

Section 1

Introduction

1.1. Motivation

Malnutrition and lack of exercise have led to a steep increase in metabolic disorders worldwide. One in eleven people in the United States have diabetes, one of the many metabolic disorders [19]. Even though diseases caused by malnutrition have become common, we still lack an accurate, inexpensive, and easily accessible method to assess a person's risk of developing metabolic diseases. This is especially true for low- and middle-income countries [7]. The two most widely used measurements are body mass index (BMI) and waist circumference. BMI only compares a person's weight to their height. Weight, however, is just the sum of fat and lean masses in the body. This is why athletes, with high lean masses, often are categorized as obese by BMI. Waist circumference is used in an effort to estimate the amount of excess adiposity in the abdomen. This parameter still falls short of fully describing whole body adiposity status. Even though height, weight, and waist circumference are currently the most accessible measurements because they only require a scale and measuring tape, they can only go so far in describing the regional distribution of fat and muscle. The ideal technology for quantifying body composition and its health risk factors would be accessible and inexpensive enough to allow for most of the population to use it, accurate to the gold standard of body composition, provide regional body composition measures, easy to use such that the operators do not need special training or licenses, and precise enough to monitor changes in body composition over time.

In this work, I test a novel method called 3D optical body composition that I hypothesized would meet these ideal conditions. I start with introducing body composition and describe how it is measured using the gold standard, a dual-energy X-ray absorptiometry (DXA) scan. Then I discuss how a 3D optical scanner works and how it is implemented to take scans appropriate for estimating body composition.

After that, I explain principal component analysis, which is used to create statistical models of the data. From there, equations can be formulated that relate the 3D optical data to the gold standard DXA data. These relationships can later be used to predict a person's body composition as if measured by DXA from just their 3D optical scans. Finally, I introduce k-means clustering to cluster the female and male populations into different phenotypes. I describe these phenotypes in terms of their unique health attributes that may be useful in understanding the phenotypical metabolic risk. If this method is successful, it would lead to opportunistic screening catered to each individual, earlier intervention, and ultimately the reduction of metabolic health-related morbidity and mortality.

1.2. Background

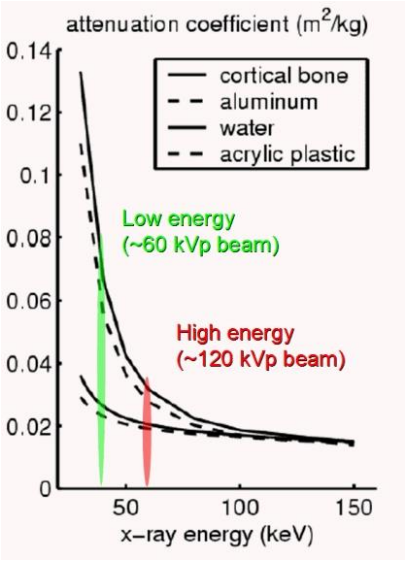
1.2.1. Body Composition

Body composition can be described using different subdivisions or compartments, from a single compartment of whole body mass, to the 5-compartment molecular model, and ultimately even an atomic model [20]. One of the most popular models represents the body as composed of two-compartments, fat mass, or triglycerides, and fat-free mass. Note that fat-free mass contains all mass that is not triglycerides. On this tissue level, this indicates that muscle is grouped with the bone. The two-compartment model is popular since it can accurately quantify fat mass even though the lean masses are grouped together. This study uses the more sophisticated three-compartment model, in which the fat mass, lean mass, and bone mineral content (BMC) are estimated [13]. The clinical gold standard for three-compartment body composition is DXA; however, this technology has accessibility limitations. I discuss it more completely in the next section.

1.2.2. Dual Energy X-ray Absorptiometry

A whole body DXA scanner is similar to a typical medical X-ray device, but instead of taking one X-ray image, it uses two low radiation X-ray exposures with different average energy levels: one low-energy level (~ 40 keV average) and one high-energy level (~ 70 keV) [4]. After the low- and high-energy X-ray beams pass through the body, the transmitted X-rays of each beam are captured by a detector. The X-ray

absorption in the patient is related to the energy-dependent attenuation coefficients of the components of the body. Figure 1.1 shows that, for a low energy X-ray beam, the attenuation coefficient of bone is much higher than the attenuation coefficient of water, which is similar to attenuation coefficient of soft tissue. Conversely, the attenuation coefficients of bone and soft tissue are more similar for a high-energy X-ray beam. By weighting and then subtracting the low- and high-energy images, a specific material's signal can be nulled. For example, to obtain a soft tissue image of the body, the low and high-energy images would be weighted and then subtracted in such a way that the resulting image would have no bone signal [9]. In other words, two equations can be solved for two unknowns. If no bone is present, fat and lean soft tissue mass can be determined. If bone is present, only bone mass and soft tissue mass with an assumed percent fat can be solved [13].



[9]

Figure 1.1. Graph of the effect of X-ray energy on the attenuation coefficient of different materials.

Using these techniques on a pixel level can generate fat, lean, and bone mineral mass DXA images as shown in Figure 1.2. DXA is currently the gold standard for measuring body composition; however, it is only located in well-supplied medical centers, making it relatively inaccessible. Also, because a DXA scanner is an X-ray device, substantial training and special licenses are typically required to operate the scanner.

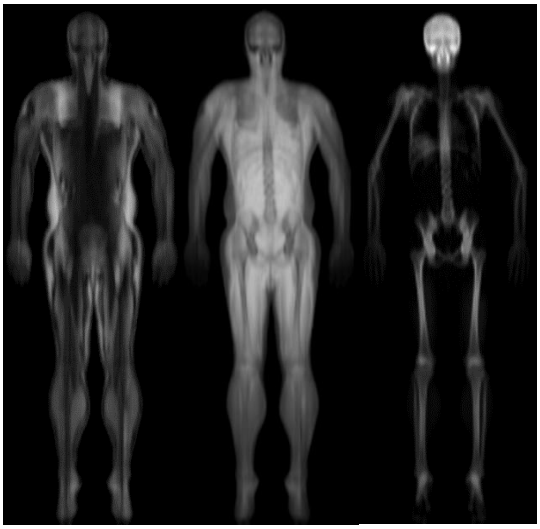
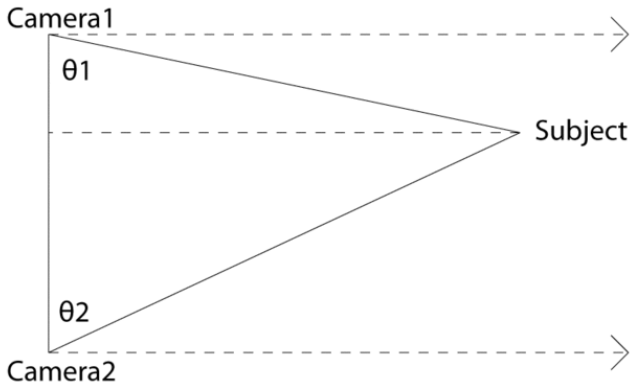


Figure 1.2. Example of DXA fat (left), lean (middle), and bone (right) images.

1.2.3. Three-Dimensional Optical Scans

3D optical devices, in contrast to DXA scanners, are cheap, easy to operate, and accessible. They are so accessible, in fact, that many fitness centers have them and popular video game devices use them. A 3D optical scanner has a projector that illuminates the subject with an infrared light pattern and two infrared cameras that capture the reflected light pattern after it has been deformed by the shape of the subject [8]. Triangulation, which is illustrated in Figure 1.3, is used to calculate the depth of each point in the field of view. Based on where the light reflected from a specific body surface point is received and recorded in each camera image, the angles ($\theta 1$ and $\theta 2$) between the surface and each camera can be determined. With the known distance between the two cameras, we know two angles and a side of the triangle formed by the transmitter, the receptors and the reflection point of the subject. From this information, the depth of the body surface point can be calculated [18]. Repeated measurements of body surface points in this fashion render the 3D depth map of the body.



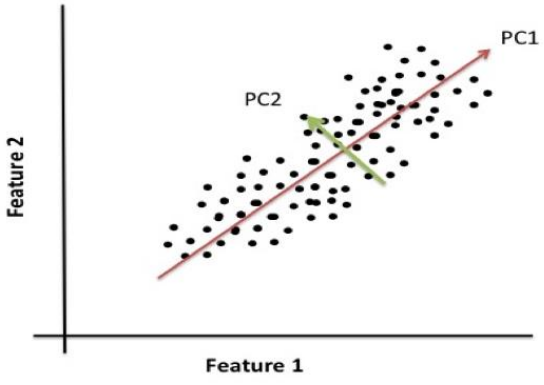
[18]

Figure 1.3. Diagram of how a 3D optical scan works.

For health purposes, 3D optical scanners are often used to obtain anthropometric measurements such as waist circumference, hip circumference, thigh girth, and arm girth. While this information can give a good insight into a person's health status, a 3D optical scan is only able to determine external body measurements.

1.2.4. Principal Component Analysis

Principal component analysis (PCA) is a dimensionality reduction method that describes the variance in the data. The first principle component (PC) contains the most variance in the data. To find the second PC, only the directions that are orthogonal to the first PC are considered, but the direction that has the next highest variance in the data is determined to be the second PC. This is the same process that is used to find all the subsequent PCs until all of the variance in the data has been accounted for [14]. Figure 1.3 shows the first two PCs for an example dataset. Often, only the PCs that describe 95% of the variance in the data are used in analysis since the last 5% of the variance most likely describes the noise of the data.



[15]

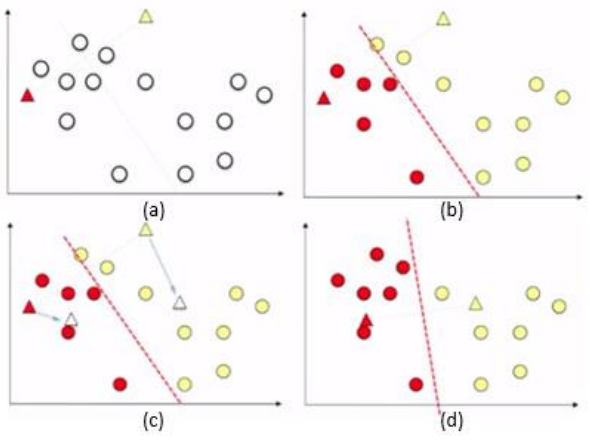
Figure 1.4. Example of the PCs of 2D data.

In this study, PCA is performed to generate the PCs that are then used in statistical models of the data. Each participant's data is assigned a coefficient for each PC of the data. The coefficient shows how the participant compares to the average data for that PC. If the participant's data is equal to the mean for a PC, their corresponding coefficient for that PC would be equal to zero. If their data is in the extremities for a PC, their corresponding coefficient for that PC would be a highly positive or negative value. With these created PCs, a new participant with data that was not used to determine the PCs can be added to the data and their PC coefficients can be estimated. PCA can also be used for DXA data, and it is possible to estimate a person's DXA PC coefficients from their 3D optical PC coefficients using linear methods like stepwise linear regression. Another feature of PCA when applied to images is that the PC coefficients can be inverted back into real space to create the associated image.

1.2.5. K-Means Clustering

K-means is a clustering algorithm that divides data into k number of clusters. The algorithm essentially finds the clusters that minimize the total within the sum of the squares, which represents the total variance within the clusters [16]. The user-inputted k number of centroids are randomly seeded in the data (Figure 1.5a). The Euclidean distance between all of the data points and the centroids are then calculated, and the centroid that is closest to each data point is determined to be the cluster that the data point is a part of (Figure 1.5b). Next, each centroid is recomputed so that it is located in the center of its determined cluster

of data points (Figure 1.5c). With these new centroids, the data points are reclustered based on the centroid to which they are now closest (Figure 1.5d). The process is repeated until no data points change clusters from one iteration to the next. K-means cluster analysis can be performed on any set of data, either raw data, after normalization, or PCs.



[17]

Figure 1.5. Steps of k-means clustering.

Section 2

Materials and Methods

This study design is a cross-sectional, retrospective, comparative analysis. Figure 2.1 shows a flow chart of the overall process in which this experiment is carried out.

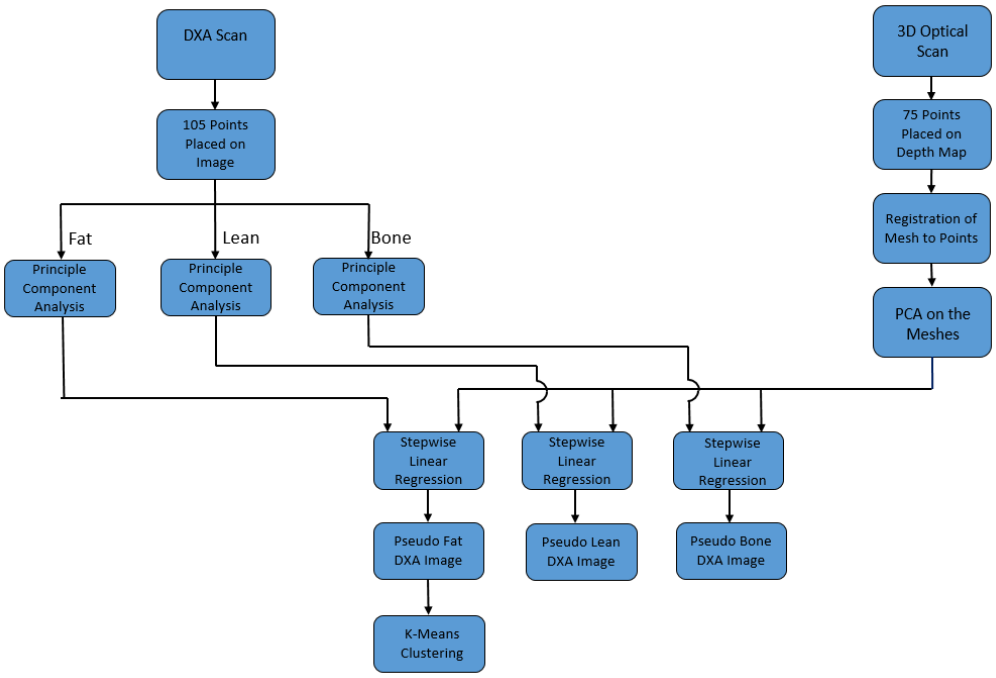


Figure 2.1. Flow chart of the experimental methods. These methods are done separately for females and males.

Whole body DXA and 3D optical scans of participants in the Shape Up! Adults Study were used in this analysis. Images were spatially registered using 105 and 75 fiducial points for DXA and 3D optical scans, respectively. Statistical appearance and shape modeling were then performed on each image type. The sex-specific population variances for shape (3D Optical) and appearance (DXA) were captured as PCs resulting in 8 PC models, 4 for females and 4 for males. Stepwise linear regression was used to predict DXA PCs from 3D optical PCs and other anthropometric and demographic measurements. The predicted

DXA PC coefficients of each participant were then inverted to create a pseudo-DXA image. Additionally, k-means cluster analysis was performed on the participants' predicted DXA fat PC coefficients to determine different body phenotypes of males and females and corresponding health risks.

2.1. Participants

The data used in this study were collected as part of the ongoing Shape Up! Adults Study. Seven hundred twenty individuals are being recruited for the Shape Up! Adults Study at both the University of California San Francisco (UCSF) and Pennington Biomedical Research Center. Each participant in the study has a whole body DXA scan, 2D and 3D optical scans, blood tests, a muscle strength test, blood pressure measurements, anthropometry measurements, and bioimpedance spectroscopic analysis. In addition, they must give written, informed consent and fill out a questionnaire to participate. An in depth description of the study can be found in [6].

2.2. Image Acquisition and Analysis

The whole body DXA scans were completed on a Hologic Horizon/A system (Hologic Inc., Marlborough, MA, USA). Following the standards of a DXA scan, each participant laid on their back in the center of the scanner with their arms by their sides and toes pointed down. If the patient was too large to fit onto the DXA table in this position, they were instead positioned so that their left arm hung off the table, causing the lower half of the left arm to be out of the field of view. A whole body DXA scan was then performed by a trained physician. Using the Hologic APEX 3.0 software, the scans were analyzed, DXA measurements of each participant were calculated, and the data was saved for later analysis. For the whole body 3D optical scan, each individual was imaged on a Fit3D Proscanner (Fit3D Inc., Redwood City, CA, USA) in form-fitting clothing as to disclose the shape of their body in their image. Each patient stood still on a circular platform with their legs slightly apart; their arms were kept down at their sides by holding onto provided handle bars. When the scan took place, the platform slowly turned 360 degrees in order to get a whole 3D view of the patient. The depth map of the participant was then automatically created on

the scanner and transferred to a computer in the .obj file format.

After image acquisition, analysis was done on both the DXA and 3D optical scans in preparation for statistical analysis. In house MATLAB 2016 (MathWorks Inc.) code was used to convert the raw Hologic DXA scan data to fat, lean, and bone DXA images for each participant. All analyses described below were done independently for females and males, as well as separately for each DXA image type (fat, lean, and bone).

Next, 105 points were placed on the same set spots of each individual in order to define the shape, texture, and appearance of each participant. Instead of manually placing all 105 points on each participant, a model based on the Coote's Random Forest Regression Voting Constrained Local Model [3] was used to place the points. The model used was trained on 1,000 images from a different study [5] and is now able to accurately place the points on new images. The model placed the points on all the participant's DXA scans using the active shape modeling toolkit software (Visual Automation Limited, Manchester, U.K.). All the points were then reviewed to check for accuracy by two trained researchers, and the points were fixed as necessary. The process of placing the points on a DXA image is shown in Figure 2.2.

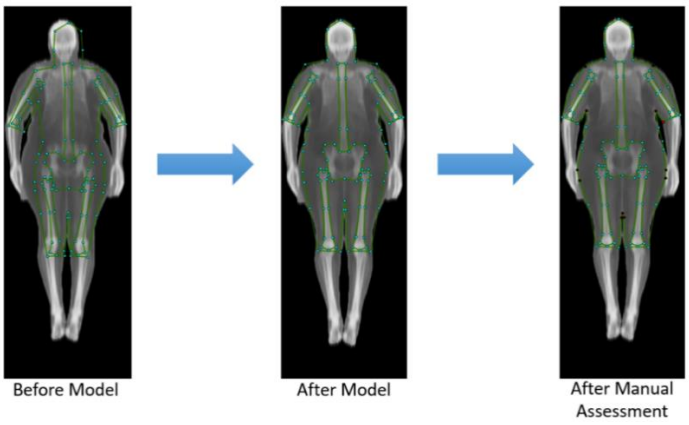


Figure 2.2. Process of point placement on DXA images. The image on the left shows the randomness of the points before the model is applied. The middle images shows the locations of the points after the model has placed the points. The image on the right shows the location of the points after manual assessment. The red points represent the points that manually corrected.

Seventy-five points were manually placed on each depth map for the 3D optical images by two trained

researchers using the program MeshLab 2016 [10]. These points were defined by the CAESAR study and are shown in Figure 2.3 [1]. After the point placement was completed, the points were used to register a mesh template to each participant, so that each participant's mesh consisted of 60,000 vertices. The points were input for the registration so that the algorithm knew where important landmarks were on a participant's body, such as the nose and the shoulders. The template could then be registered so that the templates important landmarks were fit to match the participant.

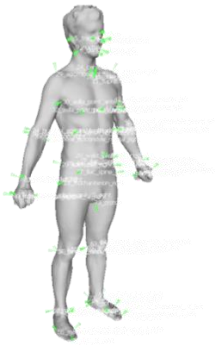


Figure 2.3. The green points represent the locations of where the 75 points were placed on the 3D depth map.

2.3. Statistical Analysis

2.3.1. Principal Component Analysis

From the annotated DXA images and the 3D optical mesh images, PCA was used to create statistical appearance and shape models using the am_build_apm tool in the program am_tools [2]. PCA was used specifically for determining the variation in the shape and texture of the participants due to differences in their body composition. The 3D data only contained shape information, but the DXA data contained shape information for each participant's body, texture information from different tissues in the body, and by combining shape and texture, appearance information was obtained. DXA analysis was performed on the appearance information of the data. The output of the PCA showed the variation in the form of eigenvectors, with each eigenvector being a different PC. For each image, each coefficient used in combination with each eigenvector was outputted. The coefficient showed how the participant compared to the average for that PC. The outputs of the PCA for both DXA and 3D optical scans were then

uploaded into the statistics software SAS™, version 9.3 (SAS Institute Inc., Cary, NC, USA). The proc CORR procedure was used to determine the correlations between the 3D optical PCs and anthropometric, demographic, blood, and DXA measurements and to determine the correlations between the DXA PCs of each image type and the same set of measurements.

2.3.2. Stepwise Linear Regression

The PCA coefficient data was also arranged into a matrix where the rows represented each participant and the columns represented both the PCs that described 95% of the DXA variance and the PCs that described 95% of the 3D optical variance. Stepwise linear regression was used to create equations that used a combination of the 3D optical PCs and other demographic variables that I saw fit to predict each one of the DXA PCs using RStudio (RStudio, Inc., Boston, MA) [11]. Equation 1 shows the overall form of the equations, where the coefficients were determined by the stepwise linear regression, because from the data provided, all measurement variables were known. An equation was created for each of the DXA PCs that described 95% of the variance, where *n* is the DXA PC number, *a* through *z* represent the weighting of each 3D optical PC used to optimally predict the DXA PCs, and *m* represents the number of 3D optical PCs required to describe 95% of the 3D optical data's variance.

$$PCn_{DXA} = aPC1_{Opt} + bPC2_{Opt} + cPC3_{Opt} + \cdots + zPCm_{Opt} + \alpha Weight + \beta Height + \gamma BMI + \rho Age + \upsilon Ethnicity \quad \textbf{(1)}$$

Because forward stepwise linear regression was used, not all measurement variables were used in every equation. A variable was only added to the previous version of the equation if it improved the Akaike information criterion (AIC) of the equation. The AIC essentially attempts to maximize the likelihood function for the model. When the AIC measurement is minimized, the likelihood function is maximized. When determining if a certain variable should be added to the equation, an evaluation of the variable's effect on the AIC is performed. If the variable does not lower the AIC, it is not added.

Finally, the equations created above were used to predict each participant's DXA PCs. These predicted

DXA PCs of each participant were then uploaded into am_tools and the am_make_image tool was used to predict each participant's DXA image [2]. Equation 2 is the equation that am_tools uses to predict the image, where $PC1_{DXA}$-PCn_{DXA} are the participant's predicted coefficients for each PC. The equation essentially converts the data from PC space to image space. All of the PCs that described 95% of the DXA variance were used in this equation.

$$DXA\ Image = APC1_{DXA} + BPC2_{DXA} + CPC3_{DXA} + \cdots + ZPCn_{DXA} \quad \textbf{(2)}$$

To determine how accurate the models are at predicting DXA fat, lean, and bone images, all of the participants' predicted DXA images were compared to their actual DXA images for each image type. This was done visually by using MATLAB to create difference images. For a quantitative assessment of the accuracy, I computed the pixel Root Mean Square Error (RMSE) of each predicted DXA image in MATLAB and averaged the RMSEs of the participants to get the overall RMSE for each image type. I concluded the estimation to be successful if the RMSE was less than 0.034 g, which is 5% of the maximum number of grams in a pixel across the population, and if the difference image had no net bias.

2.3.3. Cluster Analysis

Additionally, the predicted DXA fat PCs of all of the participants were clustered in order to determine different body phenotypes of the population. The cluster analysis was done using k-means in RStudio. To determine the optimal number of clusters for the participant population, k-means was performed with various numbers of clusters; the number of clusters was then plotted against the total within the sum of the squares. An example of this plot is shown in black in Figure 2.4. The total within the sum of the squares represents the total variance within all the clusters. The less variance within the clusters, the better, but after a certain threshold, splitting the clusters any more will not significantly reduce the variance within the clusters. This threshold was determined using the elbow method, which is represented by the blue dotted lines in Figure 2.4. Once the optimal number of clusters was determined, k-means was performed using that number of clusters as an input. Phenotypes were then created for each cluster and

health attributes of each cluster were determined using a pairwise t-test. [12]

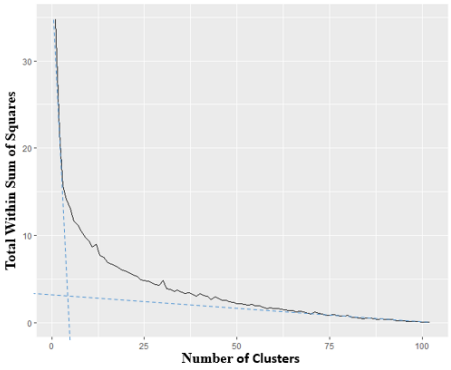


Figure 2.4. Graph that shows the total variance within the clusters as a function of the number of clusters used in k-means clustering. The blue dotted lines represent the elbow of the graph. Where the two blue lines cross is the optimal number of clusters for that data. In this example, 4 clusters would be optimal.

Section 3

Results

In order to eliminate variation in the data due to sex and be able to obtain a more accurate result, analysis was done separately for men and women.

At the time of the analysis, 215 participants had been recruited between UCSF and Pennington Biomedical Research Center, in which 176 case were used while others had to be excluded due to artifacts in either the DXA or 3D optical scan. Sixteen optical scans were excluded from analysis because the participant's hair interfered with proper analysis, the participant's head moved during the scan, or the participant was wearing non-form fitting clothing. Twenty-three DXA scans were excluded from analysis due to the presence of medical or cosmetic implants, movement during the scan, body parts besides the left arm being outside of the field of view, or the presence of external objects (other than clothing) on the body. The participants' demographics are shown in Table 3.1. The participants ranged in size, shape, and demographics, which provided a large variance for the study. Some groups, though, are underrepresented in this population, such as underweight participants and Hispanics. Future recruitment is focused on recruiting individuals in the underrepresented groups to obtain a more balanced and diverse population.

Table 3.1: Demographics of participants in the study.

	Gender	
	Female	Male
Total	104	72
Age (years)		
Mean ± Std	46.71 ± 14.94	40.75 ± 16.01
Maximum	75	77
Minimum	20	19
Ethnicity (n)		
White	55	46
Black	28	12
Oriental	15	12
Hispanic	6	2
BMI (kg/m^2)		
Mean ± Std	27.31 ± 6.57	27.73 ± 5.32
Maximum	51.29	49.17
Minimum	16.52	19.34
BMI Category (n)		
	2	0
Underweight	42	25
Normal	35	28
Overweight	25	19
Obese		

3.1. Shape and Appearance Models

For the female data, it took 38 PCs to describe 95% of the DXA fat variance, 38 PCs to describe 95% of the DXA lean variance, 39 PCs to describe 95% of the DXA bone variance, and 10 PCs to describe 95% of the 3D optical variance. For the male data, it took 32 PCs to describe 95% of the DXA fat variance, 35 PCs to describe 95% of the DXA lean variance, 35 PCs to describe 95% of the DXA bone variance, and 10 PCs to describe 95% of the 3D optical variance. It took less PCs to describe the optical variance than to describe the DXA variances, because the DXA model accounts for the shape and texture of each person, while the 3D optical model only accounts for their shape. The shape models for the first 3 PCs of the 3D optical data and the appearance models for the first 3 PCs of the DXA data are shown in Figure 3.1 for the female data and in Figure 3.2 for the male data. The model on the left of each PC image is a representation of data with a coefficient of -3 standard deviations for that PC and the model on the right is a representation of data with a coefficient of +3 standard deviations for that PC.

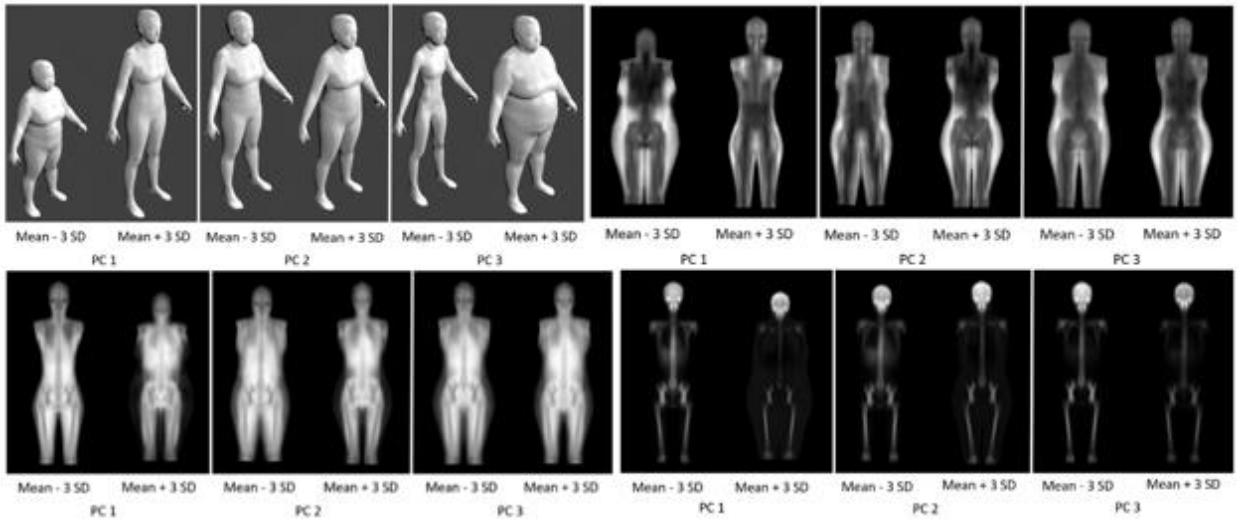


Figure 3.1. The first 3 female PCs of each image type used in this study. 3D optical PCs are at the top left, DXA fat PCs are at the top right, DXA lean PCs are at the bottom left, and DXA bone PCs are at the bottom right.

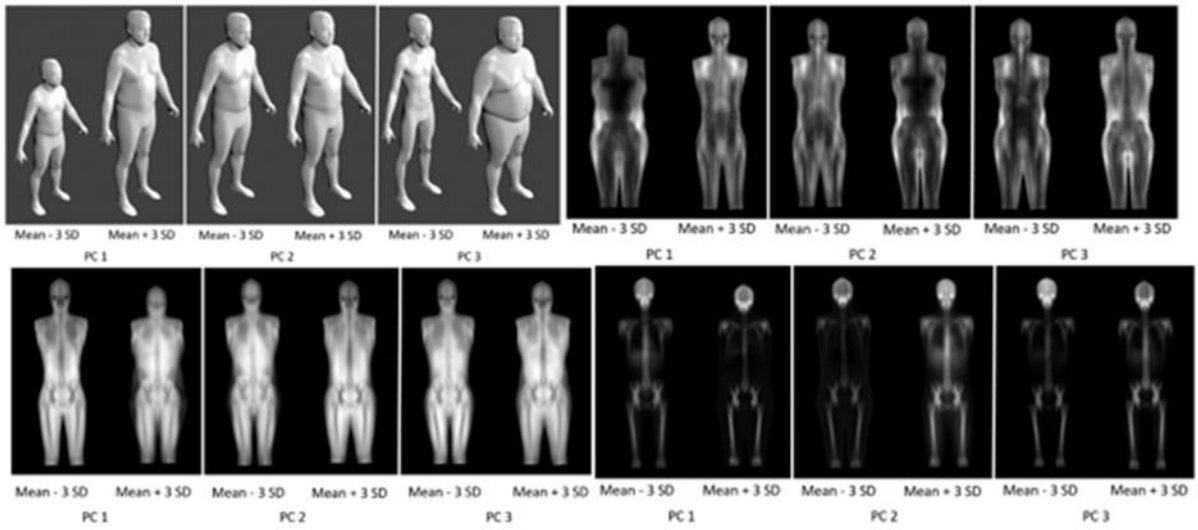


Figure 3.2. The first 3 male PCs of each image type used in this study. 3D optical PCs are at the top left, DXA fat PCs are at the top right, DXA lean PCs are at the bottom left, and DXA bone PCs are at the bottom right.

3.2. Correlations

The correlations between each PC and different anthropometry measurements, blood measurements, and DXA measurements were calculated. The PC of interest and the measurement were determined to be correlated if they had a p-value less than 0.05. Table 3.2 and Table 3.3 show all of the R-values of the PCs and measurements that were correlated. In the tables, if an R-value is in bold type, it has a p-value of

less than 0.01, and if an R-value is in bold type and shaded, it has a p-value less than 0.001. Some of the PCs, such as 3D optical female PC 3 and DXA female fat PC 1, are correlated to many nutritional health measurements, while other PCs, such as 3D optical female PC 2, are not correlated to any nutritional health measurements. Figure 3.1 shows that 3D optical female PC 3 contains significant variance in participants' weight, BMI, waist circumference, hip circumference, and upper arm circumference, which agrees with the correlations in Table 3.2. Table 3.2 also shows that 3D optical female PC 3 is also correlated with many blood and DXA measurements. 3D optical female PC 2, on the other hand, describes the variance in the way that the participants were angled in relation to the scanner. Though every effort was made to position the participants during both the 3D optical scans and the DXA scans consistently and precisely, physical differences as well as the variability in the participants' capabilities and cooperation levels could not be entirely eliminated and introduced some of the variance in the data. In DXA bone male PC 2, it can be seen that one element of variance is that participants had their legs to varying degrees and some participants had their legs together. This is also why arms were not taken considered for the DXA scans. The arm scans would not provide enough valuable information on the body composition to counteract the variance introduced.

Table 3.2. Female correlations between the first 3 PCs of the 3D optical data (top left), DXA fat data (top right), DXA lean data (bottom left), and DXA bone data (bottom right) and anthropometric, blood, and DXA measurements. Only correlations with a p-value ≤ 0.05 are shown. The R-values with a p-value ≤ 0.01 are bolded and the R-values with a p-value ≤ 0.001 are bolded and shaded.

Anthropometric Measurements	O_PC1	O_PC2	O_PC3	F_PC1	F_PC2	F_PC3
Age			0.24	**-0.24**	0.23	**-0.38**
Weight			**0.96**	**-0.70**	**-0.36**	
Height	**0.95**			0.22		0.21
BMI	**-0.27**		**0.86**	**-0.76**	**-0.38**	
Waist Circumference			**0.91**	**-0.74**	**-0.40**	**-0.29**
Hip Circumference			**0.42**	**-0.59**	**0.31**	0.24
Upper Arm Circumference			**0.44**	**-0.41**		
Blood Measurements						
Glucose	**-0.27**		**0.27**	-0.19		**-0.40**
Cholesterol						
Triglycerides	**-0.29**		**0.34**	**-0.27**		**-0.25**
High-Density Lipoproteins			**-0.34**			
Low-Density Lipoproteins						
Insulin	**-0.27**		**0.58**	**-0.50**		**-0.34**
A1C	-0.22		0.23	-0.20		-0.22
DXA Measurements						
Total Bone Mineral Density	0.22					0.22
Total Bone Mineral Content	**0.49**		**0.38**	-0.20		0.23
Total Fat			**0.85**	**-0.84**	**-0.26**	
Total Lean	0.20		**0.95**	**-0.57**	**-0.39**	
Total Percent Fat	**-0.26**		**0.73**	**-0.77**		
Total Visceral Fat Area	-0.23		**0.81**	**-0.73**	-0.23	**-0.35**
Total Subcutaneous Fat Area			0.93	**-0.84**	-0.21	-0.22

Anthropometric Measurements	L_PC1	L_PC2	L_PC3	B_PC1	B_PC2	B_PC3
Age	**0.29**	**0.28**		**0.26**	**0.30**	
Weight	**0.67**	**-0.40**		**0.68**	**-0.36**	
Height	-0.22			**-0.24**		
BMI	**0.74**	**-0.41**		**0.75**	**-0.35**	
Waist Circumference	**0.71**	**-0.42**		**0.74**	**-0.33**	
Hip Circumference	**0.61**	0.22		**0.56**	0.22	-0.21
Upper Arm Circumference	**0.43**			**0.42**		
Blood Measurements						
Glucose	0.21			0.19		
Cholesterol						
Triglycerides	**0.29**		0.22	**0.30**		**0.25**
High-Density Lipoproteins						
Low-Density Lipoproteins						
Insulin	**0.49**			**0.50**		
A1C	0.21			0.19		
DXA Measurements						
Total Bone Mineral Density		-0.19	**-0.26**	**-0.40**		
Total Bone Mineral Content		**-0.25**	**-0.25**	-0.24		-0.21
Total Fat	**0.84**	**-0.30**			**0.84**	**-0.25**
Total Lean	**0.52**	**-0.47**			**0.55**	**-0.42**
Total Percent Fat	**0.79**				**0.76**	
Total Visceral Fat Area	**0.74**	-0.22	**0.27**	0.20	**0.75**	
Total Subcutaneous Fat Area	**0.83**	-0.24			**0.83**	

Table 3.3. Male correlations between the first 3 PCs of the 3D optical data (top left), DXA fat data (top right), DXA lean data (bottom left), and DXA bone data (bottom right) and anthropometric, blood, and DXA measurements. Only correlations with a p-value ≤ 0.05 are shown. The R-values with a p-value ≤ 0.01 are bolded and the R-values with a p-value ≤ 0.001 are bolded and shaded.

Anthropometric Measurements	O_PC1	O_PC2	O_PC3	F_PC1	F_PC2	F_PC3
Age					0.27	
Weight	**0.56**		**0.78**	**-0.79**		
Height	**0.96**					
BMI	0.26		**0.89**	**-0.83**		
Waist Circumference	**0.36**		**0.80**	**-0.69**		-0.22
Hip Circumference				**-0.35**	**0.50**	
Upper Arm Circumference	**0.35**		**0.75**	**-0.65**		-0.28
Blood Measurements						
Glucose						
Cholesterol						
Triglycerides						
High-Density Lipoproteins			-0.26			
Low-Density Lipoproteins						
Insulin			**0.48**	**-0.42**		
A1C						
DXA Measurements						
Total Bone Mineral Density	**0.31**		0.24		-0.22	**0.31**
Total Bone Mineral Content	**0.62**		**0.34**			
Total Fat	**0.34**		**0.76**			-0.23
Total Lean	**0.63**		**0.65**			
Total Percent Fat			**0.58**		**0.30**	**-0.36**
Total Visceral Fat Area			**0.47**		**0.30**	**-0.34**
Total Subcutaneous Fat Area	0.25		0.80			-0.23

Anthropometric Measurements	L_PC1	L_PC2	L_PC3	B_PC1	B_PC2	B_PC3
Age		**-0.31**	**0.35**		**-0.38**	
Weight	**0.79**		0.24	**0.81**		
Height						
BMI	**0.84**		0.27	**0.84**		
Waist Circumference	**0.73**		**0.33**	**0.69**	-0.24	
Hip Circumference	**0.35**	**-0.48**		0.24	**-0.59**	
Upper Arm Circumference	**0.67**			**0.64**		
Blood Measurements						
Glucose			0.27			
Cholesterol						
Triglycerides						
High-Density Lipoproteins	-0.25					
Low-Density Lipoproteins			0.28			
Insulin	**0.43**			**0.43**		
A1C						
DXA Measurements						
Total Bone Mineral Density				0.25		
Total Bone Mineral Content	**0.33**	0.23		**0.40**		0.26
Total Fat	**0.81**		0.27	**0.81**		
Total Lean	**0.62**			**0.65**		0.23
Total Percent Fat	**0.64**		**0.30**	**0.60**		
Total Visceral Fat Area	**0.45**			**0.40**	-0.23	
Total Subcutaneous Fat Area	**0.84**		**0.31**	**0.82**		

3.3. Prediction Equations

The equations in Table 3.4 and Table 3.5 are examples of the equations created using stepwise linear regression to relate all of the 3D optical PCs, weight, height, BMI, age, and ethnicity to each of the DXA PCs. Equations were created for all the PCs that described 95% of the variance for each DXA image type, but only the first three PCs of each DXA image type are shown in Table 3.4 and Table 3.5. The R^2-value of each equation is also provided in the tables. The R^2-value describes how well the predicted PC coefficients compare to the actual PC coefficients according to the DXA model created. If the R^2-value equals 1, then the equation would perfectly predict the participants' PC coefficients.

Table 3.4. Examples of the female equations that relate DXA PCs to 3D optical PCs and other measurements. F represents fat, L represents lean, and B represents bone.

PC	Equations	R^2
F_PC1	$-10.14 - 0.03\boldsymbol{PC1_{OPT}} - 0.01\boldsymbol{PC2_{OPT}} - 0.13\boldsymbol{PC3_{OPT}} - 0.03\boldsymbol{PC6_{OPT}} + 0.03 + 0.06\boldsymbol{Height}$	0.87
F_PC2	$0.23 - 0.03\boldsymbol{PC4_{OPT}} + 0.05\boldsymbol{PC6_{OPT}} - 0.03\boldsymbol{PC7_{OPT}} - 0.06\boldsymbol{PC8_{OPT}} - 0.11\boldsymbol{PC9_{OPT}} - 0.02\boldsymbol{BMI} - 0.31\boldsymbol{Black} + 0.24\boldsymbol{White} + 0.003\boldsymbol{Age}$	0.67
F_PC3	$0.18 + 0.005\boldsymbol{PC1_{OPT}} - 0.02\boldsymbol{PC3_{OPT}} - 0.02\boldsymbol{PC4_{OPT}} - 0.02\boldsymbol{PC5_{OPT}} + 0.03\boldsymbol{PC6_{OPT}} - 0.004\boldsymbol{Age}$	0.36
L_PC1	$9.83 + 0.03\boldsymbol{PC1_{OPT}} + 0.10\boldsymbol{PC3_{OPT}} - 0.01\boldsymbol{PC4_{OPT}} + 0.01\boldsymbol{PC5_{OPT}} - 0.01\boldsymbol{BMI} - 0.06\boldsymbol{Height} - 0.11\boldsymbol{Hispanic}$	0.88
L_PC2	$0.12 + 0.02\boldsymbol{PC4_{OPT}} + 0.02\boldsymbol{PC6_{OPT}} - 0.03\boldsymbol{PC8_{OPT}} - 0.05\boldsymbol{PC9_{OPT}} - 0.02\boldsymbol{PC10_{OPT}} - 0.01\boldsymbol{BMI} + 0.14\boldsymbol{Black} + 0.12\boldsymbol{White} + 0.002\boldsymbol{Age}$	0.65
L_PC3	$-0.07 - 0.002\boldsymbol{PC1_{OPT}} - 0.02\boldsymbol{PC4_{OPT}} - 0.01\boldsymbol{PC5_{OPT}} - 0.02\boldsymbol{PC6_{OPT}} - 0.10\boldsymbol{Hispanic} + 0.002\boldsymbol{Age}$	0.30
B_PC1	$5.41 + 0.01\boldsymbol{PC1_{OPT}} + 0.004\boldsymbol{PC2_{OPT}} + 0.09\boldsymbol{PC3_{OPT}} - 0.01\boldsymbol{PC4_{OPT}} + 0.01\boldsymbol{PC5_{OPT}} + 0.02\boldsymbol{PC6_{OPT}} - 0.02\boldsymbol{PC7_{OPT}} - 0.003\boldsymbol{Weight} - 0.03\boldsymbol{Height} - 0.06\boldsymbol{Hispanic}$	0.87
B_PC2	$2.24 + 0.01\boldsymbol{PC1_{OPT}} + 0.02\boldsymbol{PC4_{OPT}} + 0.01\boldsymbol{PC6_{OPT}} - 0.01\boldsymbol{PC7_{OPT}} - 0.02\boldsymbol{PC8_{OPT}} - 0.05\boldsymbol{PC9_{OPT}} - 0.004\mathbf{Weight} - 0.01\mathbf{Height} + 0.13\mathbf{Black} + 0.11\mathbf{White} + 0.002\mathbf{Age}$	0.66
B_PC3	$-0.09 - 0.002\boldsymbol{PC1_{OPT}} - 0.03\boldsymbol{PC4_{OPT}} + 0.01\boldsymbol{PC5_{OPT}} - 0.01\boldsymbol{PC6_{OPT}} - 0.07\boldsymbol{Hispanic} + 0.002\boldsymbol{Age}$	0.29

Table 3.5. Examples of the male equations that relate DXA PCs to 3D optical PCs and other measurements. F represents fat, L represents lean, and B represents bone.

PC	Equations	R^2
F_PC1	$2.47 - 0.03\boldsymbol{PC4_{OPT}} - 0.05\boldsymbol{PC5_{OPT}} - 0.05\boldsymbol{PC8_{OPT}} - 0.05\boldsymbol{PC10_{OPT}} - 0.09\boldsymbol{BMI} - 0.33\boldsymbol{Hispanic}$	0.79
F_PC2	$0.57 - 0.07\boldsymbol{PC4_{OPT}} + 0.07\boldsymbol{PC5_{OPT}} + 0.03\boldsymbol{PC6_{OPT}} + 0.05\boldsymbol{PC8_{OPT}} + 0.06\boldsymbol{PC10_{OPT}} - 0.01\boldsymbol{Weight} - 0.27\boldsymbol{Hispanic} - 0.30\boldsymbol{Oriental}$	0.62
F_PC3	$-0.02 - 0.01\boldsymbol{PC2_{OPT}} + 0.05\boldsymbol{PC4_{OPT}} + 0.04\boldsymbol{PC5_{OPT}} - 0.04\boldsymbol{PC7_{OPT}} + 0.03\boldsymbol{PC8_{OPT}} + 0.09\boldsymbol{PC10_{OPT}}$	0.46
L_PC1	$-1.3 + 0.01\boldsymbol{PC4_{OPT}} + 0.03\boldsymbol{PC5_{OPT}} + 0.02\boldsymbol{PC8_{OPT}} + 0.03\boldsymbol{PC10_{OPT}} + 0.05\boldsymbol{BMI}$	0.79
L_PC2	$-2.28 + 0.01\boldsymbol{PC1_{OPT}} + 0.17\boldsymbol{PC2_{OPT}} + 0.17\boldsymbol{PC3_{OPT}} + 0.04\boldsymbol{PC4_{OPT}} + 0.05\boldsymbol{PC5_{OPT}} - 0.06\boldsymbol{PC6_{OPT}} - 0.03\boldsymbol{PC7_{OPT}} - 0.02\boldsymbol{PC9_{OPT}} - 0.004\boldsymbol{PC10_{OPT}} + 0.03\boldsymbol{Weight} + 0.15\boldsymbol{Hispanic} + 0.10\boldsymbol{Oriental}$	0.77
L_PC3	$-0.15 + 0.01\boldsymbol{PC2_{OPT}} + 0.01\boldsymbol{PC3_{OPT}} - 0.02\boldsymbol{PC4_{OPT}} - 0.02\boldsymbol{PC5_{OPT}} + 0.02\boldsymbol{PC7_{OPT}} - 0.03\boldsymbol{PC8_{OPT}} - 0.05\boldsymbol{PC10_{OPT}} + 0.004\boldsymbol{Age}$	0.58
B_PC1	$-1.21 + 0.02\boldsymbol{PC4_{OPT}} + 0.01\boldsymbol{PC5_{OPT}} + 0.04\boldsymbol{BMI} + 0.2\boldsymbol{Hispanic}$	0.76
B_PC2	$-4.58 - 0.01\boldsymbol{PC1_{OPT}} + 0.03\boldsymbol{PC4_{OPT}} - 0.04\boldsymbol{PC5_{OPT}} - 0.02\boldsymbol{PC8_{OPT}} - 0.04\boldsymbol{PC10_{OPT}} + 0.03\mathbf{Height} + 0.12\mathbf{Oriental} - 0.003\mathbf{Age}$	0.66
B_PC3	$-0.07 - 0.03\boldsymbol{PC4_{OPT}} - 0.02\boldsymbol{PC5_{OPT}} + 0.03\boldsymbol{PC7_{OPT}} - 0.04\boldsymbol{PC8_{OPT}} + 0.16\boldsymbol{Black} + 0.05\boldsymbol{White}$	0.53

3.4. Predicted DXA Images & RMSE Values

Figure 3.3 shows an example female's actual DXA image on the left and her predicted DXA image on the right for all three image types. Figure 3.4 shows an example male's actual DXA image on the left and his predicted DXA image on the right for all three image types. Using MATLAB, a magnitude difference image was calculated to determine the accuracy of the predicted DXA image as shown in the middle panel. As seen from the female's difference images, the texture of her predicted DXA images are slightly different from her actual DXA images while there is only a very slight difference in the size of the bodies. As seen from the male's fat difference image, his predicted fat image was slightly different in texture from his actual DXA fat image, but the prediction of his body shape was extremely accurate. His predicted lean image, on the other hand, was slightly inaccurate in predicting his body shape, but the only texture prediction that seemed to have any significant amount of error was where his bone was located. His predicted bone image inaccurately predicted the size of his head, which is not as important for body composition as other body parts, and the organ-filled areas contain most of the other differences.

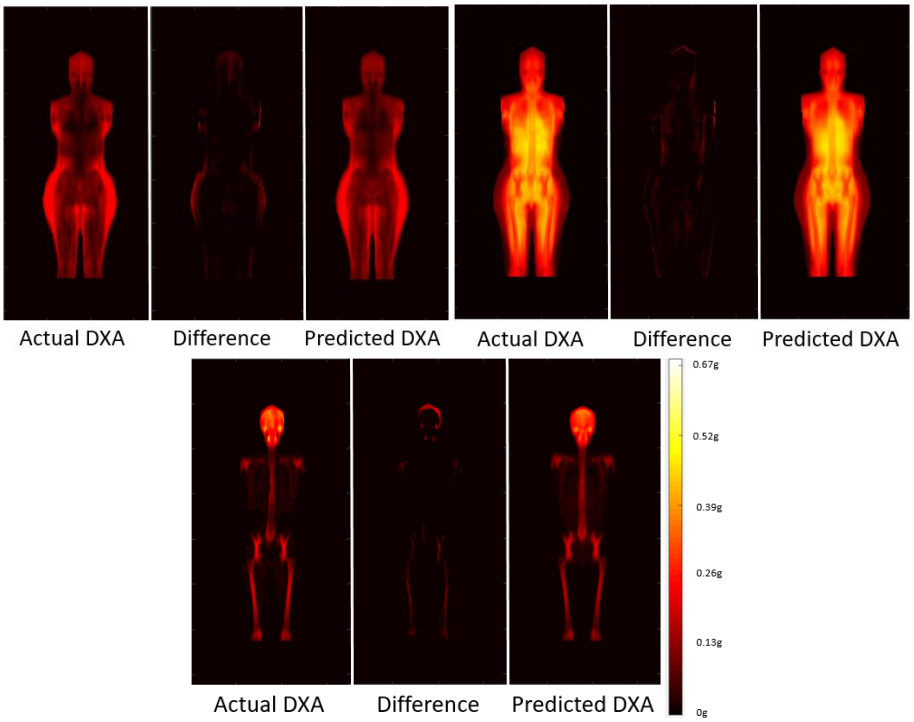


Figure 3.3. Example of a female's predicted DXA image compared to her actual DXA image for each DXA image type (Upper Left: Fat, Upper Right: Lean, Bottom: Bone). Left: Actual DXA scan Right: Predicted DXA scan Middle: Magnitude of difference between actual and predicted DXA

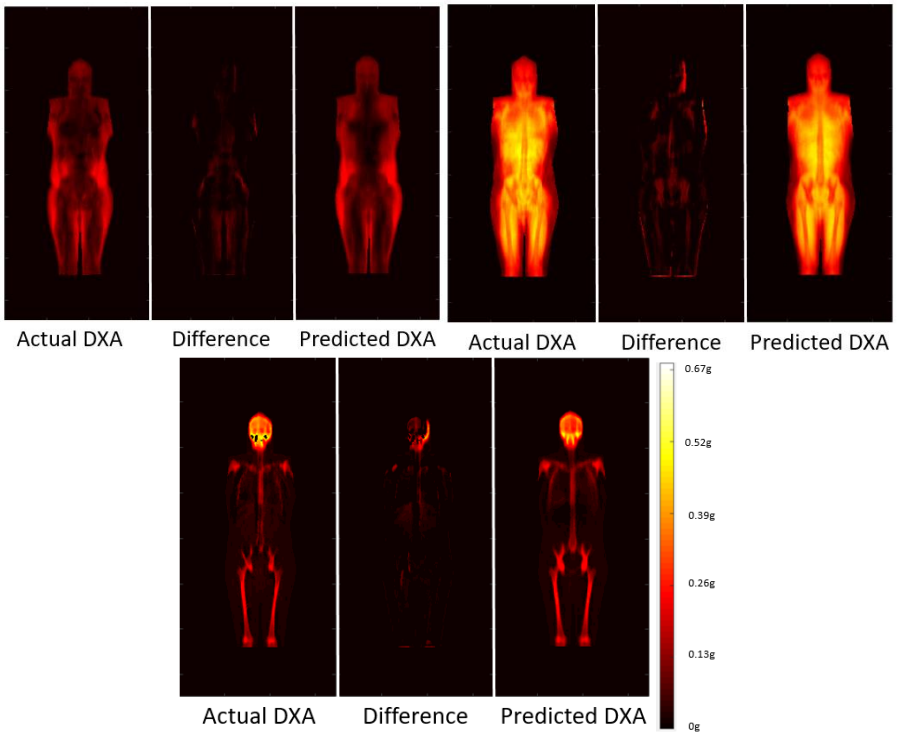


Figure 3.4. Example of a male's predicted DXA image compared to his actual DXA image for each DXA image type (Upper Left: Fat, Upper Right: Lean, Bottom: Bone). Left: Actual DXA scan Right: Predicted DXA scan Middle: Magnitude of difference between actual and predicted DXA

The RMSE of each participant was calculated to compare the predicted DXA images to their actual DXA images; subsequently, all of the participants' RMSEs were averaged to determine the RMSE of each DXA image type. These RMSE values are summarized in Table 3.6 and Table 3.7 in terms of grams of fat, grams of lean, and grams of bone.

Table 3.6. RMSE values of the females' predicted DXA images when compared to their actual DXA images.

	Fat	**Lean**	**Bone**
RMSE	0.0152 g	0.0227 g	0.0123 g

Table 3.7. RMSE values of the males' predicted DXA images when compared to their actual DXA images.

	Fat	**Lean**	**Bone**
RMSE	0.0132 g	0.0240 g	0.0177 g

3.5. Clusters

Using k-means clustering on the predicted DXA fat PCs, it was determined that there are 9 body phenotypes in the female population and 5 body phenotypes in the male population. A pairwise t-test was then performed on the averages of each cluster in order to determine unique attributes of each phenotype compared to the other phenotypes. A cluster was determined to have a unique attribute if the variable's average for that cluster had a statistically significant difference from some of the other clusters. The means of the tested variables for each female phenotype are shown in Table 3.8, and the means for each male phenotype are shown in Table 3.10. The female phenotypes and their unique attributes are shown in Figure 3.5 and the males' phenotypes and corresponding attributes are shown in Figure 3.7. The phenotype names were created from the average optical image of each cluster, which are displayed in Figure 3.6 and Figure 3.8 along with the average DXA fat images of each cluster.

Table 3.8. Table of female phenotype averages for different measurements.

	Curvy	Slender	Fit	Rectangular	Average	Love Handles	Plump	Round	Easy Chair
BMI	22.58	21.13	22.58	22.59	25.57	28.22	30	33.08	41
Age	49.73	43.43	38.5	51	43.65	37.11	55.25	51	52.89
Glucose	86.91	89.29	86.93	94.89	86.3	86.89	110.83	99	99.67
Triglycerides	57.55	72.43	71.5	82.89	70.25	97.33	96.92	87.67	134.11
High-Density Lipoproteins	70.94	69.36	70	68.79	64.07	63.66	63.17	60.72	58.49
Waist Circumference	79.42	76.99	80.94	83.47	83.69	85.79	92.03	99.23	118.93
Total Fat	19989.18	18385.23	18248.04	19504.79	24326.41	22796.52	28630.17	34607.2	48338.58
Total Lean	41991.36	42428.06	42926.31	41538.4	44868.27	44620.67	47374.5	50246.6	59712.5
Total BMD	1.1	1.09	1.1	1.04	1.12	1.1	1.13	1.03	1.12

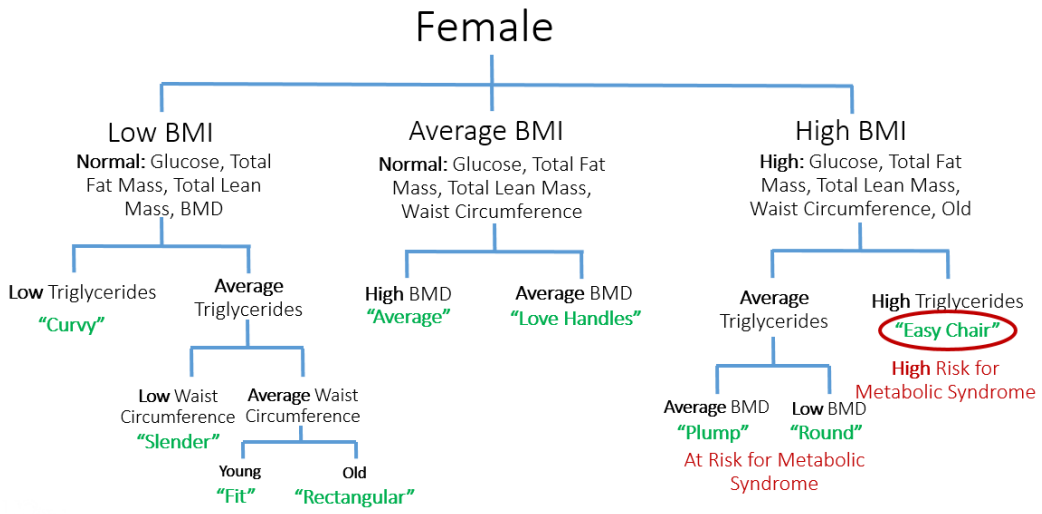


Figure 3.5. Visualization of the breakdown of the female clusters, showing the unique attributes of each cluster.

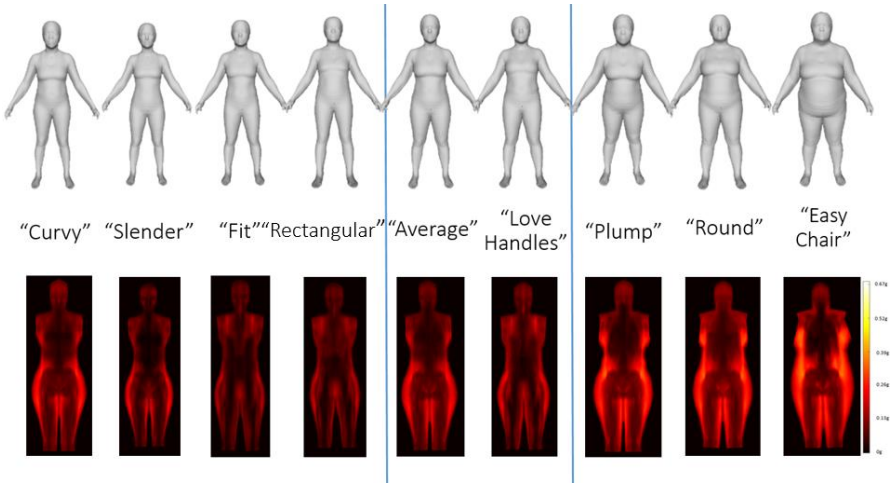


Figure 3.6. The average 3D optical and average DXA fat images for each female phenotype.

Four of the female clusters have a lower BMI than the other clusters. One of these groups had much lower triglyceride levels than the other low BMI clusters and was named "Curvy" due its small waist to hip ratio in the average image. Of the other 3 clusters with a low BMI, "Slender" had a slightly smaller waist circumference, which can also be seen from the images. Finally, "Fit" and "Rectangular" differed by age. Their average optical images agreed with this, with "Fit" looking more athletic than "Rectangular." On the other end of the spectrum, there are three female clusters that have a higher BMI than the others. "Easy Chair" is at a higher risk for having metabolic syndrome than the other two, because while all three have high glucose levels and large waist circumferences, "Easy Chair" also has high triglyceride levels. A person is determined to have metabolic syndrome if they have three or more of the attributes in Table 3.9 [13]. Although "Easy Chair's" were not quite over the limit for metabolic syndrome, they were still much higher than the other clusters and were close to the limit. None of the clusters had a high density lipoprotein count that was statistically different from the other clusters. Additionally, blood pressure measurements were not able to be analyzed because they have only been collected on half of the participants in the study.

Table 3.9. Attributes of metabolic syndrome and their corresponding measurements. If an individual has three or more of these attributes, they are diagnosed with having metabolic syndrome. Adapted from [13].

Attributes	Related Measurements
High Waist Circumference	Male: ≥102cm, Female: ≥88cm
High Triglycerides	≥ 150 mg/dL
Low HDL	Male: <40 mg/dL, Female: < 50 mg/dL
High Blood Pressure	>120/85 mmHg
High Blood Glucose	≥100 mg/dL

The main attribute that was different between "Plump" and "Round" was that "Round" had a low BMD compared to the other clusters. Although not statistically different, "Round" also had a higher BMI, waist circumference, and total fat mass than "Plump." Finally, two of the female clusters had an average BMI compared to the others. Like "Plump" and "Round," "Average" and "Love Handles" differed by BMD and although it is not statistically different, "Love Handles" has a higher BMI than "Average." "Average" was found to have healthy bones, with an average BMD higher than the other clusters.

Table 3.10. Table of male phenotype averages for different measurements.

	Slender	Well-Defined	Average	Apple-Shaped	Plump
BMI	22.8	27.08	27.15	26.94	35.99
Age	31.94	38.14	35.19	56.62	45.38
Glucose	85.13	88.38	95.75	93.46	100.92
Triglycerides	53.33	74.38	112.31	112.46	116.85
High-Density Lipoproteins	64.73	57.26	51.76	54.84	48.68
Waist Circumference	80.56	89.95	92.66	96.03	116.34
Total Fat	12841.4	16227.57	20213.76	18564.49	34085.29
Total Lean	55134.14	67965.88	66401.22	66146.38	79383.84
Total BMD	1.15	1.26	1.2	1.17	1.22

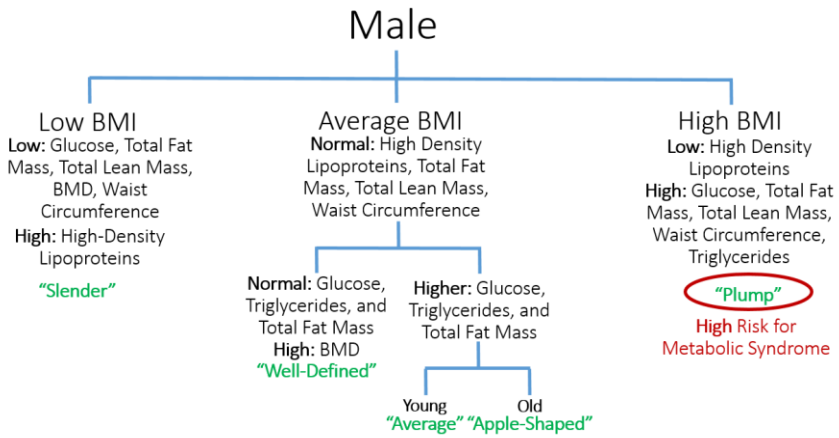


Figure 3.7. Visualization of the breakdown of the male clusters, showing the unique attributes of each cluster.

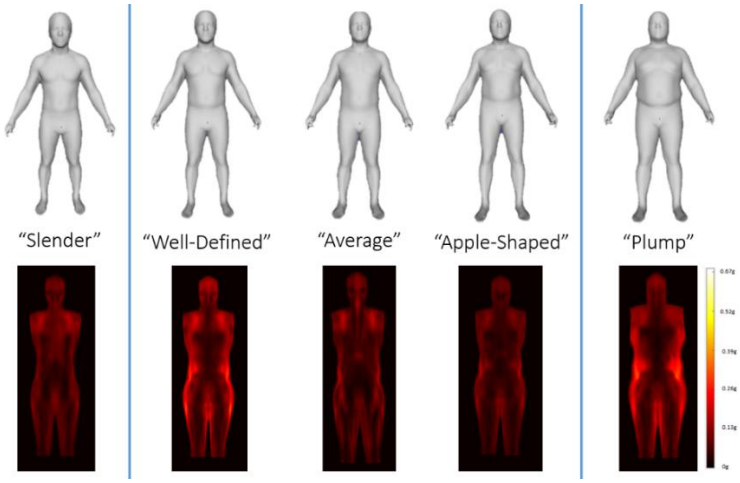


Figure 3.8. The average 3D optical and average DXA images for each male phenotype.

One male cluster had a smaller build than all of the other clusters, so it was determined to be the "Slender" phenotype. It also had some healthier blood measurements than the other groups, with a lower glucose measurement and a higher High-Density Lipoprotein count. One of the clusters had a bigger build than the other clusters, with a higher total fat mass, a higher total lean mass, and a higher BMI, and was named "Plump" because of this. This phenotype also has high glucose levels, a high waist circumference, a high triglyceride measurement, and a low High Density Lipoprotein count, making people with this phenotype at high risk for metabolic syndrome. The glucose and waist circumference averages for this phenotype were above the metabolic syndrome levels. Although the triglyceride level and high density lipoprotein averages did not quite hit the metabolic syndrome levels, they came close and were much different from the averages of the clusters that had much healthier measurements for these. Finally, there were three clusters that had average BMIs compared to the other clusters. Their other measures tended to also be average compared to the other clusters, with a lot of their measurement averages being similar to each other. Because of this, it was a bit more difficult to determine which attributes made these clusters unique. It was found that "Average" and "Apple-Shaped" had higher glucose levels, triglyceride levels, and total fat mass than "Well-Defined." "Apple-Shaped" and "Average," however, differed due to age, with "Apple–Shaped" being old and "Average" being young. From analyzing their averages, it can be seen that the "Apple-Shaped" phenotype had a slightly higher waist circumference and slightly lower BMD than the "Average" phenotype. These trends are generally seen when comparing young adults to older adults.

Section 4

Discussion

4.1. Summary

This study had two novel purposes. One was to determine if 3D optical scans can be used to accurately predict DXA images, in order to predict body composition. The second purpose was to determine what body phenotypes exist and the attributes and health risks for each phenotype. In the following subsections, I will discuss the strengths and limitations of the study, as well as future work that can be done to improve the results.

4.2. Strengths and Successes

Overall, the predicted images were accurate compared to their actual DXA images. All of the RMSE values for both males and females were less than 0.034 g, which means that the predicted DXA images met my goal of accuracy. The DXA images themselves can be very informative to an individual, telling them where they have excess fat or need to build muscle. From only a 3D Optical scan, someone can determine which areas of the body they could lose fat or gain muscle. From the cluster analysis done in this study, a person can also obtain their body phenotype and learn about any unique health attributes and health risks to which they may be prone. There were more female participants than male participants in the study, which is most likely one of the reasons why there were more clusters for the females than for the males. In addition, the female clusters seem to be more unique than the male clusters. For example, a lot of the female clusters existed in the low- and high-BMI ranges, while more of the male clusters existed in the average BMI range. This preliminary study has shown that this method will be an accurate way to measure body composition.

4.3. Limitations

Since there are only 176 participants in this study, many of its limitations are due to the limited number of participants. One deficiency is that the participant population size and representativeness were smaller and less than ideal to train the statistical models, making the models less than statistically robust. For example, there are no underweight male participants. If an underweight male was to use this model, his result would not be considered reliable because his physical characteristics are outside the range that the model was trained on. Another limitation is the lack of racial and ethnic diversity in the participant population. There are significantly more White participants. Consequently, the model would not be as reliable for a Hispanic adult as it would be for a White adult until more participants are recruited. Additionally, the required questionnaire does not ask about the participant's medical history. I might not be excluding some participants that should be excluded based on medical conditions as a result, and this may be affecting the models, making them less universally applicable. Another limitation is that this study was done on one specific DXA scanner (Hologic Horizon/A system) and one specific 3D optical scanner (Fit3D Proscanner). If the 3D optical scan was performed on a different 3D optical scanner there is no guarantee that the individual would obtain accurate results using the equipment-specific model. Finally, we did not have enough participants to fully validate our analysis. We used all participants to build the models and clusters. In summary, this analysis should be seen as a successful proof of concept more than the final work or form of the equations.

4.4. Future Work

There is much to be done to develop this novel method of body composition modeling. The starting point of future work will be continued recruitment for the Shape Up! Adults study. The goal for the study is to build the participant population to more than 700 adults, increasing the geographical areas, races, ethnicities, and body types represented, allowing a validation study to be performed. While bolstering the participant population, the background data collected should be fleshed out to include medical histories and lifestyle profiles. It would be advantageous to ensure the data collected across the program is more

consistent. Vital health statistics for each participant should be recorded in a consistent fashion. This information could be used in future clustering exercises and in evaluating participants for suitability to the study.

Because the predicted DXA images have been proven accurate, in the future ROIs can be placed on specific regions of the predicted DXA images, such as a leg or the torso, and the amount of fat, lean, or BMD can be estimated to obtain the person's estimated body composition. Additionally, while this primary study has demonstrated that predicted fat PCs can be clustered to produce additional body phenotypes, continued recruitment will ideally fill in the male phenotypes that I am deficient in and potentially result in more female phenotypes, establishing the method's robustness.

Resultant new clusters may also demonstrate attributes with more far-ranging values than those that have been determined to not be statistically different in the current clusters, such as A1C blood results, which are constantly gaining importance in the United States. The A1C blood test informs a person if they have prediabetes or diabetes. If a future cluster has a statistically higher A1C blood result than the average values across existing clusters, in addition to being able to inform individuals whether or not they are at risk for metabolic syndrome, an individual's risk of developing diabetes could be defined.

So far, clustering was also only done on predicted fat PCs. Because the clusters created by the predicted fat PCs showed that participants ended up being clustered according to their total lean mass and BMD just from their fat distribution, though, it is hypothesized that lean and bone clustering of participants would not provide an individual with any further health information. This hypothesis, however, will have to be tested in the future.

4.5. Conclusion

With this accessible and accurate method, individuals can easily obtain their body composition along with their body phenotype relating to metabolic risk groups. To my knowledge, there is no similar method that exists similar to this method that exists. This method will ultimately help increase awareness of metabolic

risks and motivate those at risk to see medical advice as early as possible.

References

1. Allen, Brett, Brian Curless, and Zoran Popović. "The space of human body shapes: reconstruction and parameterization from range scans." *ACM SIGGRAPH 2003*22.3 (2003): 587-94. Web.
2. Cootes, Tim. "Modelling and Search Software." *Tim Cootes: Software*. N.p., n.d. Web. 09 June 2017. <https://personalpages.manchester.ac.uk/staff/timothy.f.cootes/software/am_tools_doc/index.html>.
3. Cootes, Tim F., Mircea C. Ionita, Claudia Lindner, and Patrick Sauer. "Robust and Accurate Shape Model Fitting Using Random Forest Regression Voting." *Computer Vision – ECCV 2012*(2012): 278-91. Web.
4. Blake, G. M. and I. Fogelman (1997). "Technical principles of dual energy x-ray absorptiometry." Semin Nucl Med 27(3): 210-228
5. Le Marchand, Loic. "Project Information for OBESITY, BODY FAT DISTRIBUTION, AND CANCER RISK IN THE MULTIETHNIC COHORT." *National Institutes of Health RePORTER*. U.S. Department of Health and Human Services, n.d. Web. 09 June 2017. <https://projectreporter.nih.gov/project_info_details.cfm?aid=9130762&icde=34673740>.
6. Shepherd, John Alan. "Project Information for OPTICAL BODY COMPOSITION AND HEALTH ASSESSMENT." *National Institutes of Health RePORTER*. U.S. Department of Health and Human Services, n.d. Web. 09 June 2017. <https://projectreporter.nih.gov/project_info_description.cfm?aid=9082891&icde=34182270&ddparam=&ddvalue=&ddsub=&cr=3&csb=default&cs=ASC&pball=>.
7. Shepherd, John A., Steven B. Heymsfield, Shane A. Norris, Leanne M. Redman, Leigh C. Ward, and Christine Slater. "Measuring body composition in low-resource settings across the life course." *Obesity*24.5 (2016): 985-88. Web.
8. BK Ng, BJ Hinton, B Fan, AM Kanaya, and JA Shepherd “Clinical anthropometrics and body composition from 3D whole body surface scans” European Journal of Clinical Nutrition (2016), 1-6

9. Dual Energy Radiography Acquisition and Processing. (n.d.). Retrieved June 12, 2017, from http://www.upstate.edu/radiology/education/rsna/radiography/dual.php
10. *P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, G. Ranzuglia*
MeshLab: an Open-Source Mesh Processing Tool
Sixth Eurographics Italian Chapter Conference, page 129-136, 2008
11. RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA
URL http://www.rstudio.com/.
12. Tidying k-means clustering. (n.d.). Retrieved August 05, 2017, from https://cran.r-project.org/web/packages/broom/vignettes/kmeans.html
13. Wilson, J. P. (n.d.). The Search For Advanced Imaging Descriptors Of Human Body Shape And Their Association To Diabetes And Other Metabolic Disorders. University of California San Francisco.
14. Principal Component Analysis Chemometrics and Intelligent Laboratory Systems, 2 (1987) 37-52 Elsevier Science Publishers B.V., Amsterdam - Printed in The Netherlands
15. Chapman, L. (n.d.). Principle Component Analysis with Python: An Overview and Tutorial [Web log post]. Retrieved June 10, 2017, from http://blog.districtdatalabs.com/principal-component-analysis-with-python
16. Hartigan, J. A., and M. A. Wong. "Algorithm AS 136: A K-Means Clustering Algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, 1979, pp. 100–108. *JSTOR*, www.jstor.org/stable/2346830.
17. Lavrenko, V. (2014, January 19). Retrieved August 27, 2017, from https://www.youtube.com/watch?v=_aWzGGNrcic
18. McDonald, K. (2016, May 12). Structured Light 3D Scanning. Retrieved August 30, 2017, from http://www.instructables.com/id/Structured-Light-3D-Scanning/
19. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2017. Atlanta, GA: Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services; 2017.
20. Heymsfield, S. (2005). Human body composition. Champaign, IL: Human Kinetics

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Michaela K Piel

Author Signature

9/11/17

Date