Multi-View Body Image-Based Prediction of Body Mass Index and Various Body Part Sizes

Seunghyun Kim[1], Kunyoung Lee[2], Eui Chul Lee[3,*]
[1]Department of AI & Informatics, Graduate School, Sangmyung University
[2]Department of Computer Science, Graduate School, Sangmyung University
[3]Department of Human-Centered Artificial Intelligence, Sangmyung University
seunghyunk34@gmail.com   201933048@sangmyung.kr   eclee@smu.ac.kr

Abstract

This paper proposes a novel model for predicting body mass index and various body part sizes using front, side, and back body images. The model is trained on a large dataset of labeled images. The results show that the model can accurately predict body mass index and various body part sizes such as chest, waist, hip, thigh, forearm, and shoulder width. One significant advantage of the proposed model is that it can use multiple views of the body to achieve more accurate predictions, overcoming the limitations of models that only used a single image. The model also does not require complex pre-processing or feature extraction, making it straightforward to apply in practice. We also explore the impact of different environmental factors, such as clothing and posture, on the model's performance. The findings show that the model is relatively insensitive to posture but is more sensitive to clothing, emphasizing the importance of controlling for clothing when using this model. Overall, the proposed model represents a step forward in predicting body mass index and various body part sizes from images. The model's accuracy, convenience, and ability to use multiple views of the body make it a promising tool for a wide range of applications. The proposed method is expected to be utilized as a parameter for accurate sensing of various vision-based non-contact biomarkers, in addition to body mass index inference.

1. Introduction

Body Mass Index (BMI) is a widely used index for measuring obesity and is calculated by dividing weight by the square of height [1]. It is used to classify people as underweight, normal weight, over-weight, or obese. Overweight and obesity are recognized as major public health problems [2,3]. However, studies have shown that BMI derived from self-reported height and weight tends to be underestimated, particularly among overweight and obese individuals [4]. This poses a challenge for effective obesity prevention, and as such, there is a need for more reliable BMI indices.

In our society, there are various types of data that record our images or videos. This data includes members of the community, and if the obesity level of the entire community can be inferred using these images, the health level can be inferred without additional cost or effort. Additionally, with the increasing interest in health, various applications related to healthcare are emerging. For example, you can take pictures of individuals through your smartphone, but you cannot weigh them. Therefore, there is a need for research that can infer people's body index based on easy-to-obtain data.

To address this challenge, researchers have turned to predicting BMI or body index from images. Studies on predicting BMI from images can be divided into two types depending on the type of data used: face images and body images. Studies that use facial images typically use publicly available mug shot data from Illinois prisons [5] or celebrity data [6]. These studies are divided into two branches: one that extracts various features from the face and the other that uses images as input. Studies that extract various features from the face usually derive feature points such as the width, height, and proportion of the face from landmarks. On the other hand, studies that use body images usually construct their own data or acquire it through internet crawling from social network services. These studies also extract various features, such as the waist-to-hip ratio and the ratio of height to leg length.

However, there is a disadvantage when using body images in that it is difficult to compare the performance of accurate algorithms because there is no shared data for each study. To address this, this study uses train and test data provided by open data to provide a starting point for more objective comparison of algorithm performance in future research in the same field.

This paper begins by discussing the necessity and progressiveness of research related to predicting BMI from images. The following section provides an overview of the study, including the data used and the applied method. Results derived from the method are then presented, and the subsequent section analyzes and discusses these results.

Finally, the thesis concludes with a summary of the findings. It is hoped that this research will contribute to the development of effective tools for deriving more reliable BMI indices and lead to better obesity prevention efforts.

2. Related Works

Coetzee et al. wanted to know the cues used to judge weight from a face and found that facial features such as width-to-height ratio, perimeter-to-area ratio, and cheek-to-jaw-width ratio were significantly related to BMI [7].

Wen et al. calculated seven facial features, such as cheekbone to jaw width (CJWR), width to upper facial height ratio (WHR), perimeter to area ratio (PAR), eye size (ES), lower face to face height ratio (LF/FH), face width to lower face height ratio (FW/LFH), and mean of eyebrow height (MEH), and predicted BMI by conduction learning in a statistical method [8]. The active shape model (ASM) was applied to the face image to extract the reference points, and it was confirmed that the 7 derived facial features were related to BMI.

In the previous work above, the BMI prediction model performance was uncertain in the case of noisy SNS photos because the dataset consisted of only front photos with a clean background, but Kocabey et al. predicted BMI from SNS images such as profile photos. For performance evaluation, an experiment was performed to compare human performance, and a pair of profile pictures were given to evaluate which picture was more overweight [9]. The more overweight people, the more similar the system and human performance proposed in this paper.

The preliminary studies outlined above share the common approach of utilizing facial photos to predict BMI. While there are benefits to using facial photos, predicting BMI based solely on the face may result in a lack of information, as BMI is an index that directly relates to body mass. Moreover, face images can cause privacy issues. Therefore, this study takes a different direction by predicting various body-driven part sizes, including BMI, using body images instead of facial images.

Zhi Jin at al. proposes a method for predicting a person's BMI from a 2D image of the body [10]. They have used an end-to-end Convolutional Neural Network with attention guidance. The dataset was collected from Reddit website and 4190 images were collected (among them 1478 are males and 2721 are females). Only front images were used and due to the characteristics of the data, they did not share the same angle nor pose. They achieved 3.85 MAE (mean average error) in average. We hypothesized that performance could be improved by balancing the data and increasing the size of the dataset.

A. Pantanowitz at al. proposes a method for estimating a person's BMI from photographs using deep CNNs that consider various features of the person's appearance [11]. The authors evaluated the performance of their model on a test set of photographs and compared it with several baseline models, outperforming them in terms of accuracy and mean absolute error. The study utilized data from 161 individuals, 79 of whom were female, and used only a single frontal photograph for both training and testing. The photograph was segmented to create a black background with a white recreation of the human form, without any additional feature extraction. This approach offers advantages in various applications such as health monitoring and personalized nutrition and fitness programs. However, the segmentation process has the drawback of potentially deleting data regarding changes in volume due to clothing or high BMI index. To overcome this limitation, our data was structured to preserve the color space. Moreover, similar to the previous study, we believed that incorporating data from different angles would also contribute to enhancing the performance of the model.

3. Methods

3.1. Dataset

The dataset used in this research was obtained from AI Hub and was constructed with the support of the Korea Intelligent Information Society Promotion Agency with funds from the Ministry of Science and ICT. This dataset comprises of body images of 500 Korean males and 500 Korean females, covering various age groups from 20s to 60s and older. For each person, a total of 1,152 images were acquired, including 6 postures, 6 outfits, and 32 different angles. Additionally, various body measurements were labeled for each person, such as height and weight, chest, waist, hip, thigh, forearm circumference, and shoulder width. The dataset used in this study is particularly valuable because it provides a comprehensive and diverse set of body images that can be used for various research purposes related to body measurements, such as obesity assessment, anthropometry, and body shape analysis. The availability of such a large and varied dataset allows for more accurate and precise analysis and can contribute to the development of new technologies and algorithms in the field of computer vision. Furthermore, this dataset can be particularly useful for developing and testing new body measurement technologies and applications, such as virtual try-on systems, body scanning, and fitting algorithms. The labeled body measurements for each person ID in the dataset make it possible to train and evaluate these technologies accurately, providing a significant advantage over datasets without such annotations. Furthermore, as the data used in this study is open data, it can be utilized by other researchers upon request.

Among the dataset, three poses and five outfits were used for training and testing. The poses are composed with as following:

At attention posture (action #1): Standing upright with feet together, arms straight and close to the body, and eyes looking straight ahead.

A-shaped arm posture (action #2): Standing with feet together, and with arms raised and bent at the elbows to form the shape of the letter "A".

T-shaped posture (action #3): Standing with feet together and arms stretched out to the sides to form the shape of the letter "T".
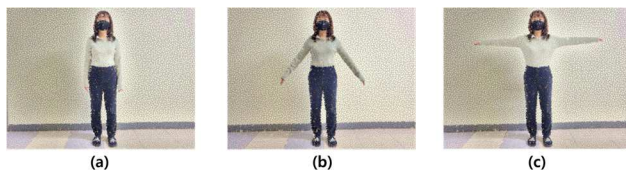

Figure 1: Conceptual images of actions. (a) Action #1. (b) Action #2. (c) Action #3.

Five different outfits are consisted with as following:

Measuring suit: a form-fitting garment such as a swimsuit that is worn during body measurements to more accurately capture the shape of the body.

Spring and autumn clothes: These refer to clothing items that are suitable for wearing during the transitional seasons of spring and autumn, when temperatures are moderate and weather conditions can vary.

Summer clothes: These refer to clothing items that are suitable for wearing during the summer season, when temperatures are typically high and weather conditions are sunny and humid.

In the case of spring and autumn clothes, each person wore two different types of clothes, so a total of 5 outfits were used. The reason for using the 15 cases mentioned above is that the postures that can better see the human form are selected first, and in the case of clothing, a more generous range is specified so that the model can be applied in various situations.

The data was then cropped based on the body landmarks, with the Y value representing the vertical axis from head to toe and the X value representing the horizontal axis from both fingertips. This approach was taken to minimize background noise and reduce the amount of unnecessary
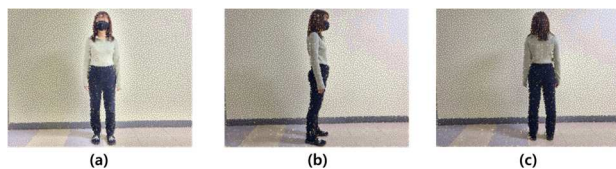

Figure 2: Conceptual images used for estimating BMI and body part sizes. (a) Frontal image. (b) Side image. (c) Rear image.

data. We opted for cropping instead of segmentation, as segmentation can introduce additional errors and create dependencies on the segmentation model. Moreover, cropping also offers the advantage of faster computation speed by reducing the number of additional calculations. Figure 2 (a), (b), and (c) illustrate examples of the original images used, while figure 3 (a), (b), and (c) show examples of the images used for training after they were cropped and resized to 224 by 224.
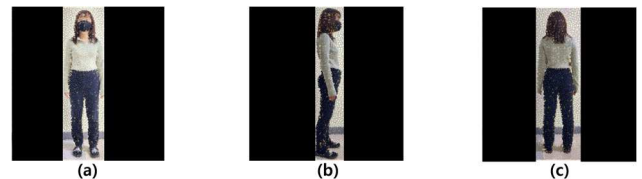

Figure 3: Conceptual pre-processed images. (a) Frontal image. (b) Side image. (c) Rear image.

3.2. Models

In this study, we propose a method for predicting various body part sizes through body pictures taken from different angles. Our model supports four methods of outputting various index outputs, including processing multi-image input and outputs a single output, processing multi-image and referenceable index input and outputs a single output, processing multi-image input and outputs various index outputs, and processing multi-image and reference index input to output various index outputs.

To predict BMI, we developed two models: the first model predicts BMI by inputting image data from three different angles, while the second model inputs both image data from three angles and float data of the object's height. The images used for each model are shown in figure 4 and figure 5, respectively.
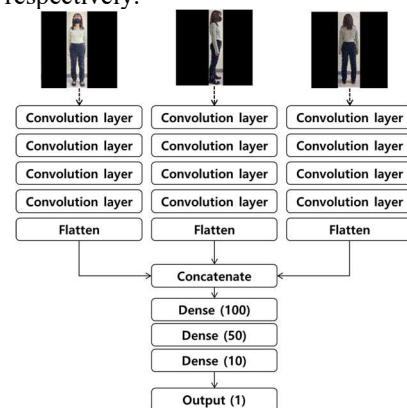

Figure 4: Architecture of images-to-BMI model. Three different images with different angles were used as inputs and a single output which is the BMI.
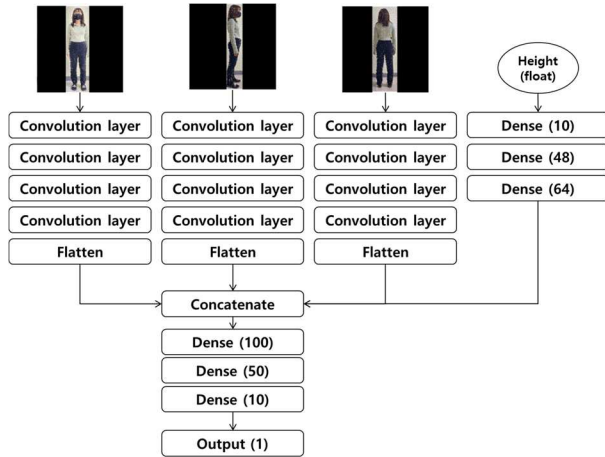
Figure 5: Architecture of images&height-to-BMI model. Three different images with different angles and height value were used as inputs and a single output which is the BMI.

We also developed a model that outputs various body indices, targeting a total of six outputs: chest, waist, hips, thighs, forearm circumference, and shoulder width. The five circumferences except shoulder width are highly related to BMI and most related to clothing, making them valuable outputs for practical applications. Instead of constructing each index into a separate model, we opted to share the weights of the basic CNN model equally and learn specific weights for each feature in additional dense layers. This approach was chosen because human body indices are correlated with each other.

Similar to the first BMI model, we developed a third model using body images measured from three angles and height values as inputs. The second case, BMI value was used as an input along the images and height that were used as well. In this case, images-to-BMI model and images&height-to-BMI models' outputs can be used as an input.
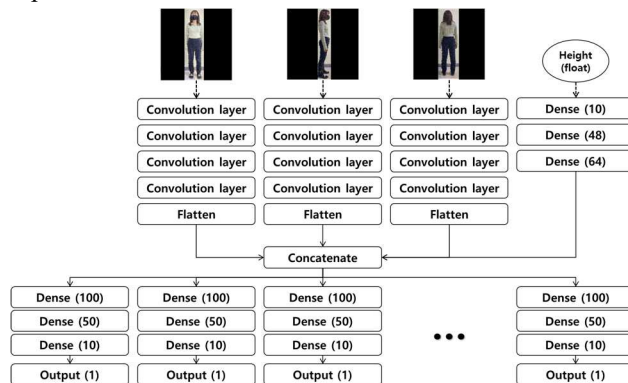

Figure 6: Architecture of images&height-to-Sizes model. Three different images with different angles and height value were used as inputs and seven different outputs were driven from the model.
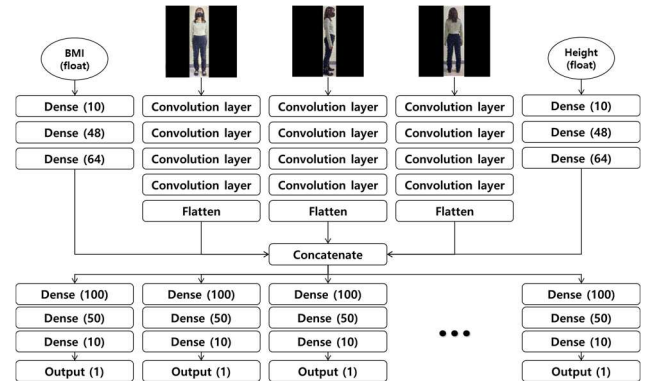

Figure 7: Architecture of images&height&BMI-to-Sizes model. Three different images with different angles, height value and BMI value were used as inputs and seven different outputs were driven from the model.

To summarize, our study presents a comprehensive method for predicting various body part sizes using body pictures taken from different angles. Our approach can be used in a range of applications, including clothing size recommendation, fitness monitoring, and health analysis.

4. Results

The test was conducted using person IDs that were not used in the training. Similar to the training, the users were asked to take three actions (attention postures), and the results were obtained as the predictions of two models for BMI and 14 predictions using two models for various body features.

Firstly, for the BMI prediction using images-to-BMI model, the MAE (Mean Average Error) was found to be 3.1199. Although this error may seem high, it was observed that the accuracy varied significantly depending on the clothes worn by the subjects, as shown in table 1. The results in table 1 indicate that the best performance was achieved when the subjects were wearing a costume during the measurement. This suggests that clothing can significantly affect the accuracy of the model's predictions. Further analysis of these results can help improve the model's performance by taking clothing into account during the training and testing phases.

The table has columns representing each action and rows representing each costume. Action #1 represents an attention posture, #2 represents an A-shaped arm posture, and #3 represents a T-shaped posture. The clothing is represented by numbers, where #1 represents the measuring suit, #2 represents the first spring and autumn clothes, #3 represents the second spring and autumn clothes, #4 represents the first summer clothes, and #5

represents the second summer clothes. All tables and graphs share the same index.

BMI (MAE)		Actions		
		#1	#2	#3
Clothes	#1	1.7157	1.7592	1.7775
	#2	3.5469	2.4377	2.7407
	#3	4.3108	4.0210	4.4000
	#4	4.8189	4.6268	2.9181
	#5	2.6872	2.7387	2.2595

Table 1. Performance of images-to-BMI model in various situations

According to the results, the overall Mean Average Error (MAE) for the images&height-to-BMI model that used both height and image as input was found to be 4.2699, which was lower than the images-to-BMI model. However, the performance of this model was still affected by clothing, as seen in table 2. In various situations, the model showed varying levels of accuracy depending on the clothes worn by the subjects during measurement.

BMI (MAE)		Actions		
		#1	#2	#3
Clothes	#1	3.9714	3.6058	4.7242
	#2	3.8489	4.4895	5.4579
	#3	4.3379	4.0284	4.8605
	#4	5.7061	3.2681	4.0324
	#5	4.2942	2.9339	4.4237

Table 2. Performance of images&height-to-BMI model in various situations

To enhance the visibility of the results, graphs were added to figure 8 and figure 9. The first graph is in the form of a bar graph and represents each result value of the model that predicts BMI from images. The second graph shows the result of a model that uses both BMI and height information. The x-axis in both graphs represents 5 types of clothing, and the color of each bar represents 3 types of action. It can be observed from the graphs that the first model is heavily influenced by the type of clothing worn.
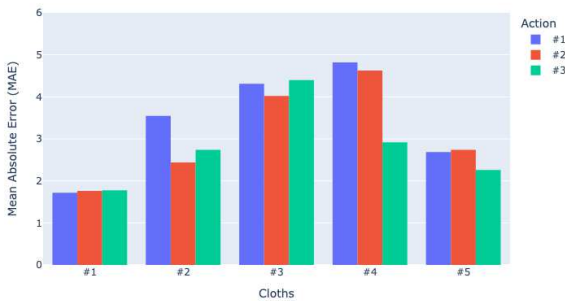

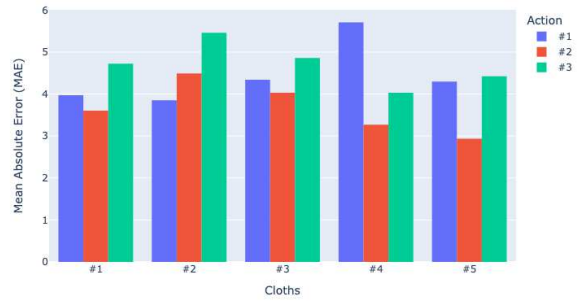Figure 8: Performance of images-to-BMI model in various situations in a bar graph

Figure 9: Performance of images&height-to-BMI model in various situations in a bar graph

In terms of the models predicting various indices, the images&height-to-Sizes model presented in the Models section had an overall MAE of 4.6199 for chest circumference, 5.0 MAE for waist circumference, 2.8599 MAE for hip circumference, 3.1099 MAE for thigh circumference, 1.45 MAE for forearm circumference, and 1.57 MAE for shoulder width. Every value is measured in centimeters. Another interesting finding of this study was that the performance of predicting various body part sizes was found to be less dependent on clothing than when predicting BMI. This suggests that clothing can have a greater impact on BMI measurements than on other body part sizes. The model's performance for each index was further analyzed in figures 10 to 15, which showed the performance of various scenarios.
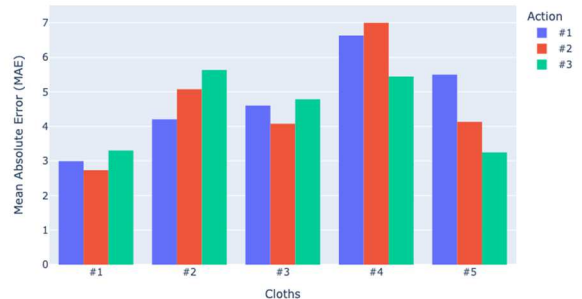

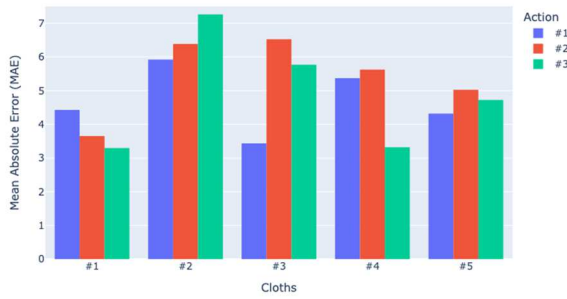Figure 10: Performance of images&height-to-Sizes model in various situations (chest circumference)

Figure 11: Performance of images&height-to-Sizes model in various situations (waist circumference)
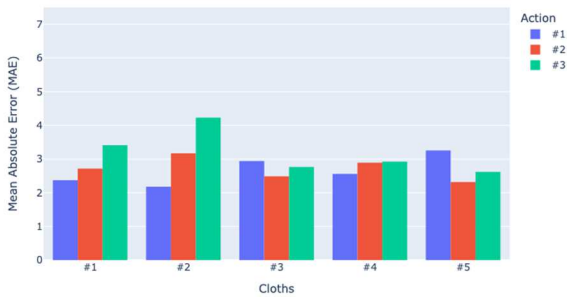

Figure 12: Performance of images&height-to-Sizes model in various situations (hip circumference)
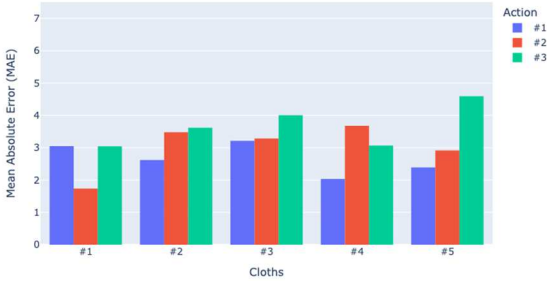

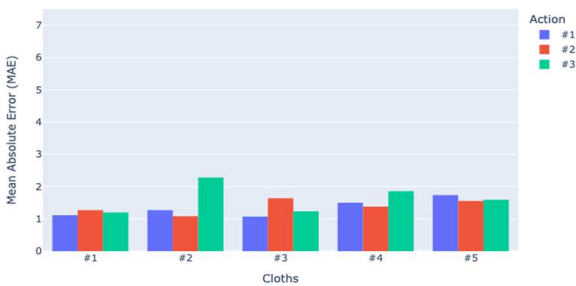Figure 13: Performance of images&height-to-Sizes model in various situations (thigh circumference)

Figure 14: Performance of images&height-to-Sizes model in various situations (forearm circumference)
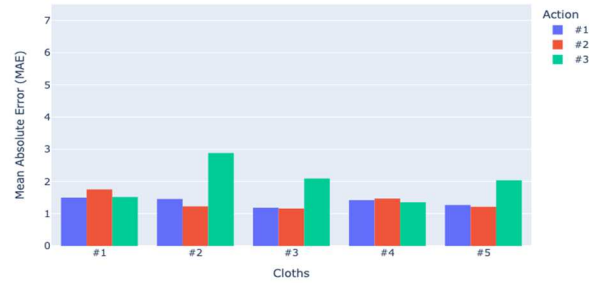

Figure 15: Performance of images&height-to-Sizes model in various situations (shoulder width)

The performance of the images&height&BMI-to-Sizes model, shown in figure 6 above, is presented below. This model additionally used ground truth BMI values as input, compared to the previous model. Each output is in centimeters as mentioned earlier, and the average MAE by index was found to be 4.0999 for chest circumference and 4.92 for waist circumference, indicating slightly better performance than when BMI was not used. However, hip circumference was 3.23, thigh circumference was 3.41, arm circumference was 1.58, and shoulder width was 1.83, confirming that the performance was slightly lower than when BMI was not used as an input. This confirms that using BMI as an additional input does not significantly affect performance change. Similarly, detailed situational performance for each index is presented in figures 16 to 21.
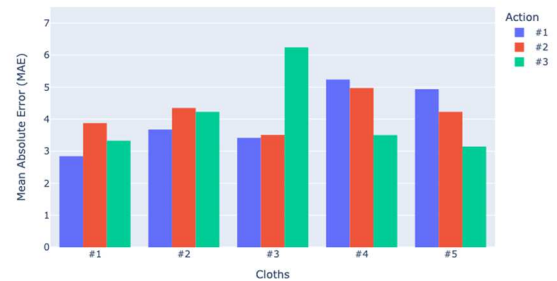

Figure 16: Performance of images&height&BMI -to-Sizes model in various situations (chest circumference)
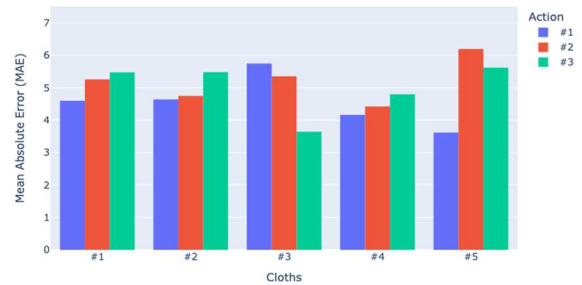

Figure 17: Performance of images&height&BMI -to-Sizes model in various situations (waist circumference)
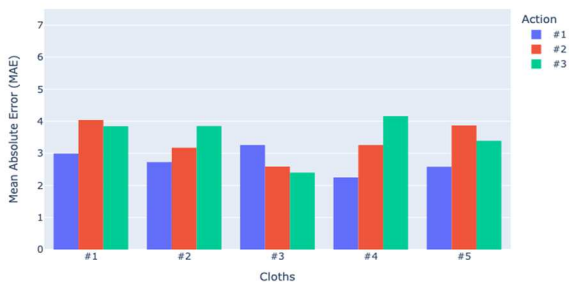
Figure 18: Performance of images&height&BMI -to-Sizes model in various situations (hip circumference)
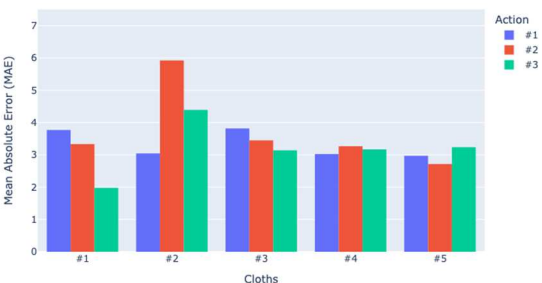

Figure 19: Performance of images&height&BMI-to-Sizes model in various situations (thigh circumference)
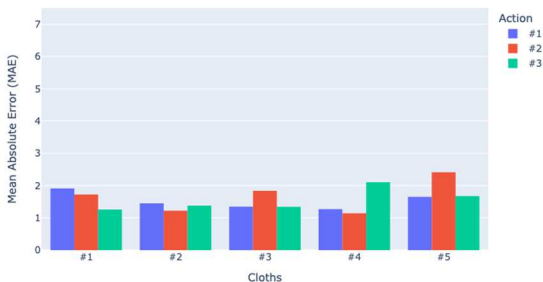

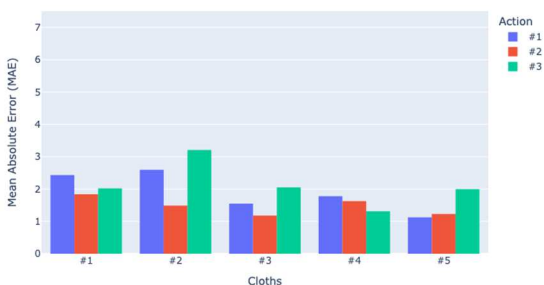Figure 20: Performance of images&height&BMI-to-Sizes model in various situations (forearm circumference)

Figure 21: Performance of images&height&BMI-to-Sizes model in various situations (shoulder width)

5. Discussion

This study represented a significant contribution to the field of measuring BMI and body part sizes using large-scale open data from body images. By using various models to compare and analyze the results, the study provides a deeper understanding of how these sizes can be derived from images in various situations. The findings of the study can also indirectly suggest the most effective type of data to use for such measurements.

However, the study also highlights a key challenge for future research. While the development of a model that can use both body and face images as inputs would be highly desirable, the lack of open data that includes BMI information along with both types of images represents a significant hurdle that should be addressed. In this regard, future studies will need to focus on building the necessary data infrastructure to support the development of such a model.

In addition to addressing this challenge, there is also room for further research in the direction of building more diverse types of backbone models to improve performance. By exploring new approaches to model design, researchers may be able to develop more accurate and efficient models for measuring BMI and other body part sizes from images. Overall, this study represents an important step forward in this field, and there is much to be gained from continued research in this area.

6. Conclusion

In conclusion, this study presents a novel method for predicting various body part sizes, including BMI, using only image inputs without the need for scales or complicated pre-processing procedures. By leveraging large-scale open images and labeled data, the method achieves impressive results, with an MAE of 3.1199 for BMI and varying MAE values for other body part sizes.

One significant advantage of this method is its convenience, as it does not require additional feature output or segmentation, making it straightforward to apply. Additionally, the study provides valuable insights into the impact of different environmental factors on the performance of various models for measuring body part sizes from images, which can inform future research in this area.

Overall, this study represents an important contribution to the field of measuring body part sizes from images and demonstrates the potential of large-scale open data for advancing research in this area. The findings of the study can be applied to a wide range of applications, including health monitoring, fitness tracking, and body size measurements for clothing and apparel.

Acknowledgement

References

[1] Yaemsiri, Slining, Meghan M. Slining, and Sunil K. Agarwal.: Perceived weight status, overweight diagnosis, and weight control among US adults: the NHANES 2003–2008 Study. International journal of obesity 35.8, 1063-1070 (2011).

[2] World Health Organization.: Obesity: preventing and managing the global epidemic. Report of a WHO consultation., (2000).

[3] Seidell, Jacob C., and Jutka Halberstadt.: The global burden of obesity and the challenges of prevention. Annals of Nutrition and Metabolism 66.Suppl. 2, 7-12 (2015).

[4] Maukonen, Mirkka, Satu Männistö, and Hanna Tolonen.: A comparison of measured versus self-reported anthropometrics for assessing obesity in adults: a literature review. Scandinavian journal of public health 46.5, 565-579 (2018).

[5] David J. Fisher, Illinois DOC labeled faces dataset, [online] Available: https://www.kaggle.com/davidjfisher/illinois-doc-labeled-faces-dataset.

[6] Dantcheva, Antitza, François Brémond and Piotr Tadeusz Bilinski. "Show me your face and I will tell you your height, weight and body mass index." 2018 24th International Conference on Pattern Recognition (ICPR) (2018): 3555-3560.

[7] Coetzee, V., Chen, J., Perrett, D. I., & Stephen, I. D.: Deciphering faces: Quantifiable visual cues to weight. Perception 39.1, 51-61 (2010).

[8] Wen, Lingyun, and Guodong Guo.: A computational approach to body mass index prediction from face images. Image and Vision Computing 31.5, 392-400 (2013).

[9] Kocabey, E., Camurcu, M., Ofli, F., Aytar, Y., Marin, J., Torralba, A., & Weber, I.: Face-to-BMI: Using computer vision to infer body mass index on social media. Eleventh international AAAI conference on web and social media., (2017).

[10] Zhi Jin, Junjia Huang, Aolin Xiong, Yuxian Pang, Wenjin Wang, Beichen Ding, Attention guided deep features for accurate body mass index estimation, *Pattern Recognition Letters*, Volume 154, 2022

[11] Pantanowitz, Adam & COHEN, EMMANUEL & Gradidge, Philippe & Crowther, Nigel John & Aharonson, Vered & Rosman, Benjamin & Rubin, D.M. Estimation of Body Mass Index from photographs using deep Convolutional Neural Networks. *Informatics in Medicine Unlocked*. 26. 100727. 2021.