

# Um Método Baseado em Inteligência Artificial para Estimar o Peso e a Altura a partir de Imagens Clínicas de Múltiplas Visões

Lucas D. Batista L.<sup>1</sup>, Ariel Soares Teles<sup>2</sup>

<sup>1</sup>Programa de Pós-Graduação em Biotecnologia (PPGBiotec)  
Universidade Federal do Delta do Parnaíba (UFDPar)

<sup>2</sup>Instituto Federal do Maranhão – Araisos, MA – Brasil

lucasbatista@ufdpar.edu.br, ariel.teles@ifma.edu.br

**Abstract.** *This study proposes a method based on deep learning and computer vision to automate height and weight measurement, reducing human-related errors in clinical assessments. Inaccurate measurements can lead to dosage and diagnostic errors, affecting the patient's health. A total of 19,172 clinical images from multiple views (posterior, anterior, and lateral) of 4,793 individual evaluations were used. The DeepLabV3 model extracted silhouettes, which were then used to train the WideResNet and Scale Equivariant WideResNet architectures. The approach combines multi-view silhouettes, specialized architectures for scale invariance, and regularizers that consider the relationship between weight and height. The method achieved a Mean Absolute Error of 3.7 kg and 3.6 cm for weight and height, respectively, based on 3,834 images or 958 evaluations in the test data. The best results were obtained where the Mean Squared Error (MSE) was the lowest. The results demonstrate the effectiveness of this approach in obtaining precise biometric measurements, highlighting its potential to enhance clinical accuracy and support patient health monitoring.*

**Resumo.** *Este estudo propõe um método baseado em aprendizado profundo e visão computacional para automatizar a medição de altura e peso, reduzindo erros inerentes aos humanos em avaliações clínicas. Medidas imprecisas podem causar erros na dosagem e diagnósticos, afetando a saúde do paciente. Foram utilizadas 19.172 imagens clínicas de múltiplas visões (posterior, anterior e laterais) de 4.793 avaliações individuais. O modelo DeepLabV3 extraiu silhuetas, posteriormente usadas para treinar as arquiteturas WideResNet e Scale Equivariant WideResNet. A abordagem combina silhuetas de múltiplas visões, arquiteturas especializadas para invariância de escala e regularizadores que consideram a relação entre peso e altura. O método obteve um Erro Absoluto Médio de 3,7 kg e de 3,6 cm para peso e altura respectivamente, utilizando 3.834 imagens ou 958 avaliações para os dados de teste. Os melhores resultados foram obtidos onde o Erro Quadrático Médio (EQM) foi o menor. Os resultados demonstram a eficácia da abordagem na obtenção de medidas biométricas precisas, ressaltando seu potencial para aprimorar a precisão clínica e auxiliar no monitoramento da saúde de pacientes.*

## 1. Motivação

A biometria humana representa uma informação essencial na grande maioria das práticas médicas. O perfil corporal do paciente está associado a diversas condições médicas, desempenhando um papel importante no processo de investigação, diagnóstico e tratamento de doenças [McDonald et al. 2013]. Em particular, o peso<sup>1</sup> e a altura são medidas antropométricas comumente utilizadas como indicadores de saúde, sendo empregadas, por exemplo, para monitorar a condição corporal ou avaliar o desenvolvimento de bebês e crianças de acordo com as condições nutricionais [McDonald et al. 2013]. Essas medidas são rotineiramente utilizadas em cálculos essenciais, como a dosagem de medicamentos [Maskin et al. 2010]. Além disso, o aumento dessas variáveis também está relacionado à redução da longevidade [Wells et al. 2022].

A importância da metrologia humana vai além da área da saúde e abrange diversos domínios, como segurança e vigilância [Kumar et al. 2022], monitoramento do fluxo de pedestres [Deleu et al. 2021], ajustes de roupas em compras online [Xia et al. 2019], e condução autônoma [Schulte et al. 2022]. No entanto, quando essas medidas são coletadas manualmente, o resultado pode apresentar imprecisões devido a alguns fatores, como: a falta de experiência dos examinadores, ao uso de instrumentos inadequados e/ou à má gestão dos dados [Medhi et al. 2012]. Além disso, para prevenir a transmissão de doenças por contato direto ou indireto, é desejável limitar a interação entre o profissional de saúde e o paciente. Consequentemente, há um interesse crescente em técnicas de medição não invasivas, que possam ser realizadas de forma rápida e sem a necessidade de palpação ou toque. A aplicação desses métodos é de fundamental importância, especialmente em ambientes restritos ou em situações em que o paciente e o profissional de saúde estão fisicamente separados, como na Telemedicina, e-Saúde e Telereabilitação [Cruz-Cunha 2016].

Embora abordagens com IA sejam comuns, muitas ainda não são práticas ou eficientes para uso clínico. No entanto, as abordagens baseadas em Inteligência Artificial (IA) identificadas na literatura frequentemente utilizam técnicas que não são práticas ou ineficientes para a prática clínica. Em um ambiente clínico, o processo de coleta de peso e altura deve ser rápido, adequado para qualquer tipo de corpo, utilizar ferramentas facilmente disponíveis e, o mais importante, garantir medições precisas. Essas características muitas vezes estão ausentes em soluções que: exigem medições antropométricas prévias para a estimativa, limitam a aplicabilidade demográfica da solução, ou dependem de scanners 3D e outras ferramentas não condizentes com o ambiente clínico [Guerra et al. 2021, Bigalke et al. 2021].

Este estudo propõe um método para estimar peso e altura a partir de imagens de vários ângulos. O método utiliza uma representação simplificada da imagem humana, especificamente a silhueta, e explora a relação entre peso e altura por meio de regularizadores durante o processo de treinamento [Jin et al. 2022]. Trata-se de um método baseado na utilização de técnicas de IA e Visão Computacional (VC) para estimar o peso e a altura de um indivíduo, fazendo uso também de arquiteturas modernas de Redes Neurais Convolucionais (RNCs), como as RNCs Equivariantes à Escala (do inglês *Scale Equivariant Convolutional Neural Networks – SE-CNNs*).

---

<sup>1</sup>O termo “peso” é utilizado neste artigo para se referir à “massa corporal”, sendo então compreendido não apenas por profissionais da área médica e nutricional, mas também pelo público em geral.

## **2. Problema de Pesquisa**

Os métodos tradicionais de coleta de medidas biométricas humanas apresentam limitações que podem comprometer tratamentos, entre outras práticas médicas. O principal problema reside na falta de praticidade e na suscetibilidade a erros durante a medição de peso e altura em ambientes clínicos, além da necessidade de palpação ou contato físico, exigidos pela maioria desses métodos. Medições imprecisas podem levar a diagnósticos errados e a tratamentos inadequados, resultando em complicações graves. Além disso, a variabilidade das características físicas dos pacientes, como sobrepeso ou condições de saúde específicas, pode comprometer a precisão das medições manuais, que muitas vezes não são adaptáveis a diferentes biotipos e cenários clínicos [Nyholm et al. 2022]. Portanto, é essencial desenvolver métodos de coleta de medidas biométricas humanas que sejam resistentes a erros e à variabilidade das características corporais, além de precisos e práticos. Isso possibilita a aplicação da medicina remota e personalizada de forma mais eficaz.

## **3. Objetivos**

### **3.1. Geral**

O objetivo geral deste estudo de mestrado foi desenvolver um método capaz de prever a altura e peso de um indivíduo dado um conjunto de imagens de múltiplas visões.

### **3.2. Específicos**

- Adaptar arquiteturas RNCs equivariantes à escala para o uso integrado com regularizadores baseados na relação entre peso e altura, melhorando o desempenho nas estimativas;
- Explorar e avaliar técnicas de segmentação e fusão de imagens através de modelos pré-treinados de aprendizado profundo e concatenação de imagens de múltiplas visões;
- Desenvolver um fluxo constituído de etapas de processamento de imagens de múltiplas visões e estimativa de peso e altura;
- Realizar experimentos para verificar o desempenho da abordagem em comparação com métodos existentes na literatura.

## **4. Contribuições do Trabalho**

As contribuições deste estudo para o estado da arte incluem principalmente:

- Uma variedade de comparações para desenvolver o método proposto, como um estudo sobre a organização de imagens de múltiplas visões e um estudo comparativo de arquiteturas de redes neurais;
- Um método inovador para estimar peso e altura que supera os métodos semelhantes encontrados na literatura;
- Percepções sobre as aplicações práticas das arquiteturas modernas de redes neurais, que ainda são pouco exploradas em situações reais ou clínicas.

## 5. Resultados e Discussão

### 5.1. Fusão Precoce das Imagens

As imagens deste estudo foram fornecidas pela empresa canadense *Solutions Biotonix Inc.*<sup>2</sup>. Um estudo para identificar a melhor abordagem de concatenação destas imagens foi conduzido. O primeiro passo foi determinar o número ideal de imagens. Mais imagens reduziram a perda, usando o Erro Quadrático Médio (EQM) para ajuste de peso. Os resultados indicaram que o uso das quatro imagens proporcionou melhor desempenho. Na segunda fase do estudo, o objetivo foi investigar o melhor arranjo das imagens. Foram testados três arranjos: formato quadrático, vertical e horizontal. A organização quadrática apresentou melhor desempenho, embora a diferença não tenha sido significativa em comparação com os outros arranjos. Como resultado, as etapas subsequentes do desenvolvimento do método utilizaram todas as quatro imagens disponíveis por indivíduo, organizadas no formato quadrático.

### 5.2. Desempenho dos Modelos

Cada arquitetura foi treinada durante 50 épocas e a reamostragem dos dados seguiu uma proporção 80/20. Os modelos foram comparados em termos de EQM, onde o menor valor define o melhor desempenho. As arquiteturas utilizadas são baseadas na WideResnet, proposta por [Zagoruyko and Komodakis 2016]. Dessa forma, esta arquitetura será referida como *WRN-16-8* e a sua versão equivariante à escala como *SE-WRN-16-8*.

Os modelos foram avaliados com base em métricas como Erro Absoluto Médio (EAM), EQM e coeficiente de determinação  $R^2$ . Os resultados indicaram que o modelo *SE-WRN-16-8*, com 16 camadas convolucionais e um fator de aumento (do inglês, *widening factor*) de 8, se destacou, apresentando um EAM de 3,7 kg para estimativas de peso e 3,6 cm para estimativas de altura, além de um  $R^2$  de 92% e 82%, respectivamente, demonstrando uma forte correlação entre as previsões e os valores reais. Os resultados superam estudos semelhantes na literatura, como em [Altinigne et al. 2020] e [Han et al. 2020], com detalhes disponíveis em [Lima et al. 2024]. Além disso, este resultado é semelhante ao erro apresentado por medições manuais de peso e altura, que podem ultrapassar 2 kg e 1,7 cm, respectivamente [Sinaga et al. 2024].

O melhor desempenho apresentado pelo modelo *SE-WRN-16-8* pode ser creditado a sua habilidade em manter a invariância em relação à escala das imagens. Essa característica permitiu que o modelo capturasse características complexas nas imagens de silhuetas, especialmente quando as silhuetas não estão perfeitamente definidas. Adicionalmente, a integração de regularizadores que exploram a relação entre peso e altura durante o treinamento do algoritmo de rede neural profunda, contribuiu significativamente para a redução dos erros, permitindo que o modelo convergisse para estimativas mais precisas.

Estudos estatísticos foram realizados para avaliar a significância das previsões do modelo. O Teste t de Student indicou que não havia diferenças significativas entre as distribuições dos valores estimados e os valores reais, com p-valores de 0,298 para peso e 0,647 para altura. Isso reforça a confiabilidade dos resultados apresentados. As representações gráficas, como gráficos de *Bland-Altman*, mostraram uma boa

---

<sup>2</sup><https://biotonix.com/en/>

concordância entre os valores previstos e reais, com a maioria dos pontos situados dentro dos limites de concordância de 95%. Esses resultados não apenas validam a eficácia do nosso método, mas também sugerem uma nova abordagem promissora para a estimativa de biometria em contextos clínicos, com o potencial de minimizar erros associados a métodos tradicionais de medição.

## 6. Conclusão

Esta pesquisa de mestrado apresenta um método promissor para a estimativa de peso e altura a partir de imagens clínicas, utilizando uma abordagem que combina técnicas modernas de aprendizado profundo, visão computacional e regularizadores. Esta abordagem, ainda não utilizada por estudos identificados na literatura, apresenta o potencial de reduzir o tempo dedicado às medições tradicionais, e aumentar a disponibilidade de profissionais para atividades de maior complexidade, promovendo ganhos indiretos na eficiência e na qualidade do cuidado. O uso de imagens de silhuetas de múltiplas visões, aliado à *Scale-Equivariant WideResNet*, demonstrou a capacidade do método em fornecer estimativas precisas e confiáveis, com resultados que superam métodos anteriores da literatura. A capacidade de estimar essas medidas biométricas humanas de forma automática não apenas reduz a dependência de medições manuais tradicionais, que podem estar sujeitas a erros, mas também alinha-se com a prática médica de e-Saúde, Telemedicina e Telereabilitação.

Futuros estudos podem ampliar a robustez do modelo, incluindo dados como coordenadas de pose para melhorar as estimativas. Essa integração poderia levar a melhorias significativas na precisão das previsões. Além disso, a realização de experimentos com exemplos adversariais poderia auxiliar na identificação de vulnerabilidades e na proteção do modelo contra variações inesperadas nas imagens de entrada, resultando em um sistema ainda mais robusto e confiável para a prática clínica.

## Referências

- Altinigne, C. Y., Thanou, D., and Achanta, R. (2020). Height and weight estimation from unconstrained images. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2298–2302. IEEE.
- Bigalke, A., Hansen, L., Diesel, J., and Heinrich, M. P. (2021). Seeing under the cover with a 3d u-net: point cloud-based weight estimation of covered patients. *International Journal of Computer Assisted Radiology and Surgery*, 16(12):2079–2087.
- Cruz-Cunha, M. M. (2016). *Encyclopedia of e-health and telemedicine*. IGI global.
- Deleu, E., Elez, S., Gadodia, A., Macvaugh, K., and Zhao, G. (2021). Using deep learning for urban pedestrian counting. In *2021 IEEE MIT Undergraduate Research Technology Conference (URTC)*, pages 1–5. IEEE.
- Guerra, R. S., Sousa-Santos, A. R., Sousa, A. S., Valdiviesso, R., Afonso, C., Moreira, P., Padrão, P., Borges, N., Santos, A., Ferro, G., et al. (2021). Prediction equations for estimating body weight in older adults. *Journal of Human Nutrition and Dietetics*, 34(5):841–848.
- Han, D., Zhang, J., and Shan, S. (2020). Leveraging auxiliary tasks for height and weight estimation by multi task learning. In *2020 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–7. IEEE.

- Jin, Z., Huang, J., Wang, W., Xiong, A., and Tan, X. (2022). Estimating human weight from a single image. *IEEE Transactions on Multimedia*.
- Kumar, A., Deeksha, K., Pooja, G. S., Reddy, T. T., and Reddy, T. A. (2022). Estimate height weight and body mass index from face image using machine learning. In *2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT)*, pages 1–5. IEEE.
- Lima, L. D. B., Teixeira, S., Bordalo, V., Lacoste, S., Guimond, S., Sousa, D. L., Pinheiro, D. N., Moreira, R., and Teles, A. S. (2024). A scale-equivariant cnn-based method for estimating human weight and height from multi-view clinic silhouette images. *Expert Systems with Applications*, 256:124879.
- Maskin, L., Attie, S., Setten, M., Rodriguez, P., Bonelli, I., Stryjewski, M., and Valentini, R. (2010). Accuracy of weight and height estimation in an intensive care unit. *Anaesthesia and intensive care*, 38(5):930–934.
- McDonald, C. M., Olofin, I., Flaxman, S., Fawzi, W. W., Spiegelman, D., Caulfield, L. E., Black, R. E., Ezzati, M., Danaei, G., and Study, N. I. M. (2013). The effect of multiple anthropometric deficits on child mortality: meta-analysis of individual data in 10 prospective studies from developing countries. *The American journal of clinical nutrition*, 97(4):896–901.
- Medhi, I., Jain, M., Tewari, A., Bhavsar, M., Matheke-Fischer, M., and Cutrell, E. (2012). Combating rural child malnutrition through inexpensive mobile phones. In *Proceedings of the 7th Nordic conference on human-computer interaction: making sense through design*, pages 635–644.
- Nyholm, B., Obling, L., Hassager, C., Grand, J., Møller, J., Othman, M., Kondziella, D., and Kjaergaard, J. (2022). Superior reproducibility and repeatability in automated quantitative pupillometry compared to standard manual assessment, and quantitative pupillary response parameters present high reliability in critically ill cardiac patients. *PLoS One*, 17(7):e0272303.
- Schulte, J., Kocherovsky, M., Paul, N., Pleune, M., and Chung, C.-J. (2022). Autonomous human-vehicle leader-follower control using deep-learning-driven gesture recognition. *Vehicles*, 4(1):243–258.
- Sinaga, H. T., Siregar, M., and Sitanggang, B. (2024). Anthropometric training impact on chws’ skills in child height measurement in indonesia. *Journal of Hunan University Natural Sciences*, 51(3).
- Wells, M., Goldstein, L. N., and Cattermole, G. (2022). Development and validation of a length-and habitus-based method of total body weight estimation in adults. *The American Journal of Emergency Medicine*, 53:44–53.
- Xia, L., Yang, J., Han, T., Xu, H., Yang, Q., Zhao, Y., and Wang, Y. (2019). A mobilized automatic human body measure system using neural network. *Multimedia Tools and Applications*, 78:11291–11311.
- Zagoruyko, S. and Komodakis, N. (2016). Wide residual networks. *arXiv preprint arXiv:1605.07146*.