Predicting body measures from 2D images using Convolutional Neural Networks

João W. M. de Souza[§†], Gabriel B. Holanda[§†], Roberto F. Ivo[§‡], Shara S. A. Alves[§‡], Suane P. P. da Silva[§‡], Virgínia X. Nunes[§], Luiz Lannes Loureiro[§], C. H. Dias-Silva[§] and Pedro P. Rebouças Filho[§†]

[§] Image Processing Laboratory, Signs and Applied Computer (LAPISCO)
[†] Federal Institute of Education, Science and Technology of Ceará (IFCE), Ceará, Brazil
[‡] Federal University of Ceará (UFC), Ceará, Brazil
Email: {wellmendes, gabrielbandeira, robertoivo, shara.shami, suanepires, virginia.nunes}@lapisco.ifce.edu.br, luiz.lannes@yahoo.com.br, chrishdias@gmail.com, pedrosarf@ifce.edu.br

***Abstract*—Nutrition is a significant determinant of health, the resolution of many nutritional issues, initially requires an anthropometry examination. Body measures provide data for studying the relationship between diet, nutritional status, and health. Manual and automatic methods can perform body measurements. The manual method usually uses an anthropometric tape. However, the automatic process uses the equipment of Dual-energy X-ray absorptiometry (DXA). Our work presents a new approach to calculate body measures using 2D Camera Images, applying Digital Image Processing, Convolution Neural Networks, and Machine Learning techniques. The dataset used contains 38 exams, for each exam, has four digital images and the dimensions of body measurements, performed by a specialist. The methods used in this work for segmentation were Dense Human Pose Estimation - CNN with the Bayesian, K-Nearest Neighbors, Support Vector Machine, Decision Threes, Adaptive Boosting, Random Forest, Multilayer Perceptron and Expectation-Maximization classifiers. The approach with Dense Human Pose Estimation and Expectation-Maximization reached the best results, with mean squared error (MSE) always bellow 4.606 $\pm$ 3.412 cm when compared with specialist measures.**

***Index Terms*—Nutrition, Body measures, Convolutional Neural Networks, Machine Learning.**

I. Introduction

Proper nutrition is fundamental for health, embodied in article 25 in the Universal Declaration of Human Rights [1]. Malnutrition can take many forms, including wasting, stunting, and micronutrient deficiencies, and as overweight, obesity, or diet-related noncommunicable diseases [2].

One of the malnutrition most current problems are obesity and overweight that is an excessive fat accumulation that presents a health risk [3]. Obesity is a significant risk factor for several chronic diseases, including diabetes, cardiovascular diseases, and cancer [4].

Worldwide obesity, has nearly tripled since 1975 [5]. In 2016, more than 1.9 billion adults, 18 years and older, were overweight. Of these, over 650 million were obese [5]. Besides that, over 340 million children and adolescents aged 5-19 were overweight or obese in 2016, but despite that obesity, is preventable [5].

Since the treatment and follow-up of obese people, as all people in a nutrition context, is a matter of concern in the current worldwide context, there are several ways to measure a person's degree of obesity. One such form is the Body Mass Index (BMI) [6], which is a measure widely used to determine the degree of obesity. This measurement uses a person's weight (in kilograms) divided by the square of his height (in meters). A person with a BMI of 30 or more is generally considered obese. A person with a BMI of 25 or above is considered overweight.

Although it is a commonly used measure, it is not accurate and applicable to everyone. Thus, it is necessary to carry out more in-depth monitoring, using body measures. The manual method usually uses an anthropometric tape to perform body measurements. The most elaborate and automatic method uses Dual-energy X-ray absorptiometry (DXA) [7].

Among the ways to perform anthropometric examinations, some of these measurements use images [8]–[12]. The limiting factor in this way is that it is characterized by being an invasive medium, in addition to being tedious, time-consuming, and requiring the professional's clinical knowledge. Therefore, obtaining body perimeters through images is something that facilitates the nutritional evaluation of people and, thus, their monitoring for the prevention of obesity and other malnutrition ways.

Some works in literature attempt to present a method that is efficient and fully automatic in estimating body measures. Initially, semi-automatic methods request a large amount of information determined by a specialist. [13], [14]. Other works presented methods that required measurement time that makes it impracticable in the clinical environment [15].

Recently, several works in the literature have obtained results that demonstrate the potential of using Convolutional Neural Networks (CNNs) for the health area. For example, to diagnosis of diseases [16], organ segmentation [17], and pose estimation [18].

The objective of this work is to present an accessible, efficient, fast, and fully automatic method to perform the prediction of body measurements from 2D images through image segmentation using CNNs and machine learning.

This paper is organized as follows: Section II presents an

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

overview of the methods used. In contrast, Section III presents the applied methodology, from the segmentation of the body to the form of measurement. Section IV displays the results of this approach, and Section V includes a discussion of these results and the future works.

II. Overview of Machine Learning and Convolutional Neural Networks (CNNs)

This section describes a literature overview of the machine learning and Convolutional Neural Networks (CNNs) methods used in our work.

A. Machine Learning

The Bayesian supervised classifier is a machine learning method based on the Bayes Decision Theory [19]. With this technique, a sample is classified through the posterior probability using the priori and class-conditional probabilities. Gaussian distribution is the most common rule applied in the Bayesian classifier, and Mean and Covariance Matrixes are the only two necessary parameters.

K-nearest neighbors (*k*NN) is a supervised learning algorithm [20]. *k*NN consists of finding the *k* closest labeled samples to unclassified samples and to use these labels to classify the unknown sample. The highest computational cost of *k*NN is during the test phase since the unlabeled example will have to be compared with all the samples contained in the input data [21].

As Bayes, the Support Vector Machine (SVM) is part of the group of supervised machine learning methods. The SVM goal is to find a function that achieves a maximum reduction in the probability that the learning output data is different from the desired output data [22]. Therefore, the SVM seeks a limit or hyperplane that achieves the highest margin of separation between different categories. For nonlinear data, different kernels aim to accomplish this separation of planes.

Decision Tree (DT) is a binary tree, where each non-leaf node has two child nodes [23]. DT can be used either for regression or classification. In regression, each tree leaf has a constant, so the approximation function is piecewise constant. Differently, in classification, each tree leaf is marked with a class label where compound leaves may have the same label.

Boosting is a powerful learning technique that solves the supervised classification learning assignment. This learning concept combines the performance of many less severe classifiers to produce a robust alternative [24]. Different variants of boosting are known as Gentle AdaBoost, LogitBoost, Real AdaBoost, and Discrete Adaboost [25]. In this work, we used the Discrete Adaboost (ADAB) variant.

Random Forest (RF) is the supervised learning method, based on the combination of decision trees [26]. RF starts from the premise that a random set of trees operate as a decision summit, in which, their correctness surpasses that achieved by the individual decisions of each tree. The RF algorithm has the advantage of not being overfitted when the number of trees increases [26].

The structure of the Multi-layer Perceptron (MLP) consists of a multilayer neural network [27]. MLP is composed of Input layers, Hidden layers, and Output layers. Different activation functions are present in each category of the network. The downward gradient and the backpropagation perform the adjustment of the weights in this network.

The Expectation-Maximization (EM) algorithm computes the multivariate probability density function parameters in the configuration of a Gaussian mixture distribution with a predetermined number of mixtures [28].

B. Convolutional Neural Networks

Dense Human Pose Estimation (DensePose), is a variation of Mask-RCNN, this being a region-based CNN that was designed to mapping all human pixels in an RGB image to a 3D representation of the human body based on the surface [29].

DensePose was implemented by entering a large-scale, manually annotated dataset and a variant of Mask-RCNN, a flexible and adaptable structure for object segmentation.

III. Methodology

This section describes the adopted methodology of this work. Firstly, we present the acquisition pattern to the human body images; then, we describe the segmentation approach. Finally, we explain how we measure each limb in the human body. Figure 1 presents a flowchart of the proposed approach.

A. Acquisition Pattern

The acquisition pattern of the human body images is essential for the measurement methodology in digital images adopted in this work. Initially, it captured four images of the same patient in the perspectives: front, back, right side, and left side. The justification is associated with the need to view the entire body. The perspectives of the human body are presented, as an example, in Figure 1(1).

It is also necessary to inform the patient's height to create a correlation between pixel/centimeter scale (pixel per centimeter).

The environment must contain natural or artificial lighting, in addition to having a homogeneous color background. The patient's clothing must show as much of the skin as possible so that no clothing gets in the way of calculating the perimeter of the human body limbs.

In front and back perspectives, the patient needs to look straight ahead. The arms need to be open at 45° of the body, with the palms of the hands flat. The legs should be aligned with the shoulders and always keeping one leg spaced from the other.

In the lateral perspectives, the patient needs to have his head straight, facing the side where his hand direction. The patient's head needs to be slightly pointed upwards so that it is possible to visualize the patient's neck correctly. The arm that is on the side of the camera needs to be doing 90° with the trunk. The patient's torso must completely omit the other arm. The legs must be aligned with the torso. Besides, the leg that is on the side of the camera must omit the opposite leg.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

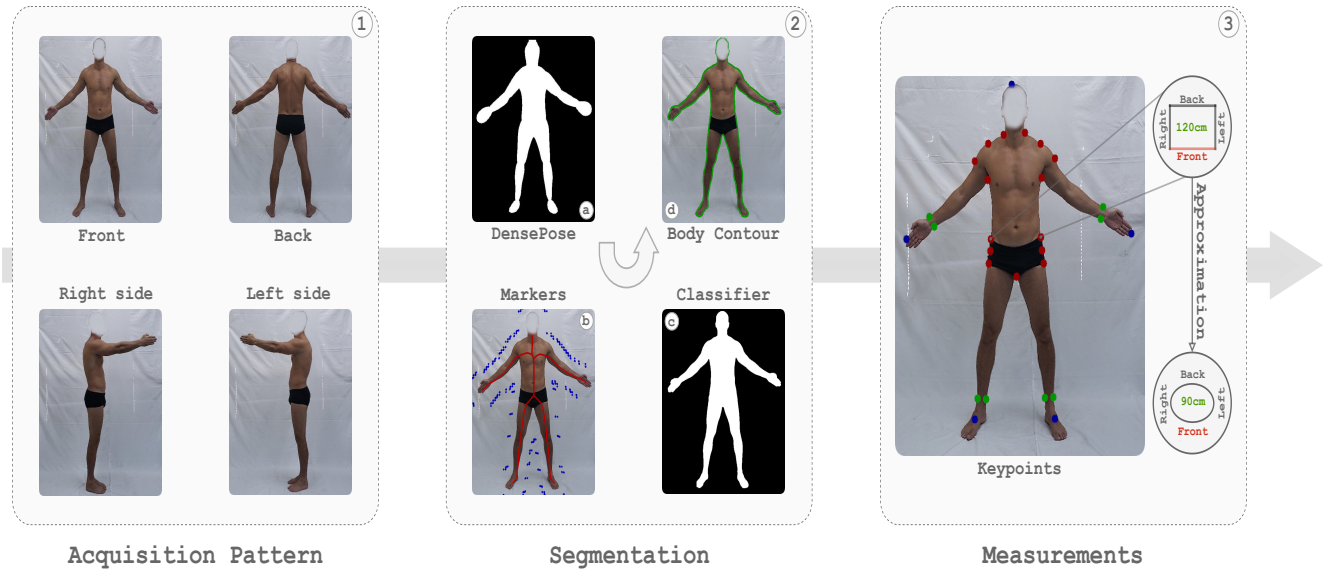

Fig. 1. Flowchart of the proposed approach: (1) Adopted pattern to capture source images; (2) Segmentation methodology using DensePose and some traditional classifiers; (3) Keypoints methodology for measurements and correlation between source images.

B. Segmentation

Following the acquisition, the next task is to segment the body from each perspective. For this, we use DensePose so that we can detect and identify the entire human body in the image. After applying DensePose the result is shown in Figure 1(2-a).

After using DensePose, we apply the dilating operation with a rectangular structuring element with a size of $15x15$ in the DensePose output image. In parallel to this operation, we applied the Zhang-Suen thinning algorithm [30] also to the DensePose output image. Here we obtain the so-called Markers, represented in Figure 1(2-b), where the blue color represents the contours points of the image obtained after the dilating operation, and the red color represents the contours points obtained from the image generated by the Zhang-Suen thinning algorithm [30].

After obtaining these Markers, we store them in two different vectors. The first vector represents points inside the human body, and the second vector represents points outside the human body. The stored value is the pixel intensities of the image's RGB channels. So these features vectors are used as input data to train each classifier. Following that, we use the chosen classifier to predict whether each pixel is part of the human body or not. Figure 1(2-c) presents the classification example result.

Finally, for each classifier, we obtain a prediction of body region and calculate external body contour to plot it on the original image. The result can be seen in Figure 1(2-d).

C. Keypoints

After segmentation and obtaining the body contour in all perspectives, before measuring, we get the Keypoints. These Keypoints divided into three categories: Exterior, Interior, and Divisional. Figure 1(3) presents Keypoints, where the Exteriors, Interiors, and Divisional are blue, red, and green dots, respectively.

The detection of these points is a very intuitive way due to the acquisition pattern used in this work. The Internal Keypoints are obtained from the combination of the external Keypoints using some logic elements. An example of that is the maximum point between the right hand and the right foot or the maximum point between the right foot and the left foot. Thus, we find the points highlighted in red in this step.

Finally, the Divisional Keypoints are obtained from the combination of the External Keypoints and Internal Keypoints and using some logic elements. The two points with the shortest distance from the two-to-two combination between the point of the right hand and the point with the next indexing in the contour, following the left and the right of the body contour. Ensuring that the body contour has a standard spacing, these two points delimit the right wrist. So, we find the points highlighted in green in this step.

We used these Keypoints to correlate the images in different perspectives, together with the person's height in pixels, since we can determine the pixel/centimeter scale. The Equation 1 presents how to determine the pixel/centimeter scale to each perspective.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

$$scale_{cm}^{px} = \frac{height_{px}}{height_{cm}} = \frac{dist(point_{midfeet}, point_{head})}{height_{cm}} \quad (1)$$

The correlation occurs from any limb of the human body, such as the trunk, right arm, left arm, right leg, and left leg. Each limb uses a certain amount of perspectives to be correlated because a limb may need only two or three perspectives for a full view of the limb. Following, present which perspectives are used for each limb:

- Trunk: Front, Back, Right and Left;
- Right arm: Front, Back;
- Left arm: Front, Back;
- Right leg: Front, Back, Right;
- Left leg: Front, Back, Left.

Figure 1(3) presents an example of this correlation where a trunk slice is displayed using all four perspectives. Then, after this correlation, we were able to sketch slices of the body members on the z-axis (height).

D. Measurements

Since all Keypoints are correlated, we can delimit each limb's boundaries. These limits can provide the contour of all members, from the complete contour of the human body, obtained in the segmentation stage. Finally, we define a step for the contour limb, and we slice the limb from the beginning to the end of each limb on the z-axis (height).

After obtaining parts of each body limb, the slices outline a quadrilateral shape, which requires an approximation to the geometric shape of the human body.

Figure 2 presents the illustration of adopted approximations. Breast and waist used correlations of four perspectives smoothing the rectangle edges shown in the top left of Figure 2. Fist used correlations of front and back views, building a circumference with mean size between these two lines, as shown in the top right of Figure 2. Forearm and relaxed biceps used correlations of front and back perspectives, building an ellipse with mean size between these two lines, as shown in the bottom left of Figure 2. Thigh and calf used correlations of three views constructing an ellipse inscribed in circumference, as shown in the bottom right of Figure 2.

IV. Results

This section presents the results in three steps. First describes the dataset, second explain the experimental setup, and third discusses our results and compares the present work best approach against specialist analysis.

A. Dataset

The dataset used in this study corresponds to a set of 38 exams of some skinfold measurements realized by a specialist, correlated with body images of the patient. Each exam was performed by a nutritionist that carry measures of fist, forearm, breastplate, waist, relaxed biceps, thigh, and calf. Also, for each exam were acquired four images containing frontal, back, right-side, and left-side perspectives of the patient body. A smartphone camera obtained images of all body perspectives with a resolution of 1920 × 1080 pixels in the depth of 8 bits.

The dataset and acquisition were analyzed and approved by the Research Ethics Committee of IFCE (Protocol No. 3.593.367), within the parameters required.

B. Experimental Setup

The experiments were performed in personal computers with the following configurations: Intel Core i7 processor at 2.9 GHz, 16 GB RAM, NVIDIA GeForce GTX 1050 TI GPU with Ubuntu 16.04.

Convolutional Neural Network (CNN) of human pose estimation, starts the process for the initialization of our approach by calculating human body measures only using images from body perspectives.

Framework [29], supported the implementation of Human Pose Estimation Convolutional Neural Network, already trained to detect a human body in digital images.

After the body detection by Human Pose Estimation were applied classifiers methods in the output of this CNN to improve the body region segmentation in digital images. The classifiers used were the Bayesian, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Trees (DT), Adaptive Boosting (ADAB), Random Forest (RF), Multilayer Perceptron (MLP) and Expectation-Maximization (EM). After this, the merge method connects the four body images and measures the fist, forearm, breastplate, waist, relaxed biceps, thigh, and calf, using images obtained in the exam described in the Dataset subsection.

C. Discussion

For validation of the body measures obtained by our approach, it was necessary to calculate for each exam the mean square error related to the measurements acquired by the nutritionist of the fist, forearm, breastplate, waist, relaxed biceps, thigh, and calf. The developed system returns different measures for each classifier because the body circumferences depend on the result of the segmentation stage. So the more accurate is the step of body segmentation in digital images, the smaller is the difference between the measures acquired by the nutritionist and measures calculated by our approach.

The Table I presents the results of the mean square error about the measures of 7 body circumferences, in centimeters, calculated for each classifier. For the general analysis, it is better if smaller is the divergence between the result found in the system relative to the obtained by the specialist.

For each body circumference presented in Table I, the Bayesian approach showed the smallest error in the Pectoral and Thigh measures. KNN, is the best only for Fist measure. AdaB, was the best in Calf. EM, was the best in measures for the forearm, waist and biceps relaxed.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

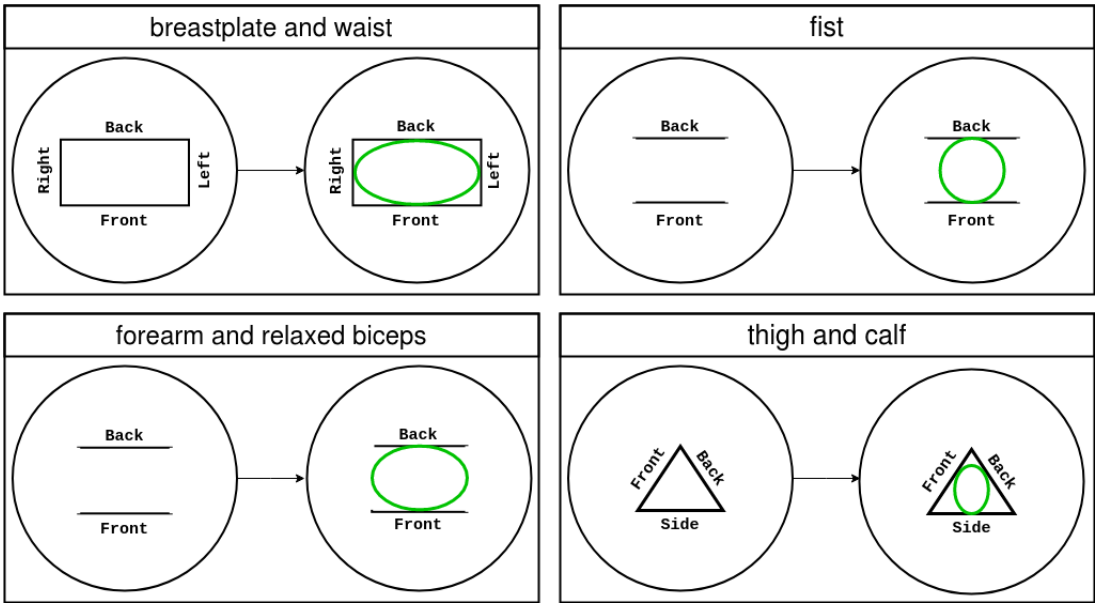

Fig. 2. Adopted approximations to the geometric shape of the human body.

TABLE I
MEAN SQUARED ERROR (MSE) USED FOR EVALUATION OF BODY MEASUREMENTS.

Method	Fist	Forearm	Breastplate	Waist	Relaxed Biceps	Thigh	Calf
Bayes	1.248±0.687	1.588±3.213	**3.806±2.358**	2.509±1.440	6.593±9.812	**2.948±2.753**	4.817±2.993
KNN	**0.641±0.414**	1.241±0.892	4.257±2.212	2.020±1.364	3.791±4.977	3.316±2.919	4.749±3.173
SVM	0.656±0.443	1.096±0.820	4.195±2.424	2.088±1.364	3.900±5.430	3.097±2.478	5.246±3.177
DT	2.752±10.095	2.193±5.43	7.766±17.642	4.823±14.524	4.431±5.892	4.027±7.088	4.495±3.959
ADAB	2.719±10.101	2.145±5.444	7.462±17.711	4.802±14.529	4.343±5.849	3.569±6.602	**4.440±3.662**
RT	2.638±10.117	2.462±6.228	4.191±2.273	2.263±1.752	4.414±5.702	4.474±5.361	4.662±2.905
MLP	0.815±0.524	1.081±0.892	4.016±2.274	2.206±1.340	3.844/-5.510	3.320±2.441	4.982±3.256
EM	0.744±0.363	**1.037±0.863**	3.971±2.309	**1.836±1.231**	**3.381±4.704**	3.744±3.734	4.606±3.412

The DT, ADAB, and RT methods presented high standard deviation values for calculating the average error in the measurements of the Fist, Forearm, Breastplate, Waist. So these methods suffered failures in the segmentation stage in some images, making the measures results far from the exam value obtained by the specialist. These failures occurred because the classification method was unable to smooth the segmentation curve of the human body, which later hindered the calculation of the approximation of the circumference that was measured.

The approach with EM was the best for more times because, in 3 different measures, it presented a smaller divergence with the result of the specialist. Also, in the other measures in which it did not obtain the best result, it showed equivalence to the best with mean squared error always bellow 4.606 ± 3.412 cm. Also, if the same test is applied several times in the proposed system, the exact measurement is always obtained. In contrast, the same nutritionist generally gets different values each time performed the measurement.

V. CONCLUSIONS AND FUTURE WORKS

Our work proposes a new approach to predict body measurements using digital image processing, dense human pose estimation - CNN and machine learning techniques to perform segmentation and, finally, make measurements.

Our contributions focus on predicting human body measures in 2D images and create a new concept for body segmentation using outputs features of Dense Pose Estimation - CNN as the input of classifiers that perform body segmentation.

The approach with Dense Human Pose Estimation and Expectation-Maximization reached the best results, with mean squared error (MSE) always bellow 4.606 ± 3.412 cm when compared with specialist measures.

In future works, we intend to develop methods to calculate body fat percentage from 2D images using CNNs and integrate this method to mobile and web applications.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

ACKNOWLEDGMENT

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. Also Pedro Pedrosa Rebouças Filho acknowledges the sponsorship from the Brazilian National Council for Research and Development (CNPq) via Grants Nos. 431709/2018-1 and 311973/2018-3.

The authors would like to thank The Ceará State Foundation for the Support of Scientific and Technological Development (FUNCAP) for the financial support (6945087/2019).

REFERENCES

[1] *The universal declaration of human rights (Resolution 217 A)*, General Assembly of the United Nations, December 1948. [Online]. Available: https://www.un.org/en/universal-declaration-human-rights/

[2] *Policy brief: The double burden of malnutrition (WHO/NMH/NHD/17.2)*, World Health Organization, 2017. [Online]. Available: https://apps.who.int/iris/handle/10665/255414

[3] D. Bagchi and H. Preuss, *Obesity: Epidemiology, pathophysiology, and prevention*, 2nd ed. Taylor & Francis, 2012.

[4] World Health Organization, "Obesity: Preventing and managing the global epidemic," World Health Organization, Tech. Rep. 894, 2000.

[5] *Obesity and overweight*, World Health Organization, February 2018. [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight

[6] A. Keys, F. Fidanza, M. J. Karvonen, N. Kimura, and H. L. Taylor, "Indices of relative weight and obesity," *Journal of Chronic Diseases*, vol. 25, no. 6, pp. 329–343, 1972.

[7] B. Cogill, "Anthropometric indicators measurement guide," *Food and Nutrition Technical Assistance Project*, January 2003.

[8] J. K. Thompson and L. M. Schaefer, "Thomas f. cash: A multidimensional innovator in the measurement of body image; some lessons learned and some lessons for the future of the field," *Body Image*, vol. 31, pp. 198 – 203, 2019.

[9] Z. Ji, X. Qi, Y. Wang, G. Xu, P. Du, X. Wu, and Q. Wu, "Human body shape reconstruction from binary silhouette images," *Computer Aided Geometric Design*, vol. 71, pp. 231 – 243, 2019.

[10] O. Affuso, L. Pradhan, C. Zhang, S. Gao, H. W. Wiener, B. Gower, S. B. Heymsfield, and D. B. Allison, "A method for measuring human body composition using digital images," *PLOS ONE*, vol. 13, no. 11, pp. 1–13, November 2018.

[11] 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, October 2019.

[12] M. Carletti, M. Cristani, V. Cavedon, C. Milanese, C. Zancanaro, and A. Giachetti, "Analyzing body fat from depth images," in *2018 International Conference on 3D Vision (3DV)*, September 2018, pp. 418–425.

[13] C. Liu, X. Fan, Z. Guo, Z. Mo, E. I.-C. Chang, and Y. Xu, "Wound area measurement with 3d transformation and smartphone images," *BMC Bioinformatics*, vol. 20, no. 724, 2019.

[14] T. Xiaohui, P. Xiaoyu, L. Liwen, and X. Qing, "Automatic human body feature extraction and personal size measurement," *Journal of Visual Languages & Computing*, vol. 47, pp. 9 – 18, 2018.

[15] J. Boisvert, C. Shu, S. Wuhrer, and P. Xi, "Three-dimensional human shape inference from silhouettes: Reconstruction and validation," *Machine Vision and Applications*, vol. 24, no. 1, pp. 145–157, January 2013.

[16] C. R. Pereira, D. R. Pereira, G. H. Rosa, V. H. Albuquerque, S. A. Weber, C. Hook, and J. P. Papa, "Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification," *Artificial Intelligence in Medicine*, vol. 87, pp. 67 – 77, 2018.

[17] R. V. M. da Nóbrega, P. P. Rebouças Filho, M. B. Rodrigues, S. P. P. da Silva, C. M. J. M. Dourado Júnior, and V. H. C. de Albuquerque, "Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks," *Neural Computing and Applications*, 2018.

[18] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes," *CoRR*, vol. abs/1711.00199, 2017.

[19] I. Rish *et al.*, "An empirical study of the naive bayes classifier," in *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, vol. 3, no. 22, 2001, pp. 41–46.

[20] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[21] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. New York: John Wiley & Sons, 2012.

[22] V. N. Vapnik, *Statistical learning theory*. New York: Wiley-Interscience, 1998.

[23] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees*. CRC Press, 1984.

[24] H. Trevor, T. Robert, and F. Jerome, *The elements of statistical learning: Data mining, inference and prediction*, 2nd ed. New York: Springer-Verlag, 2001, ch. 9, pp. 371–406.

[25] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: A statistical view of boosting," *The Annals of Statistics*, vol. 28, no. 2, pp. 337–407, 2000.

[26] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[27] S. Haykin, *Neural networks: A comprehensive foundation*, 2nd ed. Pearson Prentice Hall, 2001.

[28] J. A. Bilmes, "A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," International Computer Science Institute and Computer Science Division, University of California at Berkeley, Tech. Rep. TR-97-021, April 1998.

[29] R. A. Güler, N. Neverova, and I. Kokkinos, "Densepose: Dense human pose estimation in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7297–7306.

[30] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Image Processing and Computer Vision*, vol. 27, no. 3, 1984.