Deep learning algorithm for human body type assessment from digital images

1st Antonio Ricardo A. Brasil
Department of Electrical Engineering
Universidade Federal do Espírito Santo
Vitória, ES, Brazil
anribrasil@gmail.com

2nd Fabian Tadeu do Amaral
Departament of Sports
Universidade Federal do Espírito Santo
Vitória, ES, Brazil

3rd Licia C. S. de Lima Oliveira
Department of Pharmacology
Universidade de Vila Velha
Vila Velha, ES, Brazil

4th Girlandia A. Brasil
Department of Pharmacology
Universidade de Vila Velha
Vitória, ES, Brazil

5th Klaus Fabian Côco
Department of Electrical Engineering
Universidade Federal do Espírito Santo
Vitória, ES, Brazil

6th Patrick Marques Ciarelli
Department of Electrical Engineering
Universidade Federal do Espírito Santo
Vitória, ES, Brazil

***Abstract*—This paper presents a novel convolutional neural network (CNN) based regression approach for automated somatotype estimation from digital images. Unlike previous studies that classify individuals into predefined somatotype categories or rely on manual anthropometric measurements, our method directly estimates continuous somatotype component scores, leveraging advanced image segmentation (PointRend) to represent body shapes from digital images accurately. The innovation lies in integrating a segmentation technique with regression-based CNNs, offering higher accuracy and accessibility compared to prior automated methods. We applied advanced image segmentation (PointRend) to a dataset of 46 human body images and trained CNN models to directly regress endomorphy, ectomorphy, and mesomorphy values. Hyperparameter optimization was performed systematically to ensure optimal performance. Our CNN-based regression approach achieved comparable accuracy to traditional anthropometric methods (MSE $<$ 0.16) while eliminating manual intervention. The best result for endomorph estimation (MSE = 0.15) surpasses previous image-based preprocessing methods (MSE = 0.44). The proposed method demonstrates significant potential for automated somatotype estimation, providing an accurate, accessible alternative to traditional measurement methods. This innovation facilitates wider applications in health, sports, and biomedical engineering contexts.**

***Index Terms*—convolutional neural network, deep learning, somatotype, regression**

I. Introduction

Somatotyping, a concept developed in the early 1940s by Sheldon et al. [1], categorizes individuals based on body shape and composition. The theory proposes that the human body is composed of three components: endomorphy, mesomorphy, and ectomorphy. The proportions of these components describe an individual's body type, with endomorphy relating to rounded and fat bodies, mesomorphy to muscular and athletic bodies, and ectomorphy to linear and fragile bodies [2].

Initially, the determination of somatotypes was using photographs and manual measurements of certain body parts from images [1]. This approach had limitations due to the technology of that time, such as costly and low-quality cameras, and it was prone to human error. Later, Heath and Carter improved this method by using specialized equipment like balances, adipometers, and tape measures for direct body measurements [2]. By applying mathematical formulas, the Heath-Carter method estimates an individual's somatotype for each component with values ranging from 0 to 12.

Currently, determining an individual's somatotype and identifying its predominant component is an important method for evaluating and enhancing athletic performance [3, 4, 5]. Studies have shown correlations between predominant somatotype components and specific physical abilities; for example, mesomorphic individuals often exhibit enhanced strength performance, while ectomorphic individuals demonstrate superior flexibility [6].

Recent works have aimed to use deep learning for somatotype classification. For instance, Brasil et al. [7] proposed an approach using digital image processing techniques to classify somatotypes from digital images. The statistical region merging (SRM), the Otsu method, and heuristic algorithms were employed to estimate measurements of various body parts. A genetic algorithm combined with a neural network identified the most significant measurements for estimating somatotype components, achieving a mean absolute error (MAE) of 0.12 for one component and classification accuracy of predominant somatotype of up to 92.86%.

Building upon this work, Brasil et al. [8] evaluated different approaches using the NOMO 3D dataset, which contains 456 3D models manually classified by specialists into endomorph, ectomorph, or mesomorph classes. Transfer learning was employed with VGG-16 and InceptionResNet architectures [9, 10], with the best results achieved using VGG-16, obtaining an accuracy of 65.70% on NOMO 3D images. When applied to the dataset used in [7], an accuracy rate of 53.22% was achieved. The authors concluded that further studies with better datasets were necessary to improve the results.

Other studies, such as Chiu et al. [11], have explored somatotype estimation from 3D scanning image data. Measurements collected from these 3D models were evaluated

using support vector machines (SVM) to estimate the Heath-Carter somatotype. Gender and shape features were used as inputs, resulting in somatotype component values with a mean error below 0.5. However, further studies with larger datasets are needed, and it is important to highlight that 3D scanning devices are expensive [12]. Recently, Yoon et al. [13] introduced a deep learning approach for the classification of the human body into somatotypes using 3D anthropometric data. The authors collected 3D body scans from 217 participants to extract anthropometric measurements. The best accuracy was 72% by using MobileNetV2 by classifying into the three somatotype components: endomorphy, mesomorphy, and ectomorphy. Besides that, these two works are limited to only classification instead of estimation of the somatotype components.

While prior studies focused on the classification of somatotypes based on predefined categories or required manual anthropometric measurements, our approach innovatively leverages advanced segmentation techniques (PointRend) combined with regression-based CNN models to directly predict continuous somatotype components from two-dimensional digital images. Prior literature primarily emphasizes either three-dimensional modeling, which requires expensive hardware (e.g., 3D scanners), or traditional anthropometric methods prone to human error and manual intervention [11, 13]. Moreover, existing studies using CNNs often address only classification [8, 13], while regression-based somatotype estimation from images remains under-explored. Our proposed methodology provides a new avenue by eliminating the manual measurement step, thus contributing significantly to accessibility, scalability, and potentially faster applications.

The goal of this study is to develop a methodology based on a deep learning model that can estimate an individual's somatotype from digital images, eliminating the need for manual measurements, anthropometric devices, and mathematical formulas. While prior work employed image processing techniques to extract measurements from images or used 3D models for somatotype classification [7, 8, 11, 13], our approach differs by using convolutional neural networks (CNNs) to automatically extract image features for estimating the values of somatotype components.

There are several advantages of using CNNs for estimating somatotypes instead of manual measurements due to their capability to capture spatial relationships of the input [14]. First, the time needed to collect body data is reduced by using only body images. This automation eliminates the need for non-portable anthropometric equipment, making the process more accessible [13]. Furthermore, the neural network's capabilities eliminate the necessity of a specialist with expertise in the Heath-Carter method, democratizing somatotype recognition. Additionally, CNNs can handle large amounts of data with high accuracy and consistency, reducing human error and variability in measurements. Overall, the use of CNNs makes somatotype estimation more accessible through digital images.

Our proposed method employs the PointRend neural network for image segmentation of the human body from the background [15]. The segmented image is then used as input to a CNN for estimating the values of the somatotype components. To the best of our knowledge, this is the first study that uses a CNN to estimate somatotype components from images by regression.

The contributions of this paper are:

1) We propose a CNN-based regression method that does not require manual anthropometric measurements or pre-established classification categories by directly estimating continuous somatotype components (endomorph, mesomorph, and ectomorph) from 2D digital images.
2) We integrate the PointRend segmentation algorithm to isolate body contours from the background, enabling robust feature extraction for somatotype estimation.
3) We conduct extensive experimentation across six different data scenarios (based on gender and image orientation) and analyze the model's generalization capabilities and potential limitations in a real-world dataset.

The remainder of this paper is organized as follows. Section II describes the dataset and presents the proposed CNN-based regression methodology for somatotype estimation. Section III reports the experimental results, and Section IV presents a discussion about the results. Section V concludes the paper and discusses potential directions for future work.

II. Methods

A. Dataset

In this paper, we used the bodybuilder image dataset from the previous work of Brasil et al. [7]. This dataset provides images from bodybuilders positioned in front of a white background and with artificial lighting. To maintain a controlled environment for the image capture process, the cameras were positioned at the same height for all volunteers.

The images were captured at the International Federation of Bodybuilding (IFBB) regional championships from 2014 to 2016 in the state of Espírito Santo, Brazil. Initially, the authors collected images of 56 individuals in two different moments (28 individuals from each moment). However, 10 individuals (9 from the first collection and 1 from the second one) were excluded because the presence of other bodybuilders in the background occluded the reference markers in at least one image. As a result, the used dataset is composed of 46 individuals (15 males and 31 females). The average age of the individuals was 29.30 $\pm$ 6.94 years.

In addition to the images, specialists used the Heath-Carter method to collect the measurements directly from the bodybuilder's body. The following information was available: gender, date of birth, stature (cm), body mass (kg), triceps skinfold (mm), supraspinal skinfold (mm), subscapular skinfold (mm), medial calf skinfold (mm), biepicondylar bone diameter of the humerus (cm), biepicondylar bone diameter of the femur (cm), flexed arm circumference (cm), and calf circumference (cm). The somatotype mean was 1.8 $\pm$ 0.65 for the endomorphs, 4.3 $\pm$ 2.20 for the mesomorphs, and 1.6 $\pm$ 0.79 for the ectomorphs. Table I presents the anthropometric characteristics of the dataset.

TABLE I
ANTHROPOMETRIC CHARACTERISTICS OF DATASET.

Variable	Mean
Age (years)	29.30 (± 6.94)
Height (cm)	167.80 (± 9.62)
Weight (kg)	65.25 (± 15.99)
Triceps skinfold (mm)	16.00 (± 7.21)
Suprailiac skinfold (mm)	15.00 (± 7.64)
Subscapular skinfold (mm)	13.50 (± 8.76)
Medial calf skinfold (mm)	17.00 (± 8.86)
Bi-epicondylar diameter of humerus (mm)	6.30 (± 0.65)
Bi-epicondylar diameter of femur (mm)	8.80 (± 0.86)
Flexed arm circumference at maximum contraction (cm)	28.75 (± 4.58)
Calf circumference (cm)	36.45 (± 5.66)
Endomorphy	2
Mesomorphy	33
Ectomorphy	4
Central	7

The most common method for representing somatotypes includes the use of a two-dimensional space known as a somatochart [2]. The origin of the axes of the somatochart is located in individuals with somatotypes with the three components of the same value 4-4-4. The approximately circular region around the axes of the somatochart represents the more biologically common somatotypes to be observed. The somatochart can be divided into regions to group individuals. The most commonly used division includes 3 classes (ecto, endo, and meso) or 4 classes (ecto, endo, meso, and center).

Figure 1 shows the somatopoints of the individuals in the dataset on the somatochart and the mean somatotype in red. Each axis refers to a component of the somatotype. Note that most individuals in the dataset are close to the vertical line, indicating mesomorphic individuals. The units on each axis are based on the Heath-Carter [16] anthropometric measurements, which categorize somatotype components on a scale from 1 to 7, with 1 indicating low expression of that component and 7 indicating high expression.

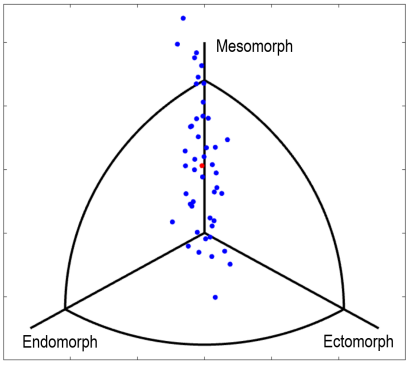
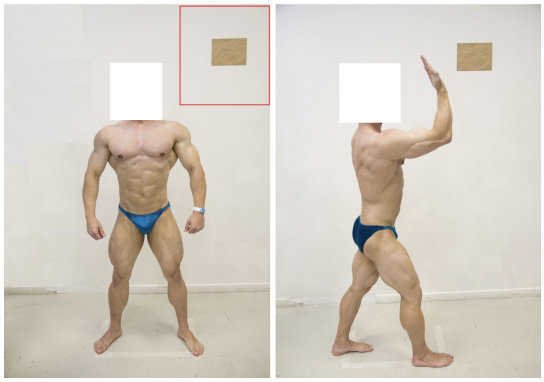


Fig. 1. Somatopoints of individuals in the somatochart [7].

The athlete's images were captured in JPG format, with two images taken of each athlete, one in each position, as shown in Figures 2a and 2b. These images provide all the necessary measurements that were automatically extracted from the images in [7] by using digital image processing techniques. The method presented in [7] requires individuals to have their legs and arms open for correct silhouette segmentation, as observed in both pictures. Additionally, tight-fitting bathing suits are necessary to prevent any alterations in the actual measurements of the body. To ensure that hair does not conflict with measurements, long hair must be tied back. A mark of known size was necessary to estimate the actual size of the measurements in the previous work, but for the current method, this mark is not relevant.



(a) Front image. (b) Side image.

Fig. 2. Examples of images from Brasil et al. [7]. The faces and tattoos were omitted to protect the privacy of the volunteers.

B. Deep learning model

Figure 3 presents the proposed method for estimating the somatotype of each component of the human body. Initially, the image with the subject is sent to a deep learning segmentation technique called PointRend [15]. This technique was used to segment the human bodies from the images and remove objects, such as tables, computers, or landmarks at the top of the image. Figure 4 shows samples from the dataset after applying the PointRend technique.

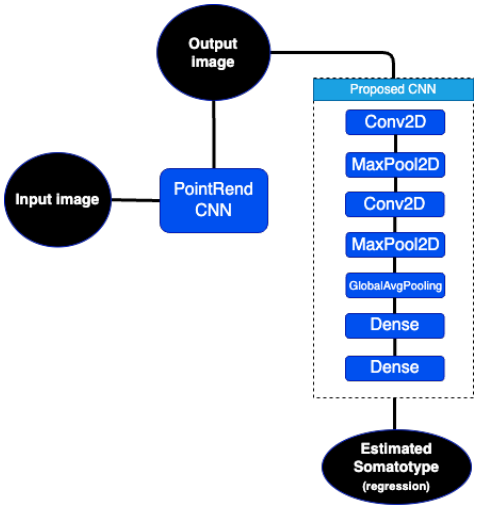


Fig. 3. Proposed method to estimate the somatotype.

During the development phase, simplified and complex models were evaluated (with more fully connected layers and

Fig. 4. Figures from the dataset after applying the PointRend technique. The faces and tattoos were omitted to protect the privacy of the volunteers.

convolutional layers, dense layers, max-pooling layers, kernel sizes, different combinations, and optimized models, such as transfer learning). However, initial experiments suggested that these variations performed worse than the chosen model.

The output image from PointRend is sent to the architecture of our model. The architecture consists of a model with 2 convolutional layers (CLs) followed by two dense layers. Each convolutional layer contains a 2D convolutional operation followed by a max-pooling 2D operation, and the last layer is followed by global average pooling. In the first two convolutional layers, 176 and 112 kernels were used, respectively. The size of the kernels was (3, 3) with ReLU activation (kernel sizes were selected during the prototyping phase to optimize performance on the validation set). The first dense layer (fully connected layer - FCL) has 48 neurons with ReLU activation, and the second dense layer contains just one neuron with linear activation for regression estimation [17]. For each component of the somatotype, a CNN was created and trained.

C. Evaluation of the model

The deep learning model performs a regression task to estimate the values of the endomorph, ectomorph, and mesomorph of each individual. To evaluate the quality of the results, we used two metrics: mean squared error (MSE) and the mean absolute error (MAE). Equations 1 and 2 show the equations of MSE and MAE, where n is the number of samples in the test dataset, $\hat{y}_i$ is the estimated value, and y_i is the target value:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2, \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|. \quad (2)$$

III. Results

The model was trained for 100 epochs with a batch size of 32. To train the neural network, we used the ADAM optimizer and MSE as a loss function. The ADAM parameters were as follows: learning rate of 0.000275 and epsilon of 1e-7. The hyperparameters (number of convolutional layers, number of dense layers, values of dense layers, values of convolutional filters, learning rate, and activation function) of the architecture were fine-tuned through the training set data with hyperparameter optimization. The evaluated values are presented in Table II, and the selected values are highlighted. One model was created for estimating each component value (endomorph, ectomorph, and mesomorph), resulting in three different models. The images were resized to 120 $\times$ 120 to reduce computational cost. We rescaled the pixel values (1/255) in the training, validation, and test steps (70% for training, 20% for validation, and 10% for testing).

TABLE II
Values evaluated with hyperparameter optimization.

	Values
Number of convolutional layers	1, **2**, 3, 4
Number of dense layers	**1**, 2, 3, 4, 5, 6, 7, 8
Values of dense layers	32, **48**, 64, 80, 96, 112, 128, 144, 160, 176, 192, 208, 224, 240, 256
Values of convolutional filters	16, 32, 48, 64, 80, 96, **112**, 128, 144, 160, **176**, 192, 208, 224, 240, 256
Learning rate	0.01, 0.001, **0.000275**
Activation function	ReLU, Sigmoid, **Linear**
Optimizer	**Adam**, SGD, RMSprop

Six different scenarios were evaluated, as shown in Table III, which contains results exclusively from the test set. The following four scenarios were used to evaluate gender separately: only using front images and only using side images. Two additional scenarios were made by mixing images of both genders: one using only front images and another using front and side images. The chosen scenarios were selected by a somatotype expert based on their visual analysis of the images. In the last scenario, the convolutional network was trained and evaluated with a dataset composed of both front and side images. Each image was processed individually, being at times a front image used as input, and at other times a side image. This allowed the model to learn from both perspectives without receiving them simultaneously. For each scenario, a convolutional neural network was built to separately train and evaluate the mesomorph, ectomorph, and endomorph components.

The results were evaluated based on the MAE and MSE metrics. The best values for each somatotype are highlighted in bold as follows: 0.15 of MSE and 0.38 of MAE for endomorph, considering the scenario with only men in front images as input, 0.76 of MSE and 0.82 of MAE for ectomorph, considering women in front images as input, and 0.77 of MSE and 0.83 of MAE for mesomorph, considering women in side images as input. The men and women in front images (5th scenario) and men and women in front and side images (6th scenario) were stratified, meaning that we maintained an equal proportion of the original dataset of women and men, with 70% for training, 20% for validation, and 10% for testing. For the cases with only men or only women, we used only one gender to train and test the models.

TABLE III
RESULTS OBTAINED WITH THE PROPOSED METHOD. THE BEST VALUES FOR EACH SOMATOTYPE COMPONENT ARE IN BOLD.

Scenario	Mesomorph		Ectomorph		Endomorph	
	MSE	MAE	MSE	MAE	MSE	MAE
Men in front images	1.66	1.29	0.95	0.97	**0.15**	**0.38**
Women in front images	0.91	0.87	**0.76**	**0.82**	0.62	0.78
Men in side images	2.10	1.44	0.84	0.91	0.30	0.54
Women in side images	**0.77**	**0.83**	0.85	0.86	0.59	0.76
Men and women in front images	3.08	1.67	0.81	0.85	0.54	0.65
Men and women in front and side images	1.24	1.01	0.90	0.89	0.53	0.71

TABLE IV
RESULTS USING PREPROCESSING FROM [7]. THE BEST VALUES FOR EACH SOMATOTYPE COMPONENT ARE IN BOLD.

Scenario	Mesomorph		Ectomorph		Endomorph	
	MSE	MAE	MSE	MAE	MSE	MAE
Men and women in front images	**1.29**	**1.13**	**1.17**	**1.08**	1.29	1.07
Men and women in front and side images	1.38	1.16	1.37	1.16	**1.19**	**1.03**

Figure 5 shows the model training and validation loss for the 6th scenario and for the endomorph component, where the loss metric is MSE and the model metric performance evaluation is MAE. Figure 5 shows that the predictions of the model, on average, differ from the actual values by approximately 0.90 units. The graph shows that the model achieved good precision in the validation set. When the MAE decreases, the model tends to have a more accurate and precise fit.

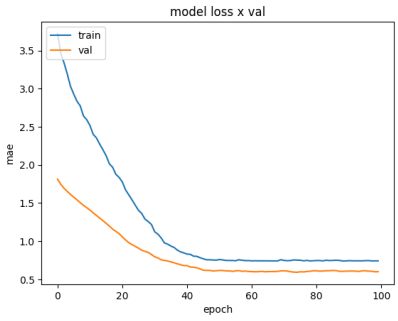


Fig. 5. Training and validation loss for the endomorph component of the 6th scenario.

We evaluated the proposed CNN with the approach of Brasil et al. [7], who preprocessed the images by using digital image processing techniques. These images were used as input for the proposed CNN instead of the images generated by the PointRend method. Table IV shows the results and (the best ones are highlighted in bold). We can see, for the same scenarios, the results of Table III are, in general, better than the results of Table IV.

IV. DISCUSSION

The results presented in Table III show an overall analysis considering six different scenarios. The scenario with the best results for the mesomorph was "women in side images", for the ectomorph was "women in front images", and for the endomorph was "men in front images". It is important to highlight that the images used as input for the convolutional neural network were subject to various external influences, such as varying lighting conditions and the presence of objects on the human body (watches). Additionally, the presence of hair on women's shoulders could have affected the results of PointRend and, consequently, the somatotype estimated by the CNN. Figure 6 shows some samples from the dataset containing the presence of hair on women's shoulders.

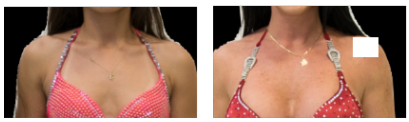


Fig. 6. Images from the dataset. Example of the presence of hair on women's shoulders. The faces and tattoos were omitted to protect the privacy of the volunteers.

By analyzing the MSE for the endomorph component across all six scenarios, we can observe a performance disparity in gender and type of image used. The best result was obtained in the scenario "men in front images", with an MSE of 0.15, whereas the highest MSE of 0.62 was observed in the scenario "women in front images", as we can see in Table III. These results suggest that specifically for endomorph estimation, models trained on male images achieved more accurate results in individualized scenarios (e.g., MSE of 0.15 for "men in front images" and 0.30 for "men in side images") than those trained on female images (MSE of 0.62 and 0.59 for front and side images of women, respectively). One possible explanation for the higher errors in female images can be the interference caused by hair covering the shoulders, which may have negatively affected the segmentation performance. As shown in Figure 6, this occlusion likely compromised the silhouette extraction and, consequently, the CNN's estimation of fat-related features such as the triceps and subscapular skinfolds (key indicators for endomorphy in the Heath-Carter method).

It is important to highlight, however, that this trend was not observed across all somatotype components. In fact, for the ectomorph and mesomorph components, female scenarios achieved the best results. For example, "women in front

images" had the lowest MSE for ectomorph, 0.76, and "women in side images" outperformed other scenarios in mesomorph estimation, with an MSE of 0.77. These results suggest that female body morphology may facilitate more accurate feature extraction for components associated with body linearity and muscularity, possibly due to more consistent visual cues in those traits. Therefore, while male images showed superior results for endomorph prediction, the performance advantage shifts depending on the somatotype component being estimated.

By analyzing other scenarios, when we compared the results for mesomorphs of the 5th scenario to those of the 6th scenario, we observed that the results improved when front and side images were used together, with an MSE of 3.08 in the 5th scenario reduced to 1.24 in the 6th scenario. On the other hand, for ectomorphs and endomorphs, the use of side images worsened or did not improve the results. This pattern suggests that the CNN performed better using the complementary information provided by both views when estimating mesomorph-related features, such as limb circumferences and bone diameters, which are more perceptible when multiple perspectives are available.

The calculation of mesomorphs depends on thigh and arm measurements, which may explain the high error values observed for the mesomorphs in each scenario in Table III. Ectomorph depends on the height and weight of the individual, and endomorph depends, for example, on the triceps skinfolds. Since the arm is clearly visible in each image, it may have positively affected the results, as the CNN likely identified relevant information in the images to perform this calculation.

The ectomorph component had an overall MAE and MSE lower than 1, with the best MSE and MAE values of 0.76 and 0.82, respectively. Unlike the previous components (endomorph and mesomorph), the best scenario for estimating the ectomorph was using images of women in front views. In general, ectomorphs tend to have a lofty forehead, narrow hips, chest, and abdomen, as well as slim legs and arms. This may explain the improved performance when using images of women in front views. When only women's images were used, the feature extraction of the convolutional layers seemed to work more effectively in identifying these relevant characteristics.

Table IV presents the performance of the neural network when using the preprocessed images from the previous method proposed by Brasil et al. [7], which relied on traditional digital image processing to extract body contours (as described in Section I). This comparison was used to identify if the PointRend segmentation improved the neural network estimation for somatotype components. Although this earlier approach provided a relevant baseline, the results across the scenarios "men and women in front images" and "men and women in front and side images" show higher MSE and MAE values for ectomorph and endomorph components when compared to our proposed method with PointRend segmentation (Table III). The results of mesomorph components in Table IV were lower than we achieved in Table III for the scenario "men and women in front images", however, the estimation errors of mesomorph components for "men and women in front and side images" were similar in both tables.

The lower performance of the baseline method can be explained since it has issues with image segmentation, introducing errors that can harm the model. Furthermore, the use of traditional digital image techniques is limited to images with controlled conditions of illumination and human poses and, at this point, PointRend is more promising to obtain better results.

In general, the somatotype components should be rounded to multiples of 0.1 or 0.5, depending on the application [18]. If any component resulting from the equations is zero or negative, the value of 0.1 is assigned to that component, as by definition, the components should be positive [18]. In this work, the values were not rounded to multiples of 0.1 or 0.5, which could affect the MAE and MSE of each scenario. This differs from the work of Brasil et al. [7], which used rounding. This adjustment can be considered for future work.

One issue with our method is the variability in accuracy across different scenarios and somatotype components. That is, some estimations are very accurate, while others are not so much, demonstrating a variability in the results. This can be observed in the values of MSE and MAE presented in Table III. For example, the mesomorph component showed a higher MSE of 2.10 in the "men in side images" scenario compared to 1.66 in the "men in front images" scenario.

This issue could likely be resolved by increasing the amount and diversity of data available to improve the training of the models, enhancing the generalization capability of the models. The small data set of 46 individuals, mostly mesomorphic, does not have too much diversity or a large number of examples that CNN can learn from. Further, the methods investigated here could likely be enhanced by dealing with externally caused image-quality issues, such as differences in lighting, some accessories like watches, or obstructions like hair over shoulders, as illustrated in Figure 6. These factors may have introduced noise and inconsistencies, impacting the segmentation accuracy of PointRend and, consequently, CNN's estimations.

Table V presents a comparison with other works. Compared to previous methodologies relying on manual anthropometric data or 3D measurements [7, 11], our regression-based CNN model uniquely automates the entire estimation process from simple digital images, making somatotype assessment more accessible and scalable. While previous methods required expensive and sophisticated equipment or expert intervention, our results demonstrate that reasonable precision can be achieved without any manual or specialized anthropometric intervention. Moreover, while classification-based CNN models, as presented in Table V, achieved moderate accuracy (72% maximum) in classifying the type of somatotype, they cannot quantitatively estimate somatotype components' values. This work proposes a model that estimates somatotype components' values, which could be highly valuable for personalized training programs and health monitoring.

The significant advantage of our approach is its potential integration into smartphone applications, wearable technologies, or low-cost diagnostic tools, democratizing access to reliable somatotype estimation beyond specialized clinical or sports environments. Additionally, the proposed method's automatic and objective nature can enhance data consistency, facilitating large-scale epidemiological studies and supporting evidence-based public health policies aimed at obesity prevention, physical activity promotion, and athletic talent identification. Our findings suggest practical implications not only for fields like sports science and physical therapy but also for biomedical engineering, physiotherapy, and rehabilitation research by allowing non-invasive, rapid, and scalable assessment of body composition.

V. Conclusion

This paper proposed a deep learning model for somatotype estimation using regression applied to digital images. The evaluation metrics (MSE and MAE) demonstrated that the CNN achieved consistent performance across different experimental scenarios, particularly for the endomorph component. The use of the PointRend segmentation technique contributed to improved silhouette extraction, enabling better regression results. In addition, the proposed approach simplified the estimation pipeline by reducing preprocessing steps and eliminating the need for manual parameter tuning. Overall, the method enhances automation, reduces human intervention, and offers a more accessible and scalable solution for somatotype analysis.

In summary, the CNN with the PointRend preprocessing technique presented in this paper shows interesting findings for somatotype estimation using only digital images. Other researchers can explore these findings and their potential impacts on fields such as sports science, personalized fitness, and healthcare. Furthermore, more research may improve the obtained results.

Future work can explore several approaches. First, the dataset used was small and had low diversity, with a large number of mesomorph individuals, which may have impacted the quality of the training of the convolutional neural network. Future work can explore a larger dataset, obtaining and expanding it to have greater diversity of individuals in different positions without a controlled environment, which can improve the model's generalizability and performance. Additionally, datasets incorporating other relevant factors, such as different body compositions and positions, could provide a more comprehensive estimation of somatotypes automatically from a CNN model. Furthermore, investigating deeper transfer learning or ensemble methods can enhance the model's accuracy and efficiency in estimating the somatotype.

Aknowledgments

The authors thank the Programa de Pós-Graduação em Engenharia Elétrica (PPGEE - UFES) and the financial support for the research from the project of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Espírito Santo (FAPES), number 598/2018.

References

[1] W. H. Sheldon, S. S. Stevens, and W. B. Tucker, "The varieties of human physique: An introduction to constitutional psychology," *Harper & Brothers Publishers*, 1940.

[2] J. L. Carter and B. H. Heath, *Somatotyping: development and applications*, Cambridge, Ed. New York, NY, USA: Cambridge University Press, 1990.

[3] W. G. Thorland, G. O. Johnson, T. G. Fagot, G. D. Tharp, and R. W. Hammer, "Body composition and somatotype characteristics of junior olympic athletes." *Medicine and Science in Sports and Exercise*, vol. 13, no. 5, pp. 332–338, 1980.

[4] M. Massidda, S. Toselli, P. Brasili, and C. M. Calò, "Somatotype of elite Italian gymnasts," *Collegium Antropologicum*, vol. 37, no. 3, pp. 853–857, 2013.

[5] B. Gutnik, A. Zuoza, I. Zuoziene, A. Alekrinskis, D. Nash, and S. Scherbina, "Body physique and dominant somatotype in elite and low-profile athletes with different specializations," *Medicina (Lithuania)*, vol. 51, no. 4, pp. 247–252, 2015.

[6] E. Terzi and A. Kalkavan, "To what extent do somatotype structures affect athletic performance in professional athletes?" *The Journal of Sports Medicine and Physical Fitness*, vol. 64, no. 7, pp. 650–660, 2024.

[7] A. R. A. Brasil, T. de Oliveira Gonçalves, J. W. B. Castro, E. C. Gonçalves, K. F. Côco, and P. M. Ciarelli, "Automatic identification of somatotype by digital images," in *Simpósio Brasileiro de Automação Inteligente*, 2021.

[8] A. R. A. Brasil, J. W. B. Castro, E. C. Gonçalves, G. A. Brasil, K. F. Coco, and P. M. Ciarelli, "Convolutional neural networks for somatotype classification of images from 3d human bodies," in *IX Congresso Latino-Americano de Engenharia Biomédica*, 2022.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[10] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[11] C.-Y. Chiu, R. Ciems, M. Thelwell, A. Bullas, and S. Choppin, "Estimating somatotype from a single-camera 3d body scanning system," *European Journal of Sport Science*, vol. 22, no. 8, pp. 1204–1210, 2022.

[12] E. Stark, O. Haffner, and E. Kučera, "Low-cost method for 3d body measurement based on photogrammetry using smartphone," *Electronics*, vol. 11, no. 7, p. 1048, 2022.

[13] J. Yoon, S.-Y. Lee, and J.-Y. Lee, "Ai somatotype system using 3d body images: Based on deep-learning and transfer learning," *Applied Sciences*, vol. 14, no. 6, p. 2608, 2024.

TABLE V
COMPARISON OF RELATED WORKS.

Work	Dataset	Methodology	Task	Results
Brasil et al. (2021) [7]	44 images (controlled conditions)	Preprocessing via digital image processing to extract anthropometric measurements; regression using handcrafted features.	Regression (manual measurement extraction)	Endomorph RMSE: 0.26
Brasil et al. (2022) [8]	456 3D models	Transfer learning using VGG-16 and InceptionResNet to classify somatotype categories.	Classification	Accuracy: $\sim$65.7% on NOMO 3D images
Chiu et al. (2022) [11]	3D scanning images	Extracts anthropometric measurements from 3D scans; SVM regression of somatotype components.	Regression	Mean error < 0.5 for somatotype components
Yoon et al. (2024) [13]	3D anthropometric data (body scans)	CNN (MobileNetV2) for classifying individuals into somatotype categories based on 3D measurements.	Classification	Maximum classification accuracy: 72%
This work	2D digital images from bodybuilder competitions (controlled conditions)	PointRend segmentation followed by a custom CNN (convolutional and dense layers) for regression.	Regression (continuous estimation)	Endomorph MSE: 0.15 (men in front images).

[14] O. Rainio, J. Lahti, M. Anttinen, O. Ettala, M. Seppänen, P. Boström, J. Kemppainen, and R. Klén, "New method of using a convolutional neural network for 2d intraprostatic tumor segmentation from pet images," *Research on biomedical engineering*, vol. 39, no. 4, pp. 905–913, 2023.

[15] A. Kirillov, Y. Wu, K. He, and R. Girshick, "Pointrend: Image segmentation as rendering," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9799–9808.

[16] B. H. Heath and J. E. L. Carter, "A modified somatotype method," *American Journal of Physical Anthropology*, vol. 27, no. 1, pp. 57–74, 7 1967. [Online]. Available: http:https://dx.doi.org/10.1002/ajpa.1330270108

[17] A. H. Magalhães, H. C. Yehia, and H. A. Magalhães, "Hyperspectral image synthesis from sparse rgb data: a comparative study combining linear regression, multilayer perceptron, and clustering," *Signal, Image and Video Processing*, vol. 18, no. 2, pp. 1625–1633, 2024.

[18] J. Carter, "The heath-carter anthropometric somatotype - instruction manual," San Diego State University, Tech. Rep., 2002.