Data Science and Engineering (DSE) Record, Volume 5, issue 1.

Estimate of Height, Weight and Body Mass Index Using Human Body Images

Pinyawat Rattanayanyong [1] and Karn Patanukhom [2]

[1] Master's Degree Program in Data Science, Chiang Mai University, Chiang Mai, Thailand
[2] Department of Computer Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand
pinyawat_r@cmu.ac.th

Abstract. This research primarily aims to develop a system that uses machine learning technology to predict human height, weight, and body mass index (BMI) from a single full-body image. We proposes a novel method that utilizes the PiFuHD model to transform 2D images into 3D models, along with processes for feature extraction, feature selection, noise reduction of 3D point clouds, training and testing machine learning models. Data were collected from a survey of male and female Thai volunteers aged 18 to 65, without physical disabilities, for evaluating abilities of the models to predict height, weight, and BMI. The effectiveness and accuracy of the machine learning methods were assessed using performance metrics such as mean absolute error (MAE). The results obtained from the testing set showed a MAE of 4.38 centimeters for height prediction, 8.56 kilograms for weight prediction, and 3.03 for body mass index. This research opens avenues for researchers and interested parties to utilize the developed concepts and methods in creating applications or systems capable of efficiently predicting human height, weight, and body mass index from images.

Keywords: Machine Learning, Artificial Intelligence, Human Height, Human Weight, Body Mass Index.

1 Introduction

In contemporary healthcare settings, the assessment of an individual's height and weight serves as a fundamental step in various medical evaluations and interventions. Beyond mere physical attributes, this data aids in calculating the Body Mass Index (BMI), a pivotal measure indicating the balance between body weight and height. The significance of BMI extends beyond aesthetic considerations, providing insights into an individual's body composition and their susceptibility to various non-communicable diseases, including but not limited to high blood pressure, heart disease, and coronary artery disease. However, traditional methods of obtaining height and weight measurements often entail direct contact with equipment, potentially exposing individuals to infectious diseases prevalent within healthcare environments.

Historically, the process of gathering patient data necessitated the use of conventional tools such as weighing scales and height measuring devices, each interaction carrying the risk of pathogen transmission. Patients, particularly those with compromised immune systems, face heightened vulnerability to infections transmitted through contact, ranging from conjunctivitis to sexually transmitted diseases. Moreover, the advent of respiratory contagions like influenza and the COVID-19 pandemic has underscored the imperative of minimizing physical contact in healthcare settings.

To mitigate these risks and revolutionize the approach to height and weight assessment, emerging technologies such as computer vision and artificial intelligence (AI) present promising solutions. By harnessing the power of AI and machine learning (ML) algorithms, coupled with sophisticated feature extraction techniques, researchers aim to develop systems capable of estimating an individual's weight and height from images alone. This innovative approach eliminates the need for traditional measuring devices, offering a contactless alternative that enhances safety for both patients and healthcare providers.

2 Literature Review

2.1 Body Weight Measurement Using Image Processing Based on Body Surface Area and Elliptical Tube Volume

Supranata. [1] Research has been done to find weight from images using Elliptical Tube Volume (ETV) and Body Surface Area (BSA). This study was conducted using 50 samples of objects containing 25 men and 25 women. Data is captured using a smartphone camera with a resolution of 4608×3456px. The distance between the object and the camera is 240 cm, and the camera's height is 127 cm. The results indicate that both BSA and ETV methods yield Mean Absolute Errors (MAE) of 4.48 kg and 4.29 kg, respectively.

2.2 Body Weight Analysis From Human Body Images

Jiang. [2] study focusing on the analysis of body weight and body mass index (BMI) through human body images. Utilizing the Body Contour and Skeleton Joint Detection (CSJ) method, the research aimed to predict weight and BMI changes. The CSJ method integrates conditional random fields with a recurrent neural network (CRF-RNN) for feature extraction. Experimentally, the study sourced 47,574 images from Reddit posts, representing 16,483 individuals, each with both "previous" and "current" images. Key findings include an 81.3% accuracy in weight change prediction, with a MAE of 3.5 for different BMI predictions and 3.76 for BMI prediction itself.

2.3 Body Weight Analysis From Human Body Images

Sriharsha. [3] study aimed to accurately estimate human height from single, uncalibrated images using anthropometric measurements and cumulative ratios in conjunction with a Multi-layer Perceptron (MLP) neural network architecture. A dataset of 264 images, featuring 22 males and 11 females across various age groups, was utilized. The images were captured at varying distances with different focal lengths. Researchers manually identified key points using the imtool in MATLAB. Results indicated a high correlation coefficient (R=0.964644) and low root mean square error (RMSE = 2.484 cm) for males, and a slightly lower correlation (R=0.8981) with a higher RMSE (8.1375 cm) for females.

3 Data and Methodology

3.1 Data Collection

The study involved collecting data from Thai volunteers aged 18 to 65 who were not physically disabled and could read, listen, and write in Thai. At least 150 participants consented to take part. Data collection included both photographs (front, left side, right side, and back views, with attire deemed polite and appropriate by the participants) and questionnaires. The questionnaires comprised two parts: one gathering general information such as age and gender, and another collecting weight and height measurements.

3.2 Exploratory Data Analysis

From the data collection survey, there were a total of 154 people participating in the research, consisting of 75 men and 79 women, divided into 4 sides, including the left side (TL), right side (TR), front-left side (L), front-right (R), back (BACK), and there are 3 camera angles: top (TOP), middle (MID), and bottom (BOTTOM), including wearing a hygienic mask (MASK) and not wearing it (NO_MASK) by There is an example as shown in Figure 1, and sometimes may take repeated photos at some angles. This gives us a total of 2,277 photos. Preliminary analysis of the data revealed the number of wearers. Clothes are separated by type as shown in Table 1, mean weight, age, height, and body mass index as shown in Table 2, number of images from each camera angle as shown in Table 3, and the number of images about wearing mask in Table 4.

Fig. 1. Example pictures of research participants.

Table 1. Table shows the number of types of clothing worn by the research participants.

Part	Type	Number
Upper	Shirt	124
	Coat	27
	Vest	3
Lower	Trousers	127
	Short	21
	Skirt	6

Table 2. Table shows average weight, age, height, and body mass index.

	Gender	Mean
Age (Year)	Male	41.4
	Female	46.9
	Summary	44.3
Weight (Kg.)	Male	68.1
	Female	61.0
	Summary	64.4
Height (Cm.)	Male	165.1
	Female	154.1
	Summary	159.4
BMI	Male	24.8
	Female	25.7
	Summary	25.3

Table 3. Table shows the number of images in each camera angle.

View	Number
MID	589
TOP	313
BOTTOM	309
L	300
R	299
BACK	159
TL	156
TR	152

Table 4. Table showing the number of pictures About wearing a face mask.

Wearing mask	Number
Mask	939
No mask	1338

3.3 Methodology

The data architecture for creating a machine learning model will be divided into three main parts, as illustrated in Figure 2.

- Data Preprocessing Section
- Data split for model training and model testing (Train Test Split)
- Model Development System

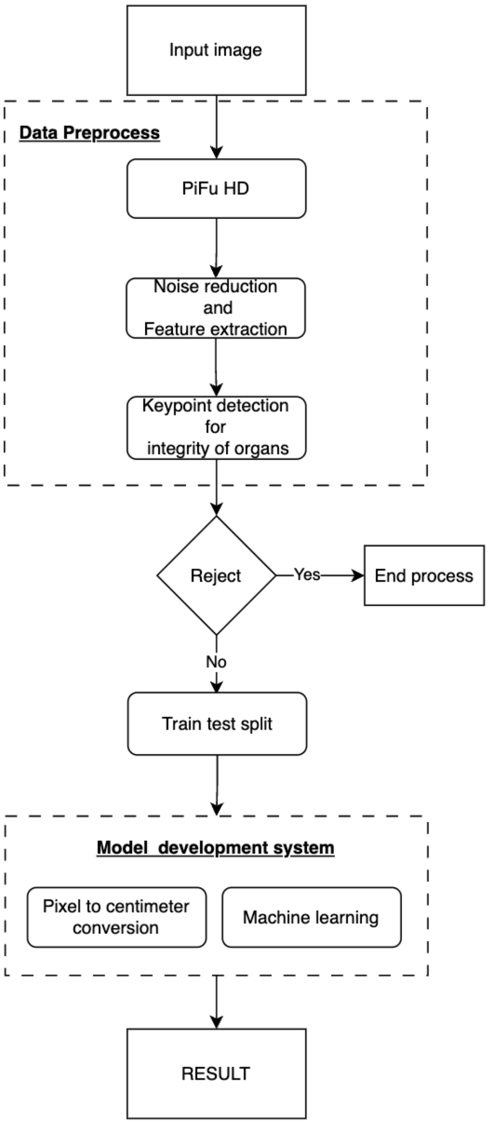

Fig. 2. Image showing data architecture for building machine learning models.

3.3.1 Data Preprocessing Section

This section consists of three parts, First PiFuHD model [4], Second image noise reduction system, and then 3D models feature extraction (Noise Reduction and Feature Extraction). Finally, there is a system for detecting important organs in 3D models (Keypoint Detection for Organ Integrity).

1) PiFuHD (Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization)

Our information is received as an image file, which currently utilizes a Model technology named PiFuHD, capable of converting 2D images into 3D models. This technology enables the extraction of features from the resulting 3D models. Therefore, the researcher has chosen to employ the PiFuHD model for converting 2D images to 3Dmodels. The input image will be a file with the extension .jpg, and importing the PiFuHD model will yield a file with the extension .obj. However, in the resulting 3D models, there may be interference such as small mesh artifacts or imperfect renderings of organs.

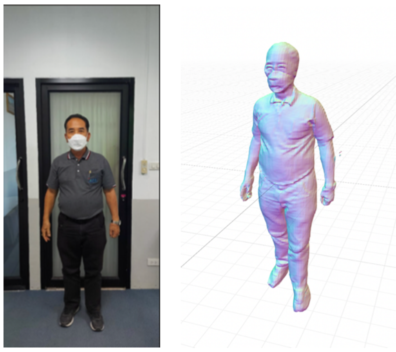

Fig. 2. Example of converting 2D images to 3D models using the PiFuHD model.

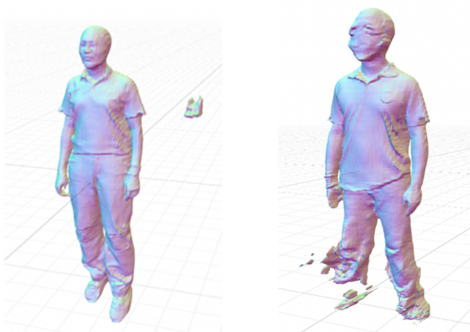

Fig. 3. Example of a 3D models that has noise and imperfect renderings of organs.

2) Noise Reduction and Feature Extraction

From the 3D model results obtained from the PiFuHD model, it can be seen that there will be some images with noise and incomplete organs. Therefore, noise must be removed, and the 3D model filtered first. This system will use a library, Trimesh, to read the .obj file and then remove noise by keeping only the largest mesh, removing small meshes, and deleting meshes that are not continuous with other meshes, resulting in an .obj file that removes noise. This process yields results in extracting additional features from 3D models that have had noise removed. The characteristics include volume (Volume), surface area (Area), height (Height), width (Width), and depth (Depth) of the 3D model.

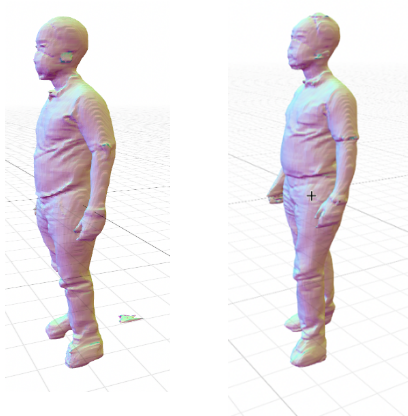

Fig. 4. Example of a 3D models before removing noise (left) and after removing noise (right).

3) Keypoint Detection for Organ Integrity

This system is responsible for filtering out 3D models with incomplete organs using the OpenPose[5] model. The OpenPose model inserts keypoints to identify the organs in 3D models generated by the PiFuHD model. Some organs may be missing, and the system checks for important organs at three levels of completeness see Table 5. If the resulting obj file is Reject not used in the next step.

Table 5. Table showing Level and Description about Keypoint Detection for Organ Integrity.

Level	Description
Perfect	OpenPose can detects all keypoints.
Acceptable	OpenPose can detects all keypoints except for the left eye, right eye, and nose.
Reject	OpenPose model fails to detect keypoints other than the left eye, right eye, and nose.

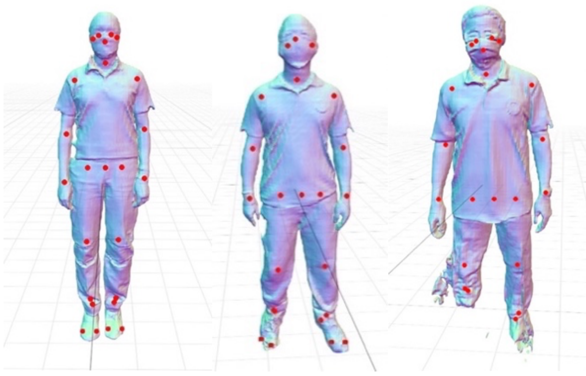

Fig. 5. Example of 3D models about Keypoint Detection for Organ Integrity Perfect(Left), Acceptable(Middle) and Reject(Right).

3.3.2 Data split for model training and model testing

This step involves dividing the data into different sets See Figure 6. The data will be split into the following ratios:

- Model Development Set: Split from 85% of all the data. The Model Development Set will undergo further division through cross-validation. It will be divided into a Model Training Set 70% of the total data, and a Model Validation Set 15% of the total data.
- Blind Testing Set: This set will constitute 15% of the total data and will be used for blind testing purposes.

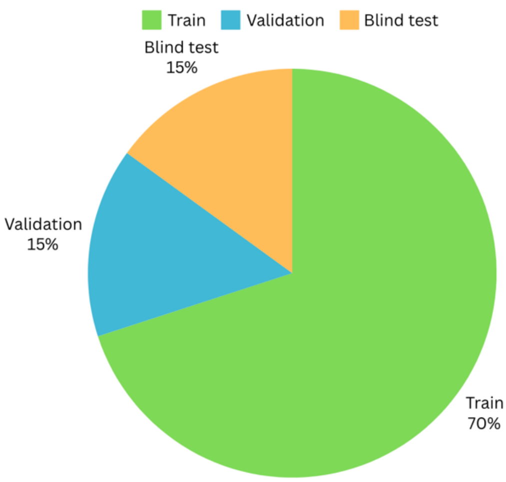

Fig. 6. The figure shows the data splitting ratio for model development.

3.3.3 Model Development System

3.3.3.1 Pixel to centimeter conversion method

Pixel to centimeter conversion method converts pixels into centimeters. This is a method that uses OpenPose to find key points for calculation. Find the height of a human in a picture. And calculate the height together with the volume obtained from the feature extraction process to find the weight and use the results of both height and weight to calculate the body mass index.

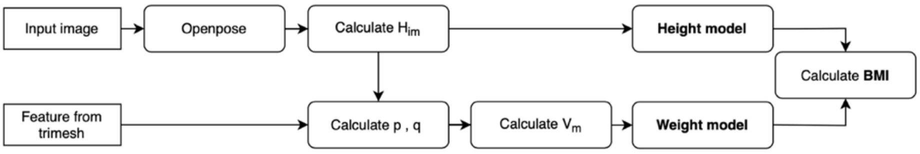

Fig. 7. Image showing the architecture of Pixel to centimeter conversion method.

To calculate the figure's height, the distance from the nose to the left heel and the distance from the nose to the right heel will be used according to equation (7), then processed and trained in the linear regression model to obtain the coefficients according to the equation. The method (8) in predicting the actual height of a person ($H_m^{(i)}$) then finding the average coefficient for comparing the image height with the actual height (p) and finding the ratio between the actual height. with the 3D models ($q^{(i)}$) according to equations (9) and (10), then calculate,The volume value of the human image ($V_m^{(i)}$) as in Equation (11). Then use the found volume value to teach it in the linear regression model to get the coefficient value as in Equation (12) to predict the weight as in equation (13) when we get the weight and height So it can be calculated body mass index.

$$H_{im}^{(i)} = \frac{D_L + D_R}{2} \quad .... (7)$$

$$H_m^{(i)} = \alpha_h H_{im}^{(i)} + \beta_h \quad .... (8)$$

$$p = \frac{1}{N}\sum_{i=1} \frac{H_m^{(i)}}{H_{im}^{(i)}} \quad .... (9)$$

$$q^{(i)} = \frac{H_{im}^{(i)}}{H_{3D}^{(i)}} \quad .... (10)$$

$$V_m^{(i)} = p^3 {q^{(i)}}^3 V_{3D}^{(i)} \quad .... (11)$$

$$W^{(i)} = \alpha_w V_m^{(i)} + \beta_w \quad .... (12)$$

$$BMI = \frac{W^{(i)}}{{H_m^{(i)}}^2} \quad .... (13)$$

$H_{im}^{(i)}$	Human height in pictures (pixels)
D_L	Distance from nose to left heel (pixel)
D_R	Distance from nose to right heel (pixel)
$H_m^{(i)}$	Human height (cm.)
α_h	Regression coefficient of the height prediction model.
β_h	Constant of the regression equation. In height prediction model.
p	Average coefficient for comparing image height with actual height.
N	Total number of data
$q^{(i)}$	Ratio between Real pictures and 3D models
$H_{3D}^{(i)}$	Human height from 3D models
$V_m^{(i)}$	Volume of human image
$V_{3D}^{(i)}$	Volume of human image from 3D models
$W^{(i)}$	Actual weight (kg.)
α_w	Regression coefficient of the weight prediction model.
β_w	Constant of the regression equation. In weight prediction model.
BMI	Body mass index

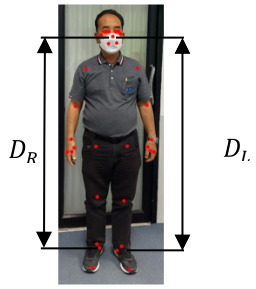

Fig. 8. Image showing distance for finding height in Pixel to centimeter conversion method.

3.3.3.2 Machine Learning

This method consists of 4 steps to find the best model and feature can predict weight, height, and body mass index, as shown in Figure 9.

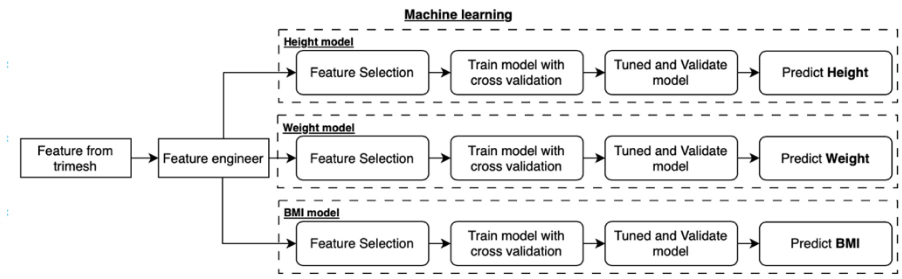

Fig. 9. Image showing the architecture of Machine learning methods.

1) Feature Engineer

This Process involves processing the attributes derived 3D model obtained during feature extraction. Initially, exponentiation is applied to features including volume, surface area, height, width, and depth. Subsequently, these transformed features are utilized to compute cumulative ratios amongst themselves. This process expands the initial set of 5 features to a total of 90, resulting in a comprehensive feature set. Furthermore, these features are integrated with additional demographic variables such as Gender and Age, culminating in a final set of 92 features.

2) Feature Selection

In Feature Selection process of choosing relevant features to improve the accuracy of a model and prevent issues such as overfitting and underfitting. There are three main types of feature selection methods [6-8] include Filter Method [9], Wrapper Method [10], and Embedded Method. In this process, the Univariate Filter Method is employed as part of the Filter Method, the Forward Selection Method is utilized as part of the Wrapper Method, and the Feature Importance Method is integrated into the Embedded Method.In the Forward Selection Method and Feature Importance Method, a variety of models are selected for feature selection, including DecisionTree (DT), Random Forest (RF), Extreme Gradient Boosting (XGB), Linear Regression (LR), Lasso Regression (LS), Ridge Regression (RD), Elastic Net Regression (EN), totaling 7 models. Feature sets are selected at intervals of 10%, ranging from 10% to 90% of the total number of features.

3) Train Model with Cross Validation

Training models with cross-validation is a method of dividing data into randomly sampled groups, based on a specified number, N. When the value N is entered, the data is divided into N subsets, with one subset retained for validation purposes. The remaining data is used to train the model, and this process iterates until all subsets have been used for validation. For instance, if N equals 5, the data is divided into 5 groups, and testing is conducted over 5 rounds. In each round, one group is used for testing while the remaining groups are used for training. This process continues until all subsets have been utilized for validation, as per the predetermined number of iterations set for this research. In this study, the researcher has set the value of K in Cross Validation to 10 groups. The models employed for training include Decision Tree, Random Forest, Extra Tree, Gradient Boosting, Extreme Gradient Boosting,Linear Regressio, and Support Vector Machine, totaling 7 models.

4) Tuned and Validation model

Once the optimal model is achieved, refinement of its hyperparameters ensues through Randomized Search Cross Validation [11], enhancing its predictive capabilities. The ultimate validation of the model's efficacy rests upon blind test datasets, yielding optimal performance. Evaluation metrics such as MAE, RMSE, and coefficient of multiple determination (R^2). To measure the results of the model.

4 Result

After compiling results from each method, including Pixel to centimeter conversion and machine learning, the machine learning approach was selected through Univariate Filter Method, Forward Selection Method, and Feature Importance Method. This resulted in a summary table of outcomes for each method, as shown in Table 6, regarding model development. Upon comparing Pixel to centimeter conversion methods and machine learning methods, it was found that the machine learning approach yielded superior results. In the Pixel to centimeter conversion method, predicting height resulted in the lowest average absolute deviation of 6.75 centimeters in the training dataset and 6.37 centimeters in the blind test dataset. Conversely, the machine learning approach yielded deviations of 1.29 and 2.80 centimeters, respectively. Regarding weight, Pixel to centimeter conversion method achieved deviations of 10.45 kilograms in the training dataset and 9.70 kilograms in the test dataset, while the machine learning approach achieved deviations of 1.48 and 6.17 kilograms, respectively. Finally, in terms of body mass index, Pixel to centimeter conversion method achieved deviations of 3.10 in the training dataset and 2.87 in the test dataset, while the machine learning approach achieved deviations of 0.48 and 0.93, respectively. Table 6 illustrates that the machine learning approach outperforms pixel to centimeter conversion method by approximately 2 to 3 times, with the best-performing model for predicting height being the Random Forest, and for predicting weight and body mass index being the Extreme Gradient Boosting, utilizing a model for feature selection is Random Forest model.

Table 6. Table summarizing the experimental results for predicting height, weight, and body mass index for each method.

					Model Development Set			Blind Testing Set		
Target	**Method**	**Estimator**	**Model**	**Feature number**	**MAE**	**RMSE**	**R^2**	**MAE**	**RMSE**	**R^2**
Height	Method 1*	-	-	-	6.748	8.187	0.168	6.365	8.057	0.119
	Method 2*	-	RF	9	1.362	2.250	0.941	3.084	4.829	0.697
	Method 3*	**RF**	**RF**	**9**	**1.285**	**2.248**	**0.942**	**2.799**	**4.381**	**0.751**
	Metohd 4*	RF	RF	9	1.398	2.330	0.937	2.808	4.446	0.744
Weight	Method 1*	-	-	-	10.446	13.857	0.178	9.69	12.235	0.217
	Method 2*	-	RF	82	3.548	5.667	0.864	7.563	10.915	0.365
	Method 3*	**RF**	**XGB**	**9**	**1.479**	**4.573**	**0.911**	**6.168**	**8.557**	**0.610**
	Metohd 4*	EN	RF	9	3.232	5.344	0.879	6.988	10.077	0.459
BMI	Method 1*	-	-	-	3.098	4.127	-0.878	2.867	3.67	-0.435
	Method 2*	-	RF	82	1.168	1.793	0.877	2.474	3.301	0.453
	Method 3*	DT	RF	9	1.150	1.800	0.876	2.192	3.065	0.528
	Metohd 4*	**RF**	**XGB**	**46**	**0.475**	**1.354**	**0.930**	**2.108**	**3.032**	**0.538**

*Method 1: Pixel to centimeter conversion

*Method 2: Machine Learning Method with Univartiate Filter Method

*Method 3: Machine Learning Method with Forward Selection Method

*Method 4: Machine Learning Method with Feature Importance Method

5 Conclusion and Future Work

From the development of machine learning models to predict human height, weight, and body mass index using images, this article discusses summary, limitations, problems, and various obstacles encountered during development, along with suggestions and feedback for further improvement and enhancement.

5.1 Summary

In the process of converting 2D images to 3D models using the PiFuHD model, challenges arose particularly in cases where organs were absent, attributed to factors such as camera angle and background variability. Notably, neither the presence of a mask nor gender exhibited significant impact on organ loss. The influence of clothing on missing organs remains inconclusive. In the realm of predictive modeling for height, weight, and BMI, machine learning techniques have demonstrated superiority over traditional Pixel to centimeter conversion methods. Among these techniques, a random forest model yielded the most accurate predictions for height (MAE = 2.80 cm), while Extreme Gradient Boosting excelled for weight and BMI, with MAE values of 6.17 kg and 0.93, respectively. Incorporating additional variables such as gender, mask usage, and attire significantly enhanced prediction accuracy, particularly notable among female subjects. Notably, mask wearing did not impact forecast outcomes, while the shooting angle emerged as a critical factor directly influencing model accuracy.

5.2 Limitations

The system developed is a prototype that necessitates further development, particularly in components such as OpenPose, which is no longer actively developed, and PiFuhd, which occasionally produces inaccuracies in generating 3D models that cannot be rectified directly within the model. Additionally, the limited amount of data collected for surveying may lead to potential overfitting issues. Furthermore, the selected feature set may not represent the optimal features, as the selection process is based on a predetermined percentage rather than the best possible features. Lastly, the complexity of machine learning models poses challenges in providing detailed explanations about their decision-making processes.

5.3 Challenges and Obstacles

The ability to collect data is hindered by several factors. Firstly, obtaining permission to use healthcare facilities for data collection poses a challenge. Consequently, there is a lower volume of data collected from adolescent participants compared to other age groups due to their relatively fewer health issues and the necessity of using bulky and immobile equipment at these facilities. Moreover, certain angles for photographing subjects are obstructed by background distractions, rendering some images unusable. Additionally, the reliance on medical personnel to interpret data values, with only one staff member available for this task, contributes to delays in data collection, often resulting in incomplete daily targets.

Furthermore, participants' attire presents a challenge as verbal persuasion is required for data collection, impeding the control over participants' clothing choices. Explaining the necessity of collecting personal data to participants is also difficult, as some individuals express discomfort, leading to frequent time wastage. Additionally, participants' inability to stand at designated points due to communication barriers and spatial constraints may introduce errors during image processing.

The training of models necessitates substantial resources, including processing systems, tracking systems, and data storage space, thereby impeding development progress. Moreover, the prolonged process of obtaining research permits contributes to delays in research activities.

5.4 Future Work

In the future, consider adopting additional modeling approaches such as artificial neural networks and refining feature selection criteria. This could include specifying additional parameters for camera height and enhancing clothing customization controls.

References

1. Supranata, Tan & Davin, Poh & Jeremy, Darren & Pratiwi, Anggy & Wulandari, Meirista. (2018). Body Weight Measurement Using Image Processing Based on Body Surface Area and Elliptical Tube Volume. 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE). Bali, Indonesia.
2. M. Jiang and G. Guo (2019). Body Weight Analysis from Human Body Images. *IEEE Transactions on Information Forensics and Security, 14*(10), 2676-2688.
3. K. V. Sriharsha and P. J. A. (2019). Alphonse. Anthropometric based Real Height Estimation using Multi layer peceptron ANN architecture in surveillance areas. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). Kanpur, India.
4. Shunsuke Saito (2020). *PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization*. University of Southern California.
5. Z. Cao, G. Hidalgo, T. Simon, S. -E. Wei and Y. Sheikh. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(1), 172-186.
6. A.Ann Romalt. (2020). An Aanalysis on Feature Selection Methods, Clustering and Classification used in Heart disease Prediction – A Machine Learning Approach. Multidisciplinary review journal. *Journal of Critical review, 7*(6), ISSN-2394-5125.
7. Jianyu Miao. (2016). A Survey on Feature Selection.. Procedia Computer Science. DOI: 10.1016/J.PROCS.2016.07.111.
8. Ghojogh, Benyamin & Samad, Maria & Mashhadi, Sayema & Kapoor, Tania & Ali, Wahab & Karray, Fakhri & Crowley, Mark. (2019, May 9). Feature Selection and Feature Extraction in Pattern Analysis.
9. Supongmen Walling, Sibesh Lodh. (2023). Performance Evaluation of Supervised Machine Learning Based Intrusion Detection with Univariate Feature Selection on NSL KDD Dataset.
10. Ferri, F.J. (1994). Comparative study of techniques for large-scale feature selection. *Machine Intelligence and Pattern Recognition, 16*, 403-413.
11. Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research, 13*, 281–305.