Body Weight Estimation using 2D Body Image

Rohan Soneja[1], Prashanth S[2], R Aarthi[3]
Department of Computer Science and Engineering
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham, India

***Abstract*—Two dimensional images of a person implicitly contain several useful biometric information such as gender, iris colour, weight, etc. Among them, body weight is a useful metric for a number of usecases such as forensics, fitness and health analysis, airport dynamic luggage allowance, etc. Most current solutions for body weight estimation from images make use of additional apparatus like depth sensors and thermal cameras along with predefined features such as gender and height which generally make them more computationally intensive. Motivated by the need to provide a time and cost efficient solution, a novel computer-vision based method for body weight estimation using only 2D images of people is proposed. Considering the anthropometric features from the two most common types of images, facial and full body, facial landmark measurements and body joint measurements are used in deep learning and XG boost regression models to estimate the person's body weight. The results obtained, though comparable to previous approaches, perform much faster due to the reduced complexities of the proposed models, with facial models performing better than full body models.**

***Keywords*—Body weight estimation; deep learning; xgboost regressor; anthropometric features; computer vision**

I. Introduction

The purpose of this work is to estimate the weight of a person given only a two dimensional image. Many use cases necessitate the estimation of body weight without the physical measurement or presence of a person directly. For example, in health analysis to check the weight through mobile devices for a quick estimation, in forensics to gain additional identification features, in airports to estimate weight to aid dynamic baggage allowance, for physicians working remotely for rural patients, etc. Additionally, in social networks containing advertisements, considering the huge volume of images of people uploaded on the Internet daily, body weight can be taken as another admetric.

A novel cost-effective method that is computationally less intensive is proposed for estimating body weight using only the 2D images of a person and features extracted from the images. The person's image has the facial and the full body components of the person. Different procedures are applied to each type of image to obtain the weight of the person in the corresponding image.

For the model using only facial image a novel dataset, IDOC-Mugshots [1], containing over 70,000 frontal face images of prison inmates was used. This dataset, available on Kaggle, was obtained from the Illinois Department of Correction. In addition to that, VIP-Attribute dataset [2] containing 1000 images of celebrities was scraped from popular websites. For the model using full-body images Visual-Body-to-BMI dataset [3] containing around 6000 images scraped from the subreddit r/ProgressPics was used. A deep learning model and an XGBoost regressor model were trained on the features extracted from both types of images and the results obtained showed the feasibility of using only 2D images for body weight estimation. The results obtained were analysed when using only facial images or full body images. They were comparable to previous work even with the reduced feature set and computationally less intensive methods used.

The remaining sections in the paper are organised as follows: first, we review related work done in the area and their results; this is followed by a brief description of the proposed architecture; then the implementation using features from face and full body is explained; finally a summary of the results and the conclusions drawn along with scope for future work are provided.

II. Related Work

Previous approaches for this task either used additional apparatus such as RGB-D sensors or thermal cameras in addition to a camera to obtain more features than just the 2D image. In other cases, other features are used as input to the learning models that aren't obtained from the image such as gender, age, etc. These methods require prior data collection about the person or additional equipment and hence are neither fast nor cost efficient.

Some work has analyzed body weight or BMI from face images [4] [5] [6]. Wen et al. [7] first proposed a computational method for BMI prediction from face images based on the MORPH-II dataset, which obtained mean absolute errors for BMI in the range of 2.65-4.29 for different categories based on ethnicity. They also analyzed the correlations between facial features and BMI values. An Active Shape Model is used to extract facial features which are used to predict BMI using various regression techniques. Barr et al. [8] used facial landmarking to figure out adiposity (facial fatness) which positively correlates to the weight of the person. This method is less acurate in exteme underweight and obese cases though. A support vector machine regression model was used. Windhager et al. [9] showed that shape of the face has direct correlation with several body characteristics such as height and weight and can be determined by facial landmarking and spatial scaling. A total of 71 landmarks and semi landmarks were digitized to capture facial shape. Regression and geometric morphometric toolkit tool were used to estimate facial fatness. Additionally as a feature set, height measurement using anthropometer and saliva sample testing done apart from facial front photograph. Tai et al. [10] used Kinect sensors to estimate BMI using facial data on a regression model. Recently, Haritosh et al. [11] used convolution neural networks and artificial neural networks to estimate the weight using the facial image extracted using Viola-Jones detector.

There are a few studies on estimating human body weight or BMI from body related data [12] [13] [14], such as body measurements, 3-dimensional (3D) body data and RGB-D body images. Jiang et al. [3] used images of entire front body by scraping data from a Reddit page called r/ProgressPics. To estimate BMI they used anthropometric measurements from body contour segments with the help of skeletal joints and estimated the weight based on those features.

The body weight was studied directly by Velardo et al. [15] from anthropometric data collected by National Health And Nutrition Examination Survey, Centers for Disease Control and Prevention using a polynomial regression model to estimate the weight within 4% error using 2D and 3D data extracted from a low-cost Kinect RGB-D camera output. Pfitzner et al. [16] also used RGB-D camera data; this demonstrated a body weight estimation by volume extraction from RGB-D data with an accuracy of 79% for a cumulative error of ±10%. Compared to a physician's estimation, this approach is already more suitable for drug dosing. Pichler et al. [17] estimated human body volume in clinical environment by eight stereo cameras around a stretcher and bioelectrical impedance analysis. Nguyen et al. [18] estimated body weight using a side view feature and a support vector regression model to obtain an average error of 4.62 kg for females and 5.59 kg for males.

III. Proposed Architecture

Most of the previous research has been to estimate BMI using facial images and in some cases body weight. While that is shown to have a high correlation to the BMI and body weight, it is prone to high error. Without the use of external hardware, two methods to measure the body weight using a limited feature set are proposed. One is using face data, and the other is an extension of that which uses full body image data (including facial features).

Previous studies have shown that estimation of a person's weight given only the image of the person's face is possible due to the correlation of the weight with facial fatness, i.e. adiposity [19]. This can be measured by taking various measurements across the face such as the height and width of the face, length and width of the nose, etc.

Next, given the full body image of a subject, a novel method is chosen to extract features from the image. Instead of making anthropometric measurements with respect to body contours, measurements are made with respect to the bone joints of the person which are found to be correlated to the body weight [20]. Using these measurements result in several advantages.

- With current models, bone joint coordinates are easier to extract from images instead of the body contour, i.e., less computationally intensive.
- Bone joint coordinates are usually more accurate than body contour segments, i.e., more reliable measurements.
- Even in case of reliable body contour measurements, unlike that, bone joint coordinates are not affected by baggy/tight garments, i.e., less clothing bias.

These measurements are used in conjunction with the facial measurements as input features with the weight of the person being the target feature.

Fig. 1. Face Landmarks.

Finally, in both cases, a split of the feature set is used to train the machine learning model while the remaining is used to evaluate it, using the mean absolute error as the evaluation metric.

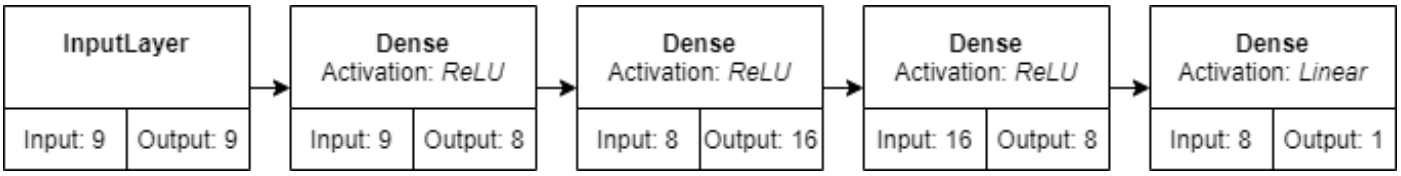
IV. Implementation

A. Features from Face

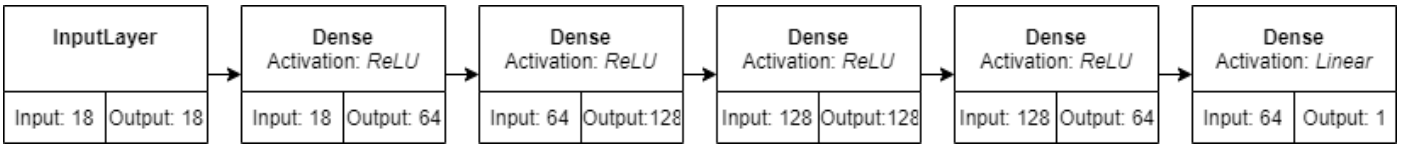
Weight estimation with only a 2D facial image was done by extracting measurements of various noticeable regions of the face such as eyes, eyebrows, nose, mouth and jawline. This was applied to two datasets, VIP-Attribute dataset [2] and a novel IDOC-Mugshots dataset [1]. Due to the varied nature of images and issues caused by background noise, first, face localisation is performed using the face detector in Python's dlib library. The cropped-out region of interest is then used for landmarking to detect key facial structures and thereby their measurements. For this task, the facial landmark detector included in the dlib library is used, which is internally implemented using an ensemble of regression trees [21]. This estimates 68 coordinates (x,y) on the face (Fig. 1) which are used to measure facial features such as:

- Left eyebrow width (18-22)
- Right eyebrow width (23 - 27)
- Left eye width (37 - 40)
- Right eye width (43 - 46)
- Nose width (32 - 36)
- Nose length (28 - 34)
- Outer lip width (49 - 55)
- Inner lip width (61 - 65)
- Face height (28 - 9)
- Face width (1 - 17)

Since the measurements are made from a 2-dimensional image there is a lack of perception of depth causing those



(a) For facial features



(b) For full body features

Fig. 2. Deep Learning Model Architectures.

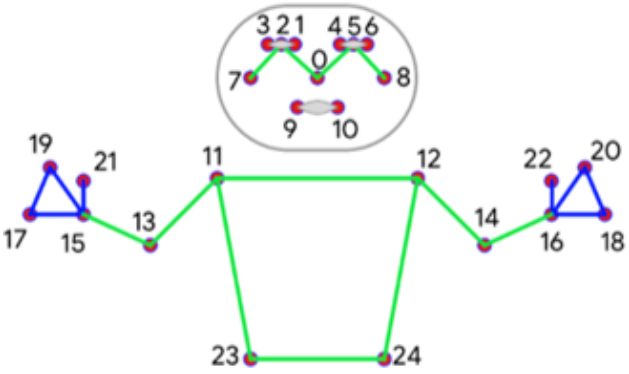


Fig. 3. Body Landmarks.

farther away in the image to have smaller measurements which hampers the performance of the model. Hence all the measurements were converted into ratios by diving the lengths with the width of the face.

These extracted ratios exhibiting facial fatness (adiposity) are used as features for training. Hence no additional features other than the image itself is used for weight estimation. A deep learning model (Fig. 2a) and an XGBoost Regressor (40 estimators) were used on both datasets to estimate the body weight.

B. Features from Full Body

Initially, the BlazePose model [22] is used to detect important landmarks (Fig. 3) of the most prominent human body in the image that is provided. Since the dataset contains images with only one subject, the body joints detected are for the person of interest only. The dataset contains a lot of images that do not have the joints visible clearly. Based on the additional visibility parameter provided by MediaPipe's Pose API, images that do not cross a certain threshold for all the joints that are considered are eliminated from training. There are four pairs of joints that are used, viz. shoulder, hip, elbow, and wrist. Various Euclidean distances are measured between the joints and they are scaled down with respect to the inter-shoulder distance (11 - 12) to ensure uniformity of measurements between images of various resolutions and subject to image ratios.

- Left-shoulder to left-hip (11 - 23)
- Right-shoulder to right-hip (12 - 24)
- Left-hip to right-hip (23 - 24)
- Left-shoulder to right-hip (11 - 24)
- Right-shoulder to left-hip (12 - 23)
- Left-shoulder to left-elbow (11 - 13)
- Left-elbow to left-wrist (13 - 15)
- Right-shoulder to right-elbow (12 - 14)
- Right-elbow to right-wrist (14 - 16)

Finally, the features used for training are the aforementioned ratios along with the facial measurements as described in the facial model. It is important to note that the facial measurements taken on full body images are not accurate due to the fact that the face of the person takes up a very small region of the image as compared to the rest of the body.

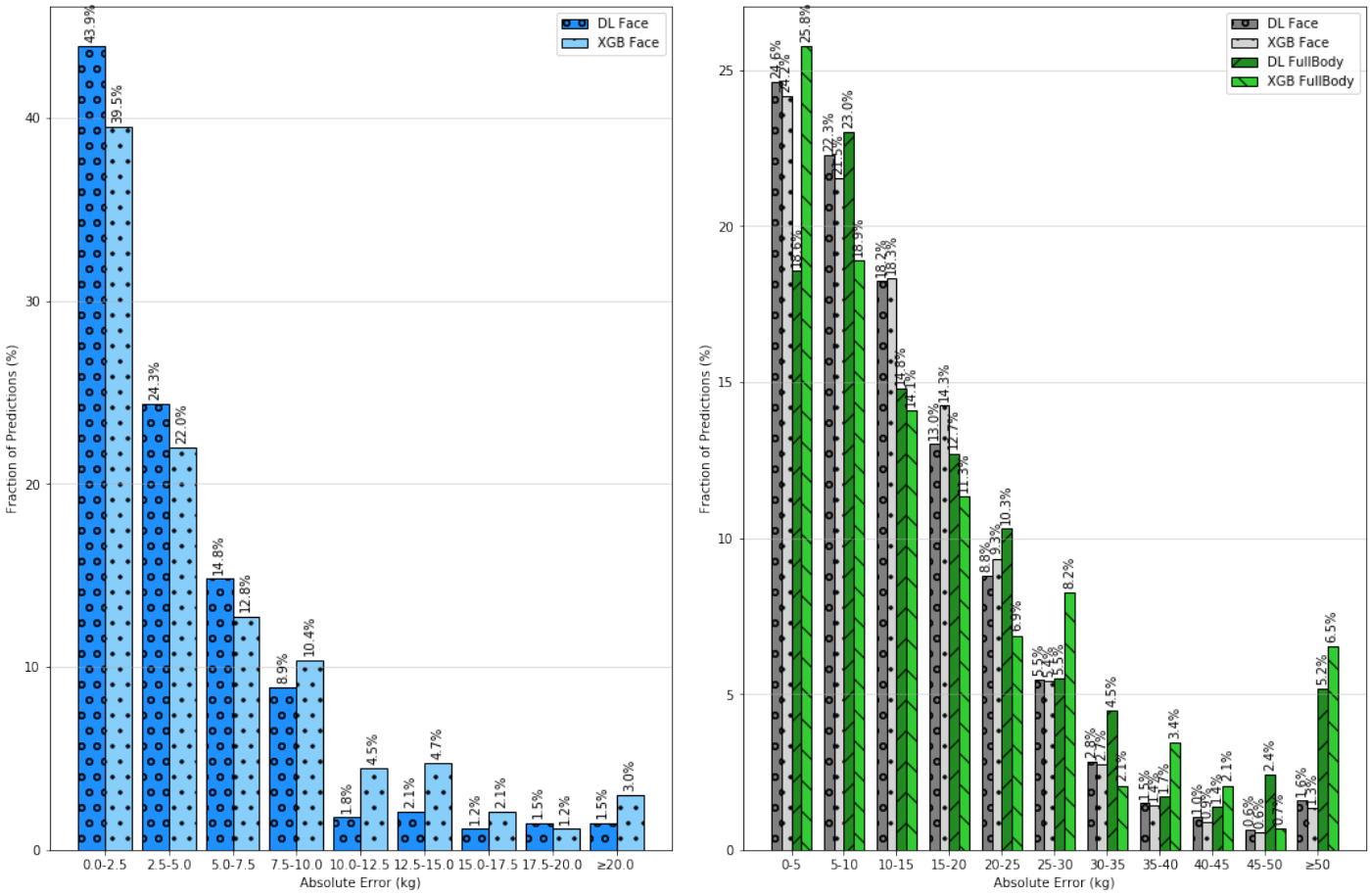
A deep learning model (Fig. 2b) and an XGBoost Regressor (7 estimators) were used on the dataset to estimate the body weight.

V. Results

The metric chosen to evaluate the model performance is Mean Absolute Error (MAE). MAE is calculated by taking the mean of absolute value of errors between the predicted weight and the actual weight of the person.

$$MAE(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Table I shows the MAE for each of the models and datasets. Overall and in case of facial features, VIP-Attribute dataset has yielded the best results with the deep learning model with $MAE = 9.8kg$ and in case of full body features, XGBoost performs slightly better than the deep learning model with $MAE = 18.2kg$. Fig. 4 shows the distribution of errors made by the various models for each of the datasets.



(a) On VIP-Attribute Dataset

(b) On IDOC-Mugshots and Visual-Body-to-BMI Datasets

Fig. 4. Distribution of Errors for Weight Estimation.

TABLE I. MEAN ABSOLUTE ERRORS

S. N.	Dataset (only face)	Model	MAE (kg)
1.	VIP-Attribute	Deep Learning	9.8
2.	VIP-Attribute	XGBoost Regressor	11.9
3.	IDOC Mugshots	Deep Learning	13.5
4.	IDOC Mugshots	XGBoost Regressor	13.5

(A) MAE FOR FACE DATASET

S. N.	Dataset (full body)	Model	MAE (kg)
1.	Visual Body to BMI	Deep Learning	18.6
2.	Visual Body to BMI	XGBoost Regressor	18.2

(B) MAE FOR BODY DATASET

If the graph is skewed to the left, it indicates a better model performance since the errors are small in that case. The relative spike on the right end can be explained by the the fact that the last bars represent errors more than the specified value and not just within a particular range. Fig. 5 shows the comparison between the ground truth body weights with their corresponding estimations as provided by the ML models in the form of a scatter plot. The dotted line represents the ideal model with no error and the surrounding solid lines represent an absolute error of 15 kg. It can be seen that the models tend to underestimate weights more than 100 kg and slightly underestimate weights less than 60 kg.

Although it may seem unintuitive that the facial model performs better than the full body model (about 40% of predictions having error less than 2.5 kg (Fig. 4a) as opposed to around 25% of predictions having error less than 5 kg (Fig. 4b)), it is justified as follows.

- The facial datasets are much cleaner and regular due to faces being in the exact same positions for all images.
- Due to extensive occlusion in Visual Body to BMI owing to varying postures, object obstructions, etc., a lot of the measurements are approximated.
- The VIP-Attribute dataset has lesser range of body weights compared to the other datasets (Fig. 5).
- The full body feature set considers only bone joints which are inherently less correlated to the body weight as opposed to the body contour.
- The version of the Pose API that was used did not include bone joints below the waist, hence reducing

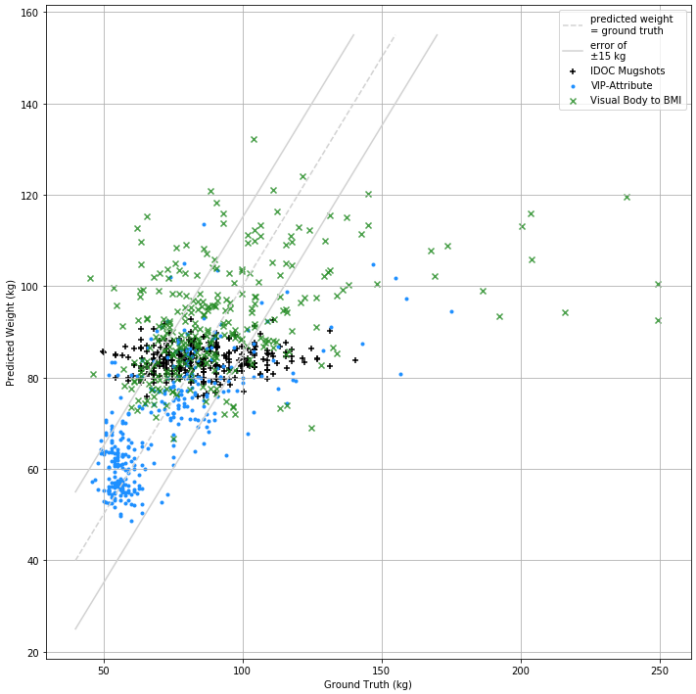


Fig. 5. Ground Truth v. Predicted Weight Scatterplot for Deep Learning Models.

the availability of femur length in the feature set, from which the models may have benefitted.

- The Visual Body to BMI dataset contains several instances of weights lying on the extreme ends (Fig. 5) due to the nature of the source of the dataset and hence in some cases, the error is too high (even more than 100 kg as evident from Fig. 4b).

With respect to the computational intensity, from experimentally timing the runtime of the feature extraction methods, it was found that extracting landmarks (using BlazePose [22]) instead of body contour (using CRF-as-RNN [23]) was 40 times faster, given the same hardware and running conditions. Of course, once the body contour is found, it requires more preprocessing (in the form of contour smoothening, pixel counting, length extraction, etc.), making it even more time consuming. Also, considering our model's reduced complexity using simpler features, we report a marginally higer MAE compared to 8.51 kg as reported by Dantcheva et al. [2] for the same VIP-Attribute dataset. Haritosh et al. [11] report an even higher MAE of 13.29 kg owing to the separate Reddit-HWBMI facial dataset that was used.

VI. Conclusion

In this work, the body weight of a person was estimated given just the image of the subject. Two types of datasets are employed viz. facial images and full body images. Features are extracted using publicly available libraries and they are used to train deep learning and XGB regressor models. The results obtained were compared and it was found that it is a viable and efficient method to estimate the body weight using just the person's image as long as the weights are not on the higher extreme. To counter that problem, in the future, an efficient body contour detector may be developed that uses the landmarks to facilitate itself so that the model is not made computationally intensive. Once the Pose API is updated to support depth attributes to the joint coordinates, that too can be used as a feature among other currently unexplored features to improve the overall MAE.

References

[1] Elliot, "Idoc-mugshots dataset," 06 2018. [Online]. Available: https://www.kaggle.com/elliotp/idoc-mugshots

[2] A. Dantcheva, F. Bremond, and P. Bilinski, "Show me your face and i will tell you your height, weight and body mass index," in *2018 24th International Conference on Pattern Recognition (ICPR)*, 8 2018, pp. 3555–3560.

[3] M. Jiang and G. Guo, "Body weight analysis from human body images," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2676–2688, 10 2019.

[4] K. Vikram and S. Padmavathi, "Facial parts detection using viola jones algorithm," in *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2017, pp. 1–4.

[5] T. Keshari and S. Palaniswamy, "Emotion recognition using feature-level fusion of facial expressions and body gestures," in *2019 International Conference on Communication and Electronics Systems (ICCES)*, 2019, pp. 1184–1189.

[6] N. Parameswaran and D. Venkataraman, "A computer vision based image processing system for depression detection among students for counseling," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, pp. 503–512, 04 2019.

[7] L. Wen and G. Guo, "A computational approach to body mass index prediction from face images," *Image and Vision Computing*, vol. 31, no. 5, pp. 392 – 400, 2013.

[8] M. Barr, G. Guo, S. Colby, and M. Olfert, "Detecting body mass index from a facial photograph in lifestyle intervention," *Technologies*, vol. 6, no. 3, p. 83, 8 2018.

[9] S. Windhager, F. L. Bookstein, E. Millesi, B. Wallner, and K. Schaefer, "Patterns of correlation of facial shape with physiological measurements are more integrated than patterns of correlation with ratings," *Scientific Reports*, vol. 7, no. 1, p. 45340, 5 2017.

[10] C. Tai and D. Lin, "A framework for healthcare everywhere: Bmi prediction using kinect and data mining techniques on mobiles," in *2015 16th IEEE International Conference on Mobile Data Management*, vol. 2, 6 2015, pp. 126–129.

[11] A. Haritosh, A. Gupta, E. S. Chahal, A. Misra, and S. Chandra, "A novel method to estimate height, weight and body mass index from face images," in *2019 Twelfth International Conference on Contemporary Computing (IC3)*, 8 2019, pp. 1–6.

[12] K. Padmavathi and S. Nithin, "Comparison of image processing techniques for detecting human presence in an image," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 383–388.

[13] K. Hena, J. Amudha, and R. Aarthi, "A dynamic object detection in real-world scenarios," in *Proceedings of International Conference on Computational Intelligence and Data Engineering*, N. Chaki, N. Devarakonda, A. Sarkar, and N. C. Debnath, Eds. Singapore: Springer Singapore, 2019, pp. 231–240.

[14] S. T. and P. B. Sivakumar, "Human gait recognition and classification using time series shapelets," in *2012 International Conference on Advances in Computing and Communications*, 2012, pp. 31–34.

[15] C. Velardo and J.-L. Dugelay, "Weight estimation from visual body appearance," 10 2010, pp. 1 – 6.

[16] C. Pfitzner, S. May, and A. Nüchter, "Evaluation of features from rgb-d data for human body weight estimation," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 10 148 – 10 153, 2017, 20th IFAC World Congress.

[17] K. Santner, M. Rüther, H. Bischof, F. Skrabal, and G. Pichler, "Human body volume estimation in a clinical environment," 01 2009.

[18] T. V. Nguyen, J. Feng, and S. Yan, "Seeing human weight from a single rgb-d image," *Journal of Computer Science and Technology*, vol. 29, no. 5, pp. 777–784, 9 2014.

[19] L. Wen, G. Guo, and X. Li, "A study on the influence of body weight changes on face recognition," in *IEEE International Joint Conference on Biometrics*, 2014, pp. 1–6.

[20] C. Velardo, "Anthropometry and soft biometrics for smart monitoring," 2012, pp. 34–46.

[21] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1867–1874.

[22] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "Blazepose: On-device real-time body pose tracking," *arXiv preprint arXiv:2006.10204*, 2020.

[23] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, "Conditional random fields as recurrent neural networks," *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2015.179