



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA
DOUTORADO EM ENGENHARIA DE TELEINFORMÁTICA

NAVAR DE MEDEIROS MENDONÇA E NASCIMENTO

**SHAPEDNET: A REGRESSION CONVOLUTIONAL NEURAL NETWORK
APPROACH FOR BODY COMPOSITION ASSESSMENT BASED ON
INTERPRETABLE DBF-SCAN TECHNIQUES**

FORTALEZA

2023

NAVAR DE MEDEIROS MENDONÇA E NASCIMENTO

SHAPEDNET: A REGRESSION CONVOLUTIONAL NEURAL NETWORK APPROACH
FOR BODY COMPOSITION ASSESSMENT BASED ON INTERPRETABLE DBF-SCAN
TECHNIQUES

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia de Teleinformática. Área de Concentração: Engenharia de Teleinformática.

Orientador: Prof. Dr. Pedro Pedrosa Rebouças Filho.

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- N196s Nascimento, Navar de Medeiros Mendonça e.
Shapednet: A Regression Convolutional Neural Network Approach for Body Composition Assessment Based on Interpretable DBF-SCAN Techniques / Navar de Medeiros Mendonça e Nascimento. – 2023.
123 f. : il. color.
- Tese (doutorado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia de Teleinformática, Fortaleza, 2023.
Orientação: Prof. Dr. Pedro Pedrosa Rebouças Filho..
1. Body Composition Assessment. 2. Body Fat Percentage. 3. DXA. 4. Multi-task Learning. 5. Deep Learning. I. Título.

CDD 621.38

NAVAR DE MEDEIROS MENDONÇA E NASCIMENTO

SHAPEDNET: A REGRESSION CONVOLUTIONAL NEURAL NETWORK APPROACH
FOR BODY COMPOSITION ASSESSMENT BASED ON INTERPRETABLE DBF-SCAN
TECHNIQUES

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de doutor em Engenharia de Teleinformática. Área de Concentração: Engenharia de Teleinformática.

Aprovada em: 07 de Julho de 2023

BANCA EXAMINADORA

Prof. Dr. Pedro Pedrosa Rebouças
Filho (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. ATSLANDS REGO DA ROCHA
Universidade Federal do Ceará (UFC)

Prof. Dr. JOAO PAULO PORDEUS GOMES
Universidade Federal do Ceará (UFC)

Prof. Dr. CARLOS ANTONIO BRUNO DA
SILVA
Universidade de Fortaleza (UNIFOR)

Prof. Dr. JOAO PAULO PAPA
Universidade do Estado de São Paulo (UNESP)

In loving memory of my stepfather, Nilo Soares,
Your unwavering support and boundless pride
fueled my journey. Though you're gone, your
spirit guides every word.
With eternal gratitude,

ACKNOWLEDGEMENTS

I wish to extend my sincere gratitude to the individuals whose unwavering support and encouragement have been pivotal in the completion of this doctoral thesis

To my esteemed mother, Jacqueline Medeiros Soares, whose enduring love and sacrifices have profoundly influenced my personal and academic journey. Your unwavering faith in my abilities has been a beacon of strength and inspiration throughout this endeavor.

In remembrance of my stepfather, Nilo, whose steadfast belief in my potential and pride in my achievements continue to resonate deeply within me. Though his physical presence is no longer with us, his spirit and encouragement remain ever-present in the culmination of this academic pursuit.

To my father, José Haroldo do Nascimento for his unwavering support, invaluable guidance, and insightful counsel. His wisdom and encouragement have been instrumental in shaping my academic aspirations, and for this, I am deeply grateful.

I am indebted to my esteemed advisor, Prof. Pedro Pedrosa Rebouças Filho, for their scholarly guidance, expertise, and unwavering commitment to my intellectual development. Their mentorship has been indispensable, and I am honored to have had the opportunity to learn under their tutelage.

To my beloved partner, Nádila Diniz Nascimento, whose unwavering patience, understanding, and encouragement have sustained me through the rigors of doctoral research. Their unwavering support has been a source of strength and inspiration, for which I am profoundly grateful.

To my esteemed colleagues in the LAPISCO research lab, Leandro Marinho, Pedro Feijó, Suane Pires, and Roger Moura, whose collaborative spirit, scholarly discussions, and shared passion for knowledge have enriched my academic journey. Their insights and camaraderie have been invaluable, and I am privileged to have been a part of such an esteemed community. My deepest gratitude extends to other members who have helped throughout this journey.

Special appreciation is extended to my friends at IFCE Jaguaribe, namely Daniel Cavalcante, Antônio Augusto, Rodrigo Fernandes, and Luis Fernandes, for their unwavering friendship, support, and encouragement. They have made our odyssey at Jaguaribe easier. Their companionship during our journey has provided me with weekly encouragement, especially during our travels. Their presence has enriched my doctoral experience immeasurably, and for

this, I am deeply appreciative.

I want to extend my heartfelt gratitude to my surf buddies at Caponga, George Felipe, Danrley Soares, Jhonatan Dantas, and Pedro Henrique. Your friendship and shared love for surfing have been a vital source of rejuvenation during times when I needed to reconnect with myself. Our moments together in the waves have provided me with the clarity and focus necessary to navigate the challenges of doctoral research. Thank you for your unwavering support and camaraderie.

To all those who have contributed to this academic endeavor, whether through direct assistance or silent encouragement, I offer my heartfelt gratitude. Your support has been integral to the realization of this milestone, and I am deeply thankful for your unwavering presence throughout this journey.

"It's not who we are underneath, but what we do that defines us." (Rachel Dawes in the Dark Knight Series)

ABSTRACT

A estimativa precisa da composição corporal é essencial para avaliar a saúde e o bem-estar. Esta tese apresenta uma nova técnica para avaliação da composição corporal, o ShapedNet. Este método é uma rede neural profunda de aprendizado multitarefa que simultaneamente estima o Percentual de Gordura Corporal (PGC), identifica e localiza indivíduos usando uma única foto. Para validá-lo, foram feitas comparações com o método de referência padrão, a Absortometria de Raio-X de Dupla Energia (DXA). O processo de validação envolveu um conjunto de dados de 1273 adultos, variando em idade (18 a 65 anos), sexo (54,59% mulheres) e níveis de PGC (DXA 9,3% - 57,6%). A avaliação incluiu medidas de PGC usando DXA e duas seções fotográficas. Além disso, criou-se um conjunto de dados de acompanhamento, no que foram repetidos todos os exames e fotos com os participantes após três meses. Esta abordagem superou os melhores resultados relatados para estimativa de gordura corporal baseada em visão computacional usando fotos, em uma margem de 1,19 a 1,89 pontos. O ShapedNet obteve um Erro Percentual Absoluto Médio (MAPE) de 4,91% e um Erro Absoluto Médio (MAE) de 1,42. Foi avaliado o uso de abordagens baseadas em gênero e não baseadas em gênero, sendo que a última superou a primeira em uma margem de 1,07 a 2,12 pontos. O ShapedNet estima a PGC com 95% de confiança, dentro de uma margem de erro de 4,01% a 5,81%, ou uma diferença de 1,18 a 1,66 em relação às estimativas do DXA. Coeficientes de correlação acima de 0,9 foram observados entre o ShapedNet e o DXA em todos os experimentos realizados. Para aprimorar a interpretabilidade do método proposto, foi criada a técnica de visualização DBF-SCAN. Esta fornece visualizações valiosas sobre a distribuição e composição da gordura corporal, destacando a importância relativa de diferentes regiões do corpo na determinação da gordura corporal. Essa visualização orientada ao paciente complementa a estimativa quantitativa fornecida pelo ShapedNet. Esta pesquisa apresenta um método para a avaliação da composição corporal, avaliado por análises de erro, correlação e técnicas de visualização. O método ShapedNet está alinhado com o método de dois compartimentos na teoria da avaliação da composição corporal e oferece uma abordagem indireta inovadora. Além da estimativa da composição corporal, os avanços apresentados nesta tese contribuem para o desenvolvimento de modelos de Inteligência Artificial Explicável (XAI) e ampliam as aplicações do aprendizado multitarefa em diversos domínios.

Palavras-chave: avaliação da composição corporal; percentual de gordura corporal; dxa; aprendizado multitarefa; aprendizado profundo; shapednet; dbf-scan.

ABSTRACT

Accurate body composition estimation is essential for assessing health and well-being. This thesis introduces a novel techniques for improving body composition assessment, the ShapedNet. This method is a multi-task learning deep neural network that simultaneously estimates Body Fat Percentage (BFP), identifies, and localizes individuals using a single photo. To validate its accuracy, rigorous comparisons were made against the gold standard reference method, Dual-Energy X-ray Absorptiometry (DXA). The validation process involved a dataset of 1273 healthy adults, ranging in ages (18-65 years old), sexes (54.59% women), and BFP levels (DXA 9.3% - 57.6%). The evaluation included BFP measurements using DXA, two photo sections, and other clinical exams. Additionally, a follow-up dataset was created by repeating all exams and photos with the participants three months later. Our approach outperformed the best reported results for computer vision-based body fat estimation using photos by a margin of 1.19 to 1.89 points. ShapedNet achieved a Mean Absolute Percentage Error (MAPE) of 4.91% and Mean Absolute Error (MAE) of 1.42. We also evaluated using both gender-based and non-gender-based approaches, the later outperformed the gender-based by a margin of 1.07 to 2.12 points. ShapedNet estimates BFP with 95% confidence, within an error margin of 4.01% to 5.81%, or a difference of 1.18 to 1.66 in DXA estimates. High correlation coefficients greater than 0.9 were observed between ShapedNet and DXA in all experiments conducted. To enhance the interpretability of the proposed method, we design the DBF-SCAN visualization technique. DBF-SCAN provides valuable visual insights into the distribution and composition of body fat, highlighting the relative importance of different body regions in determining body fatness. This subject-oriented visualization complements the quantitative estimation provided by ShapedNet. This research introduces a robust and accurate method for body composition assessment, supported by rigorous validation, correlation analysis, and visualization techniques. The ShapedNet method aligns with the two-compartment method in body composition assessment theory and offers an innovative indirect approach. Beyond body composition estimation, the advancements presented in this thesis contribute to the development of Explainable AI (XAI) models, and broaden the applications of multi-task learning in various domains.

Palavras-chave: body composition assessment; body fat percentage; dxa; multi-task learning; deep learning; shapednet; dbf-scan.

LIST OF FIGURES

Figure 1 – Different types of body composition models: 1C, 2C, 3C and 4C.	28
Figure 2 – The 2-, 3- and 4-compartmental models of human body composition (left hand side), alongside the validation hierarchy (right hand side).	28
Figure 3 – Evaluation by dual-energy X-ray absorptiometry.	29
Figure 4 – Example architecture of a CNN.	37
Figure 5 – Illustration of the convolution operation. The input matrix is highlighted in blue, the filters of the convolutional layer are shown in red, and the resulting output volume from the convolution is depicted in green.	39
Figure 6 – Activation map generated by a convolutional layer with 2 filters.	40
Figure 7 – The ReLU activation function.	40
Figure 8 – Illustration of the effect of the pooling layer with the maximum operator. . .	41
Figure 9 – Dropout illustration. in (a), conventional flow of information during a forward pass. In (b), this flow is interrupted (set to zero, or Dropped out) in some neurons.	42
Figure 10 – Parameters of a Batch Norm layer.	43
Figure 11 – Calculations performed by Batch Norm layer.	43
Figure 12 – The architecture of the LeNet-5 Convolutional Neural Network.	45
Figure 13 – The effective receptive field of a stack of two 3×3 convolutions.	46
Figure 14 – Two illustrations of initial architecture: (a) Simple Inception module and (b) Inception module with dimensionality reduction.	47
Figure 15 – Here is an illustration of the architecture of the Inception module with asymmetric convolutions.	48
Figure 16 – An example of the architecture of a Residual Block.	50
Figure 17 – A generic illustration of the overall architecture when combining Residual Connections with the Inception Module.	50
Figure 18 – Illustration of the architecture of a Dense Block.	53
Figure 19 – Illustration of the NASNet modules.	54
Figure 20 – Illustration of a computational graph	55
Figure 21 – A general overview how ShapedNet works and what it outputs.	57
Figure 22 – The higher-level architecture of ShapedNet.	58
Figure 23 – The lower-level architecture of ShapedNet.	60

Figure 24 – Input photo in (a), ShapedNet output in (b) and output of DBF-SCAN in (c).	63
Figure 25 – Cuboid transformation to create the tensor $y_{m,n,c,k}^{(57)}$	65
Figure 26 – Visual representation of the compression made by the partial linearization of $\alpha_{\hat{Y}_{bf}}^{(58)}$	68
Figure 27 – Mapping region of input space, X and heatmap space, $L_{BF-SCAN}^r$	69
Figure 28 – Application of the heatmap matrix in the partial linearization of an input image.	70
Figure 29 – Dynamic Visualization of the evolution of the DBF-SCAN and heatmap matrix, $L_{BF-SCAN}^r$, during Training.	70
Figure 30 – Dataset distribution.	72
Figure 31 – Dataset DXA distribution by sex.	72
Figure 32 – Dataset attributes distribution regarding anthropometric measurements.	73
Figure 33 – Correlation matrices divided by sex.	75
Figure 34 – Relation weight in DXA vs BMI.	75
Figure 35 – Summary of all details about the database (BCDB23).	77
Figure 36 – Correlation Analysis between DXA and ShapedNet in Experiments T_1 and T_2 in Two Exam Sections fo Gender-based and Non-gender-based Approaches.	92
Figure 37 – Results of DBF-SCAN in a female subject (E04). In (a) it the ouput of ShapedNet, (b) the body fatness likelihood and (c) a reference grid for analyses.	96
Figure 38 – Results of DBF-SCAN in a male subject (E27). In (a) it the ouput of Shaped-Net and (b) the body fatness likelihood	97
Figure 39 – In-depth Analysis of DBF-SCAN Results for Subject E04.	99
Figure 40 – In-depth Analysis of DBF-SCAN Results for Subject E27.	100
Figure 41 – Analyses of the right and left side of the body in DBF-SCAN (Subject E27).	101
Figure 42 – Side by Side Comparison of Subject E04 and Her Heatmap Matrix $L_{BF-SCAN}^r$.	102
Figure 43 – Example of transfer learning procedure.	104
Figure 44 – Proposed methodology to incorporate Transfer Learning for Deep Learning Regression.	104

LIST OF TABLES

Table 1 – Dataset distribution in average, min, std, and max.	72
Table 2 – Overall results on test, validation, and train sets. The metrics are MAPE, MAE, and MSD, and the groups of analyses are Male, Female, and Non-gender. . .	83
Table 3 – Error Analysis with 95% Confidence Intervals: MAPE, MAE, and MSD. . .	83
Table 4 – Results of the Tukey’s HSD hyphotesis test. MD denotes Mean Differences and the values of p highlighted in bold represent the groups are significantly different at the 95% confidence level	83
Table 5 – Comparison of Errors MAE, MAPE and MSD for the first and second Sections in T1.	85
Table 6 – Results of the Tukey’s HSD hypothesis test in first and second session. p is highlighted to represent the groups that are significantly different at the 95% confidence level.	85
Table 7 – Comparison of Error Analysis for the first and second Sections in T1, with 95% Confidence Intervals: MAPE, MAE, and MSD.	86
Table 8 – Comparison of MAPE, MAE, and MSD Errors over Time, with Follow-up Assessments (T_2).	88
Table 9 – Results of the Tukey’s HSD hyphotesis test between the first and section of photos in the follow-up assesment. p highlighted in bold represent the groups that are significantly different at the 95% confidence level.	88
Table 10 – Comparison of Error Analysis with Follow-up Assessment, with 95% Confidence Intervals for MAPE, MAE, and MSD. $v = CI_{max} - CI_{min}$	89
Table 11 – Hyphotesis test with The Follow-up Assesment. ShapedNet vs DXA, in both photo sections. p highlighted in bold represent the groups that are significantly different.	90
Table 12 – Comparison of $\rho_{x,y}$ and R^2 for the first and second sections with follow-up assesment.	91
Table 13 – Comparison of ShapedNet and DXA regarding gender-based and non-gender-base approaches. Reported values are mean and standard deviation of BFP estimates.	94
Table 14 – Reported metrics MAE and MAPE for the non-gender based model. The 95% confidence intervals are reported as well.	105

Table 15 – Reported metrics MAE and MAPE for the male-gender based model. The 95% confidence intervals are reported as well.	105
Table 16 – Reported metrics MAE and MAPE for the female-gender based model. The 95% confidence intervals are reported as well.	105
Table 17 – Comparative Analysis: ShapedNet vs. SOTA Approaches for BFP estimate. Metrics: MAE, MAPE for Male, Female, Non-Gender. N: Number of Subjects. RM: Reference Method. Dash indicates missing information.	106

LIST OF ABBREVIATIONS

1C	One-compartment
2C	Two-Compartments
2D	2-Dimensions
3C	Three-Compartment
3D	3-Dimensions
AD	Automatic Differentiation
AI	artificial intelligence
BFF	Body Fat Mass
BIA	Bioelectrical Impedance Analysis
BMI	Body Mass Index
BMC	Bone Mineral Content
BFP	Body fat percentage
BC	Body Composition
CI	Confidence Interval
CNN	Convolutional Neural Networks
CT	Computer tomography
DNN	Deep Neural Network
DXA	Dual-Energy X-ray Absorptiometry
FFM	Fat-Free Mass
FM	Fat Mass
HSD	Honestly Significant Difference
IFCE	Federal Institute of Ceará
LBM	Lean Body Mass
MAE	Mean Absolute Error
MAPE	Mean Percentage Error
MM	Muscle Mass
MLP	Multi-layer Perceptron
MRI	Magnetic Resonance Imaging
MSD	Mean Signed Deviation
NAS	Neural Architecture Search

ReLU	Rectified Linear Unit
SOTA	State-of-the-Art
ST	Skinfold thickness
SVM	Support Vector Machines
TBMC	Total Body Mineral Content
TBW	Total Body Water
UFRJ	Federal University of Rio de Janeiro
VBC	Visual Body Composition
VGG	Visual Geometry Group
VBM	Visceral Body Mass
WHO	World Health Organization

CONTENTS

1	INTRODUCTION	17
1.1	Motivation	18
1.2	State of the art	19
1.3	Goals	20
1.3.1	<i>Specific Goals</i>	21
1.4	Contributions	21
1.5	Scientific Production	22
1.5.1	<i>Paper published in scientific journals</i>	22
1.5.2	<i>Paper published in international and national conferences</i>	23
1.5.3	<i>Papers accepted for publication/presentation</i>	24
1.5.4	<i>Papers submitted or under revision</i>	24
1.6	Thesis Structure	24
2	THEORETICAL BACKGROUND	26
2.1	Body composition	26
2.1.1	<i>Body Composition models</i>	27
2.1.2	<i>Methods for body compositions assessment</i>	27
2.1.2.1	<i>DXA – The gold standard</i>	28
2.1.2.2	<i>Well-know approaches to assess body composition</i>	30
2.1.2.3	<i>Computer vision-based method to estimate BFP</i>	33
2.2	Convolutional Neural Networks	34
2.2.1	<i>Multilayer Perceptron</i>	34
2.2.2	<i>A brief background about CNNs</i>	36
2.3	State of The Art Architectures	44
2.4	Automatic differentiation in Deep Learning	54
3	BODY FAT ESTIMATION WITH SHAPEDNET	57
3.1	ShapedNet: a multi-task learning deep neural network	57
3.2	DBF-SCAN: Visual representation of the distribution of body fat in human body	63
3.3	Database and Experiments	71
3.3.1	<i>Characteristic of BCDB23</i>	75
3.3.2	<i>Experimental setting</i>	76

4	RESULTS AND DISCUSSIONS	80
4.1	A quantitative analyses of ShapedNet results	80
4.1.1	<i>Accuracy Assessment and Method Comparison: Evaluating Error against DXA</i>	80
4.1.2	<i>Precision Evaluation of the First and Second Sections: Examining Equivalence in Error Measurements</i>	83
4.1.3	<i>Concordance Evaluation with DXA over Time with Follow-up Assessments</i>	86
4.1.4	<i>Limitations of ShapedNet</i>	94
4.2	A qualitative analyses of DBF-SCAN	95
4.2.1	<i>Limitations of DBF-SCAN</i>	101
4.3	Using Transfer Learning to estimate Body Fat Percentage	102
4.4	Comparing with State-of-the-Art (SOTA) Methods	106
5	CONCLUSIONS	109
5.1	Regarding the goals	110
5.2	Future works	111
	REFERENCES	112

1 INTRODUCTION

As stated in Article 25 of the Universal Declaration of Human Rights (GENERAL ASSEMBLY OF THE UNITED NATIONS, 1948), proper nutrition is crucial for good health. Malnutrition can take various forms, including wasting, stunting, micronutrient deficiencies, and diet-related non-communicable diseases (WORLD HEALTH ORGANIZATION, 2017). At present, one of the primary challenges posed by malnutrition is obesity and overweight, which involve the excessive accumulation of body fat and represent a significant health risk.

The prevalence of obesity has increased dramatically worldwide since 1975, as reported by the World Health Organization (WHO). In 2016, there were more than 1.9 billion adults aged 18 and over who were overweight, with over 650 million of them being obese. Moreover, more than 340 million children and adolescents aged 5 to 19 were overweight or obese. In many countries, overweight and obesity-related diseases cause more deaths than underweight-related conditions (WORLD HEALTH ORGANIZATION, 2021). Obesity is associated with a higher risk of chronic diseases, including cancer, cardiovascular disease, and diabetes.

It is alarming to realize that the majority of the world's population resides in countries where obesity and overweight cause more deaths than low body mass. These data emphasize the urgency of combating this global epidemic and implementing effective measures for the prevention and treatment of obesity (WORLD HEALTH ORGANIZATION, 2021). Therefore, it is of great importance to democratize exams that can provide useful information regarding body composition. While there is a variety of methods to calculate a person's body composition, the gold standard is Dual-Energy X-ray Absorptiometry (DXA). Despite being commonly used as a reference method for body composition assessment, including body fat (GUERRA *et al.*, 2010; HAAPALA *et al.*, 2002), DXA is not readily available in some clinics, even in major cities, and is generally more expensive than other exams.

Understanding human body composition has been a subject of investigation for over 80 years now (BEHNKE *et al.*, 1942). Over the decades, remarkable research has been carried out, including the works of Siri (), Jackson e Pollock (1978), and Wang *et al.* (1996). Despite these efforts, the problem is not yet fully understood. Presently, the scientific community is calling for methods that are more accurate, precise, and user-friendly. In addition, Artificial Intelligence (AI) has been a rapidly evolving field for more than sixty years, and recently, dedicated efforts have been made to explore the potential of AI techniques for body composition

estimation (PARIS *et al.*, 2020; WANG *et al.*, 2020; LIU *et al.*, 2021). This trend indicates a growing interest in this approach.

Therefore, the main contribution of this thesis is the development of multi-task learning a method, based on Convolutional Neural Networks (CNN), that can simultaneously identify a person, determine their coordinates in the image, and estimate their body fat percentage (BFP). Another way we aim to explore this topic is by addressing the following question: **Can a CNN be designed to analyze a person’s photo and simultaneously perform classification and regression tasks?** Furthermore, the method was designed to replicate results similar to DXA by meeting a quality standard of less than 10% differences in BFP estimation. Additionally, we developed a technique to visualize regions of interest for body fatness in an individual.

1.1 Motivation

Body composition and body fat estimation are important areas of research with significant implications for public health. Excessive body fat has been linked to a range of health problems, including cardiovascular disease, type 2 diabetes, and certain types of cancer (WANG *et al.*, 2011). Conversely, a lean body composition and adequate muscle mass have been associated with better health outcomes, such as improved physical function and reduced risk of chronic disease (JANSSEN *et al.*, 2002; SRIKANTHAN; KARLAMANGLA, 2011). Accurate measurement of body composition is therefore essential for identifying individuals at risk and developing effective interventions.

There are various methods for body composition and body fat estimation, each with its own advantages and limitations. Traditional methods include anthropometric measurements, such as skinfold thickness (POLLOCK *et al.*, 1979), and imaging techniques, such as dual-energy X-ray absorptiometry (DXA) (KNAPP, 2015). While these methods are widely used and validated, they can be costly, time-consuming, and require specialized equipment and expertise.

In recent years, there has been growing interest in the use of artificial intelligence (AI) for body composition and body fat estimation. AI techniques, such as deep learning algorithms, have shown promise in accurately estimating body composition from various types of data, such as digital images and Magnetic Resonance Imaging (MRI) and Computer Tomography (CT) (TORRES *et al.*, 2019; WANG *et al.*, 2020). These methods have the potential to offer non-invasive and cost-effective alternatives to traditional methods, as well as the ability to analyze large datasets with greater efficiency.

Another area of interest in body composition research is the role of diet and exercise in modulating body composition and body fat. A study by Wilson (Knoxville, Tennessee: University of Tennessee, 2005) evaluated the accuracy of various body composition measurement techniques in older adults, finding that DXA was the most accurate method for measuring total body fat and lean mass. Therefore, the authors emphasize one main drawback of DXA is the exposure to radiation.

Overall, research into body composition and body fat estimation is crucial for identifying and managing health risks related to excess body fat, as well as understanding the impact of diet and exercise on body composition. The use of AI techniques offers new possibilities for accurate and efficient measurement of body composition, while traditional methods continue to be important tools in the field.

1.2 State of the art

Numerous studies have proposed various approaches to calculate body fat percentage. Shao (2014) focused on predicting body fat percentage using intelligent hybrid approaches. Swainson *et al.* (2017) estimated visceral adipose tissue mass and body fat percentage using five anthropometric measures. Ferenci e Kovacs (2018) utilized neural networks to predict body fat percentage based on anthropometric measurements, using a dataset of 862 individuals over 18 years of age. Chiong *et al.* (2021) presented a method using Support Vector Machines (SVM) to estimate body fat in individuals by considering relative error. Harty *et al.* (2020) proposed the development of new equations for estimating body fat.

On the other hand, Chiong *et al.* (2021) developed a relative error SVM (IRE-SVM) to predict body fat, incorporating feature selection techniques to enhance prediction accuracy. In the study by UÇar *et al.* (2021), the estimation of body fat percentage using hybrid machine learning methods raised some concerns. For instance, in the first dataset consisting of 252 samples, the Multi-layer Perceptron (MLP) model exhibited a higher risk of overfitting.

Some studies have attempted to estimate body fat percentage through the application of Machine Learning, leading to variations in approaches and objectives. However, many of the studies available in the literature lack low-cost techniques for obtaining the essential data required for accurate body fat estimation.

The use of Convolutional Neural Networks (CNNs) has been explored in various studies in the literature, demonstrating its potential in the healthcare field. For instance, CNNs

have been applied successfully in organ segmentation (NÓBREGA *et al.*, 2018), disease diagnosis (PEREIRA *et al.*, 2018), and human pose estimation (XIANG *et al.*, 2017). Some of these studies specifically focus on developing effective and fully automated methods for estimating body measurements. Initially, there were semi-automatic methods that relied on extensive expert-defined information (LIU *et al.*, 2019; XIAOHUI *et al.*, 2018). Other proposed methods required impractical measurement times, making their use in clinical settings unfeasible (BOISVERT *et al.*, 2013).

One example is the study by Paris *et al.* (2020), which used a neural network to estimate body composition from digitalized CT scan images. The authors trained and validated the network on a dataset of over 800 subjects and showed excellent agreement between human and network-predicted segmentations for skeletal muscle, intermuscular, visceral, and subcutaneous adipose tissue. The network was able to analyze diverse body composition phenotypes and clinical cohorts with high accuracy, providing opportunities to predict health outcomes in clinical populations.

Another recent study on the topic is Wang *et al.* (2020), which used a Deep Neural Network (DNN) based on MRI to analyze body composition. The authors trained the network on a dataset of over 600 subjects and compared the results to those obtained using traditional MRI-based methods. The DNN achieved high accuracy in segmenting various body tissues such as fat, muscle, and bone, demonstrating the potential of AI techniques for accurate and efficient body composition analysis.

One of the key contributions in the field of body composition analysis made by Paris *et al.* (2020). This study addresses the laborious and costly process of manual analysis of CT scans for quantifying body composition. The authors successfully developed a framework based on deep learning for automated segmentation of skeletal muscle and adipose tissues, demonstrating excellent agreement between human and network-predicted segmentations. This automated approach offers a promising solution to enhance our capacity to predict health outcomes in clinical populations, contributing to advancements in the field of body composition assessment (PARIS *et al.*, 2020).

In addition to these studies, there are numerous other examples of the use of machine learning for body composition estimation, including various types of neural networks (MAJMU-DAR *et al.*, 2022; TIAN *et al.*, 2020). These studies collectively highlight the growing interest and potential of AI in the field of body composition estimation, offering new possibilities for

accurate and non-invasive assessment of body composition.

Our approach proposes an innovative method that not only estimates body fat using only the person's photo but provides detection and coordinates of the person as well.

1.3 Goals

The main goal of this thesis is to develop a method based on Convolutional Neural Networks that utilizes a single photo to simultaneously estimate human body fat percentage, identify a person, and visualize the importance of body regions regarding the individual's body fatness.

1.3.1 *Specific Goals*

To achieve the main goal, the following specific goals will be addressed

- Design a CNN-based algorithm for multi-task learning purposes, through multi-regression, classification, and localization tasks.
- Develop a Deep Learning architecture for precise body fat percentage estimation using only a single photo as input;
- Validate the model in a comprehensive dataset that encompasses characteristics such as repeatability and concordance;
- Validate the proposed method by comparing the estimated body fat percentage with the gold standard reference method, DXA;
- Validate gender-inclusive approaches to ensure unbiased and inclusive body fat estimation across diverse individuals;
- Design a visualization technique to highlight and quantify the relative importance of different body regions in determining body fatness. This technique will provide visual insights into the distribution and composition of body fat, allowing for a better understanding of the factors influencing an individual's body fat percentage.
- Develop an visualization method to enhance interpretability in image regression tasks;

By accomplishing these specific goals, this research aims to provide an innovative and accurate method for estimating body fat percentage using a single photo, facilitating personalized health assessments and contributing to the understanding of body composition and its implications for health and well-being.

1.4 Contributions

The main contributions of our research are enlisted below:

- Accurate body fat percentage estimation: This research introduces a deep learning architecture for precise estimation of body fat percentage from a single photo, enabling personalized health assessments.
- Validation and comparison to gold standard reference: The proposed method will be rigorously validated by comparing body fat percentage estimates to DXA. This establishes its accuracy, precision, and reliability for estimating body composition.
- Visualization of body fat distribution: This research presents a novel technique for visually quantifying body fatness and its regional distribution, improving our understanding of its health implications.
- Consideration of gender diversity: The research aims to validate approaches that consider gender diversity by removing gender bias during modeling. This inclusive approach ensures that the developed method is applicable to diverse individuals, promoting inclusivity and equal access to accurate body fat estimation for everyone.
- Advancement of body composition research: Overall, this research contributes to the field of body composition assessment by pushing the boundaries of technology and deep learning algorithms. By developing an innovative method classified as an indirect method of two compartments, regarding body composition assessment.
- Expanded multi-task learning: Our research encompasses tasks beyond body composition estimation. It enables simultaneous identification, localization, and regression, benefiting domains like computer vision, medical imaging, and biometrics.
- Explainable AI (XAI) contribution: Our research contributes to XAI by developing a visualization technique that highlights the importance of body regions in determining body fatness. This interpretable method enhances transparency and trust in AI decision-making, promoting explainability in AI models.
- Development of a novel method: The research develops an innovative method based on Convolutional Neural Networks. This method simultaneously estimates body fat percentage, identifies individuals, and visualizes the importance of different body regions in determining body fatness.

1.5 Scientific Production

Throughout the course of this thesis, the following scientific contributions were made:

1.5.1 *Paper published in scientific journals*

- REBOUÇAS FILHO, PEDRO P. ; GOMES, SAMUEL L. ; E NASCIMENTO, NAVAR M. MENDONÇA ; MEDEIROS, CLÁUDIO M.S. ; OUTAY, FATMA ; DE ALBUQUERQUE, VICTOR HUGO C. . Energy production predication via Internet of Thing based machine learning system. *Future Generation Computer Systems*, v. 97, p. 180-193, 2019.
- MARINHO, LEANDRO B. ; NASCIMENTO, NAVAR DE M.M. ; SOUZA, JOÃO WELLINGTON M. ; GURGEL, MATEUS VALENTIM ; REBOUÇAS FILHO, PEDRO P. ; DE ALBUQUERQUE, VICTOR HUGO C. . A novel electrocardiogram feature extraction approach for cardiac arrhythmia classification. *Future Generation Computer Systems*, v. 97, p. 564-577, 2019.
- SOUSA, P. H. F. ; NASCIMENTO, N. M. M. ; ALMEIDA, J. S. ; REBOUÇAS FILHO, P. P. ; ALBUQUERQUE, V. H. C. . Intelligent Incipient Fault Detection in Wind Turbines based on Industrial IoT Environment. *Journal of Artificial Intelligence and Systems*, v. 1, p. 1-19, 2019.
- NASCIMENTO, NAVAR MEDEIROS M.; MARINHO, LEANDRO B. ; PEIXOTO, SOLON ALVES ; DO VALE MADEIRO, JOÃO PAULO ; DE ALBUQUERQUE, VICTOR HUGO C. ; FILHO, PEDRO P. REBOUÇAS . Heart Arrhythmia Classification Based on Statistical Moments and Structural Co-occurrence. *CIRCUITS SYSTEMS AND SIGNAL PROCESSING*, v. 1, p. 1-20, 2019.

1.5.2 *Paper published in international and national conferences*

- LOUREIRO, L. L.; NASCIMENTO, NAVAR M.M. ; OHATA, ELENE F. ; ALVES, SHARA S. A. ; CAPISTRANO JUNIOR, V. L. M. ; REBOUCAS FILHO, PEDRO P. . Development Of A Method For Evaluating Body Composition And Perimeters Using Cell PhonePhotos. In: 2022 ACSM Annual Meeting World Congresses, 2022, California. Official Journal of the American College of Sports Medicine, 2022.
- ALVES, SHARA S.A. ; OHATA, ELENE F. ; NASCIMENTO, NAVAR M.M. ; DE

- SOUZA, JOAO W. M. ; Holanda, Gabriel B. ; LOUREIRO, LUIZ LANNES ; REBOUCAS FILHO, PEDRO PEDROSA . Gender-based approach to estimate the human body fat percentage using Machine Learning. In: 2021 International Joint Conference on Neural Networks (IJCNN), 2021, Shenzhen. 2021 International Joint Conference on Neural Networks (IJCNN), 2021. p. 1.
- REBOUÇAS FILHO, PEDRO PEDROSA; SILVA, SUANE PIRES PINHEIRO DA ; OHATA, ELENE FIRMEZA ; ALMEIDA, JEFFERSON SILVA ; SOUSA, PEDRO HENRIQUE FEIJÓ DE ; NASCIMENTO, NAVAR MEDEIROS M. ; SILVA, FRANCISCO HERCULES DOS SANTOS . A New Strategy for Mobile Robots Localization based on Omnidirectional Sonar Images and Machine Learning. In: XXXII Conference on Graphics, Patterns and Images, 2019. Anais Estendidos da Conference on Graphics, Patterns and Images (SIBGRAPI), 2019. p. 168.
 - A., M. A. ; FERREIRA JUNIOR, M. V. G. ; MARINHO, LEANDRO B. ; NASCIMENTO, N. M. M. ; SILVA, SUANE P. P. ; ALVES, S. S. A. ; RAMALHO, G. B. . Evaluation of Heart Disease Diagnosis Approach using ECG Images. In: International Joint Conference on Neural Networks (IJCNN), 2019, Budapest. International Joint Conference on Neural Networks (IJCNN), 2019.
 - NASCIMENTO, N. M. M.; BARRETO, G. A. ; NASCIMENTO, C. L. ; REBOUCAS FILHO, P. P. . Temporal Classification of Turbofan Engine Health using Elman Recurrent Network. In: XIV Brazilian Congress of Computational Intelligence, 2019, Belém, Pará. XIV Brazilian Congress of Computational Intelligence, 2019.

1.5.3 Papers accepted for publication/presentation

- Nascimento, N. M. M., Rebouças Filho, P. P. *Advancements in Body Composition Assessment: Unveiling the Power of ShapedNet and Interpretable DBF-SCAN Techniques*. in Proceedings of the International Joint Conference of Neural Networks (IJCNN) - Workshop int-XAI, 2023.
- Alves, S. S. A., Ohata, E. F., Sousa Junior, P. C., Barroso, C. B., Nascimento, N. M. M., Loureiro, L. L., Bittencourt, V. Z., Capistrano Junior, V. L. M., Rocha, A. R. da, Rebouças Filho, P. P. *Sex-based approach to estimate human body fat percentage from 2D camera images with Deep Learning and Machine Learning*. Measurement, 2023.

1.5.4 Papers submitted or under revision

- NASCIMENTO, N. M. M., REBOUÇAS FILHO, P. P. *The ShapedNet: A Novel Architecture based on Deep Regression and Semantic Segmentation to Estimate Human Body Fat*. in Proceedings of the International Conference on Computer Vision (ICCV), 2023.

1.6 Thesis Structure

In Chapter 2.2, a background is provided to establish the necessary knowledge for this thesis, including a brief overview of body composition assessment in nutrition and the theoretical foundations of CNNs. Chapter 3 presents the design and mathematical foundations for developing the novel methods, ShapedNet and DBF-SCAN. The results of the proposed methods, compared to DXA, are discussed in Chapter 4, covering accuracy, precision, and concordance analyses. And, the conclusions regarding goals and future works are discussed in Chapter 5.

2 THEORETICAL BACKGROUND

This chapter provides a comprehensive overview of the key concepts and methods relevant to the study. It begins with an introduction to body composition, including various models and methods used for assessment. The gold standard method, DXA, is discussed, alongside with well-known approaches and computer vision-based methods for estimating body fat percentage. Moving on to neural networks, the chapter explores convolutional neural networks and their architecture, including an introduction to the multilayer perceptron and a background on CNNs. The state-of-the-art architectures in the field are also examined. Additionally, the chapter touches upon automatic differentiation in deep learning, highlighting its significance in training and optimizing neural networks. Overall, this chapter provides the necessary theoretical foundation to understand the subsequent analyses and methodologies presented in the study.

2.1 Body composition

The term Body Composition (BC) refers to the distribution and quantities of components in the total body mass (HOLMES; RACETTE, 2021). Information about human body composition has been based on chemical analysis of organs and dissection of cadavers. These analyses allowed for the quantification of fat, total water, minerals, and body protein content. From this, reference data were obtained for the development of body composition models (HEYMSFIELD *et al.*, 1997). The human body composition model, proposed by Wang et al. (WANG *et al.*, 1992), is divided into five distinct levels: atomic, molecular, cellular, tissue, and whole body. The sum of the components that make up each level equals the body mass (WANG *et al.*, 1992).

In level I, the body mass consists mainly of six elements, such as oxygen, carbon, hydrogen, nitrogen, calcium, and phosphorus, which account for over 98% of body weight. These elements serve as the foundation of body composition analysis, leading to the proposed five levels. The atomic level is the starting point and allows for the division of body mass into fat and lean mass (WANG *et al.*, 1992).

Level II focuses on molecular components, including water, lipids, protein, carbohydrates, bone minerals, and soft tissue minerals. Assessment models at this level can measure various combinations of these components. Level III includes extracellular solids, extracellular water, and cells, which can be further divided into fat mass and fat-free mass. In level IV, the

body is divided into components such as skeletal muscle, adipose tissue, viscera, bones, and solid organs. Finally, level V considers the whole body as a unit, taking into account size, shape, area, and density. This level can be further divided into regions like appendages, trunk, and head. Anthropometric measurements, such as circumferences and skinfold thickness, are often used to describe the trunk and limbs (LOHMAN *et al.*, 1988).

2.1.1 Body Composition models

Initially, BC was described as Two-Compartments (2C): Fat Mass (FM) and Fat-Free Mass (FFM) (KURIYAN, 2018). Other methods include Three-Compartment (3C), Four-Compartment (4C), or multicomponent models (KURIYAN, 2018).

Two-compartment model: Divides body weight into FM and FFM based on assumptions from cadaver analysis. Methods include anthropometry, BIA, and PA (KURIYAN, 2018).

Three-compartment model: Adds a component dividing FFM into lean tissue mass and bone mineral content. Dual-energy X-ray absorptiometry (DXA) is an example method (KURIYAN, 2018).

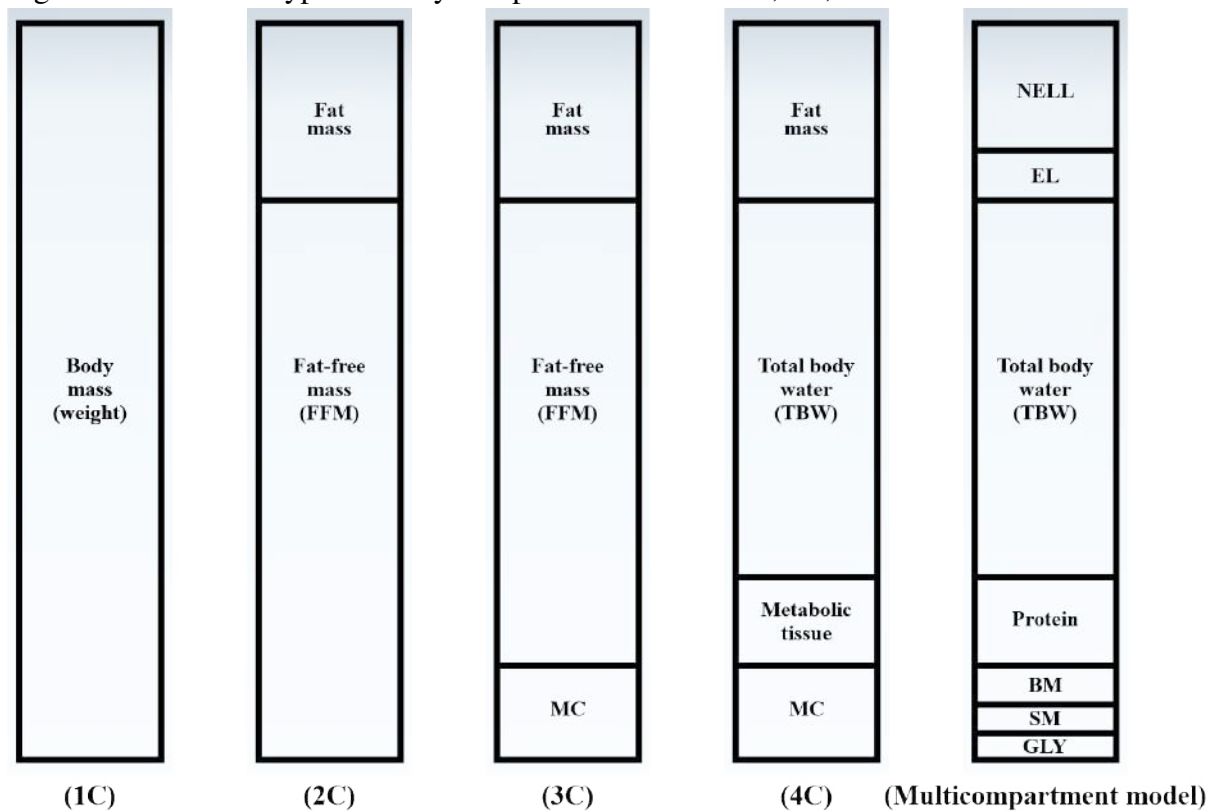
In Figure 1 different types of body composition models are presented, they are: One-compartment (1C), 2C, 3C, 4C, Body Fat Mass (BFF), Bone Mineral Content (BMC), Fat-Free Mass (FFM), FM, Lean Body Mass (LBM), Muscle Mass (MM), Total Body Water (TBW), Total Body Mineral Content (TBMC), Visceral Body Mass (VBM). The 4C model Includes FM, water, BMC, and protein. Practical use is limited due to complexity and requirements (KURIYAN, 2018). In Figure 2 a general view about their differences is shown.

2.1.2 Methods for body compositions assessment

Body composition assessment methods can be classified as direct, indirect, and doubly indirect (HEYWARD, 1996). The only truly direct method is dissection, but it has limited utility. Some studies have been conducted on cadavers to investigate body composition (JANSSEN *et al.*, 2000; EDMONDS *et al.*, 1975; BEHNKE *et al.*, 1942). Among these, one of the primary direct assessments of body composition was the Brussels Cadaver Analysis Study (CAS), in which over thirty cadavers were evaluated between 1979 and 1983.

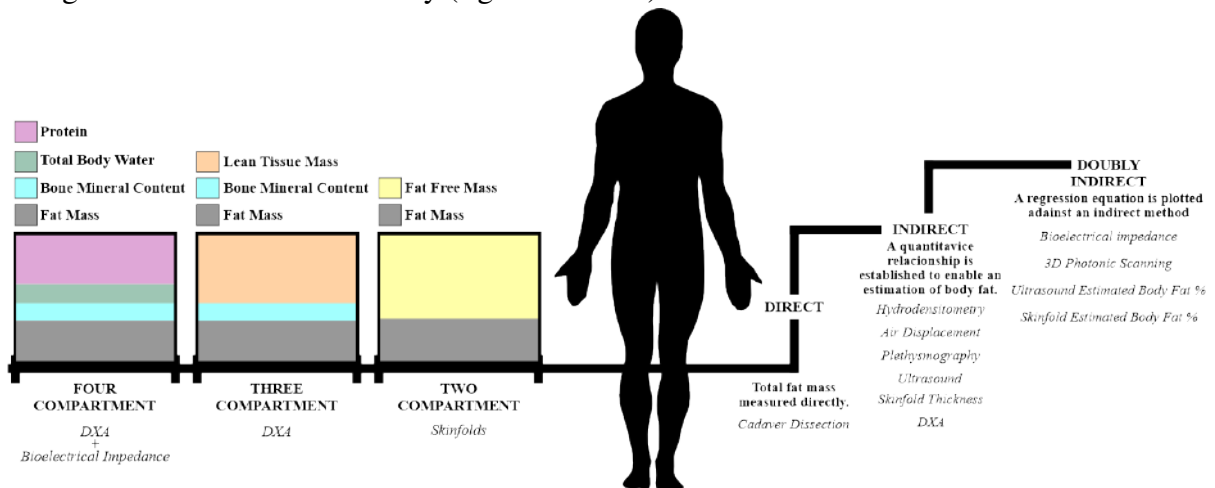
Indirect methods are based on quantitative assumptions of each body component (HEYWARD, 1996). Among these methods, pH, PDA, DXA, computed tomography, and

Figure 1 – Different types of body composition models: 1C, 2C, 3C and 4C.



Source: Adapted from Kuriyan (2018).

Figure 2 – The 2-, 3- and 4-compartmental models of human body composition (left hand side), alongside the validation hierarchy (right hand side).



Source: Adapted from Kasper *et al.* (2021).

magnetic resonance stand out as they can quantify body components with high accuracy and are mainly used to validate doubly indirect techniques (HEYWARD, 1996). Despite their high reliability, these methods require specialized professionals to operate the equipment, considerable time for the procedure, and high cost (BRODIE *et al.*, 1998; HEYWARD, 1998).

2.1.2.1 DXA – The gold standard

Historically, pH was used as the gold standard assessment technique. However, the equipment, space, expertise, and time required, combined with the burden on participants, made this method impractical in most non-research settings (BRODIE *et al.*, 1998). In this context, DXA is becoming increasingly important due to its advantages in terms of precision, simplicity, and relatively low cost compared to procedures like magnetic resonance and computed tomography. Additionally, DXA has the notable advantage of providing a segmented assessment that includes CMO and BMO (KNAPP, 2015).

DXA involves measuring the tissue absorption of high- and low-energy X-ray beams passing through the individual lying in a supine position on the scanning bed, as shown in Figure 3. The instrument utilizes the attenuation, or weakening, of these X-ray beams to provide accurate estimates of BMC and soft tissue composition. This attenuation is measurable and strongly depends on the thickness, density, and chemical composition of the underlying tissue. DXA assessment is practical, does not require active involvement from the subject, and poses minimal risk. The radiation exposure from a whole-body DXA scan is equivalent to 1-10% of a chest X-ray. However, a qualified professional is needed to perform the evaluation (BRODIE *et al.*, 1998; HEYWARD, 1998; KNAPP, 2015).

Figure 3 – Evaluation by dual-energy X-ray absorptiometry.



Source: Adapted from CAROLINA ARTHRITIS CENTER (2016).

With DXA, professionals can quantify FM, BMD, BMC, FFM, which reduces the assumptions underlying methods such as BD and PDA (BRODIE *et al.*, 1998; HEYWARD, 1998). Additionally, the equipment is also used to estimate total and regional body fat (BRODIE

et al., 1998; HEYWARD, 1998). Thus, in addition to being indicated for measuring BMD, which allows for the diagnosis and monitoring of osteoporosis, it is suggested that its use be extended to monitor overweight/obese patients. DXA measures excess adiposity and is considered a highly accurate method (KNAPP, 2015). Although promising, routine use of DXA for the diagnosis of obesity is not yet recommended, as there are few clear state, ents regarding its clinical indication for assessing body composition in patients outside the research setting (KNAPP, 2015).

DXA is primarily used to measure fat mass, specifically triglycerides, whereas conventional Magnetic Resonance Interference quantifies the volume of adipose tissue. Although the majority of body fat is stored in adipose tissue, it should be noted that MRI and DXA assess slightly different aspects and do not capture precisely the same information (MAJMUDAR *et al.*, 2022). DXA has gained widespread popularity in clinical research due to its ability to provide a favorable balance between quality, cost, and speed (KULLBERG *et al.*, 2009).

In summary, evaluating the concordance of DXA with a current state-of-the-art method for body composition assessment is crucial to validate any method's accuracy, identify biases, and ensure reliable measurement of body composition parameters. It helps determine the level of agreement between DXA and the reference method, assess the validity of DXA measurements, and identify any systematic discrepancies. Therefore, according to UÇar *et al.* (2021), standardization issues can affect its results, such as variations in equipment, data collection methods, and software analysis. Therefore, there is a need for convenient methods to predict body fat percentage. Then, understanding the limitations and potential sources of error in a given method measurements is important for appropriate interpretation and application of body composition data in research and clinical practice related to nutrition, exercise, and health interventions.

2.1.2.2 *Well-know approaches to assess body composition*

Doubly indirect methods in nutrition assessment are based on anthropometric measurements, such as body circumferences and skinfold thickness, combined with Bioelectrical Impedance Analysis (BIA). These methods utilize estimation equations derived from densitometry measurements, with densitometry being a commonly used reference method (JANSSEN *et al.*, 2000). Doubly indirect methods offer practical and cost-effective approaches for assessing body composition, particularly focusing on measurements of body fat. Anthropometric measurements, as non-invasive quantitative measures, provide valuable insights into an individual's nutritional

status. They are widely employed in clinical settings and large epidemiological studies due to their simplicity, low cost, and ease of use (CASA DEI; KIEL, 2022; KURIYAN, 2018)

Anthropometric measures encompass various parameters, including height, weight, Body Mass Index (BMI), body circumferences, and skinfold thickness. These measurements serve as indicators of adiposity, assist in health assessment, and provide information on body fat and muscle stores. Height, weight, and BMI are fundamental parameters for estimating nutritional status and assessing growth. Height can be easily measured using various equipment and is essential for growth evaluation and estimation of other body. The BMI, calculated by dividing weight (in kilograms) by the square of height (in meters), is a widely used index for categorizing individuals into weight categories and is extensively employed in population-based studies (Kuriyan, 2018). The World Health Organization classification is commonly used, which categorizes BMI as underweight ($<18.5 \text{ kg/m}^2$), normal weight ($18.5 - 24.9 \text{ kg/m}^2$), overweight ($25 - 29.9 \text{ kg/m}^2$), obesity class I ($30 - 34.9 \text{ kg/m}^2$), obesity class II ($35 - 39.9 \text{ kg/m}^2$), and obesity class III (40 kg/m^2) (WORLD HEALTH ORGANIZATION, 1998).

It is important to note that while BMI is a simple and cost-effective measure, it has limitations when assessing individual body composition. It fails to consider important factors such as the distribution of body fat, muscle mass, and bone structure, resulting in low sensitivity and specificity. Furthermore, the relationship between BMI values and body fatness is not linear and differs between men and women (ROTHMAN, 2008). Due to these limitations, alternative measures of adiposity, such as waist circumference or waist-to-height ratio (STEINBERGER *et al.*, 2005).

Waist circumference, measured at a specific anatomical landmark, serves as an indicator of intra-abdominal fat. It is obtained using a non-elastic tape with a precision of 0.1 cm, in an upright position during the final expiratory phase, at the midpoint between the lower rib cage and the iliac crest (WORLD HEALTH ORGANIZATION, 1995). Specific cutoff values for waist circumference are utilized to assess the risk of obesity-related diseases, with values $\geq 102 \text{ cm}$ in men and $\geq 88 \text{ cm}$ in women (WORLD HEALTH ORGANIZATION, 1998).

In summary, doubly indirect methods in nutrition assessment, combining anthropometric measurements and BIA, offer practical and cost-effective approaches for evaluating body composition. Anthropometric measures, including height, weight, BMI, body circumferences, and skinfold thickness, serve as valuable indicators of adiposity and provide essential information for health assessment, nutritional interventions, and future disease risk prediction. However, it is

important to consider the limitations of certain measures, such as BMI, and to explore alternative measures, like waist circumference, to enhance the accuracy of body composition assessment.

Skinfold thickness (ST) is a low-cost method to estimate body fat percentage, which measures the thickness of skinfolds at different body locations using a skinfold caliper. This method is based on the principle that the amount of subcutaneous fat is proportional to the total body fat. The measurement of ST involves compressing a double layer of skin and subcutaneous adipose tissue away from the underlying muscle and then placing a pressure-sensitive caliper on the standardized fold site. The caliper adjusts to the thickness of the skinfold and provides a measurement in millimeters (mm) (HOLMES; RACETTE, 2021).

The specific anatomical sites for ST measurements (triceps, biceps, chest, axilla, subscapular, abdominal, iliac crest, thigh) and the number of sites vary depending on the equation used to estimate ST and BFP (HOLMES; RACETTE, 2021). Common ST equations are specific to sex and age, including three (JACKSON *et al.*, 1980; JACKSON; POLLOCK, 1978), four (DURNIN; RAHAMAN, 1967), or seven skinfold sites (JACKSON; POLLOCK, 1978; JACKSON *et al.*, 1980). To convert ST to percentage each author propose equations for body density and BFP estimation. Other equations, also include age, circumferences, and BMI for both sexes (BEHNKE *et al.*, 1942; GALLAGHER *et al.*, 2000).

Bioimpedance is a non-invasive technique used to assess various physiological parameters by measuring the impedance of biological tissues. It is based on the principle that different tissues have distinct electrical properties due to variations in their cellular composition and water content (KYLE *et al.*, 2004). Bioimpedance analysis involves passing a small, safe electrical current through the body and measuring the impedance, which consists of resistance and reactance. By analyzing these impedance values, BIA can provide estimates of body composition, including measures such as fat-free mass, fat mass, and total body water (KYLE *et al.*, 2004). BIA is widely used in clinical and research settings for assessing hydration status, monitoring changes in body composition, and evaluating nutritional status. It offers several advantages, such as being non-invasive, portable, and relatively low-cost compared to other methods like dual-energy X-ray absorptiometry. Additionally, BIA can provide quick results and does not expose individuals to ionizing radiation (LUKASKI, 2013). However, BIA has certain limitations, including sensitivity to factors like hydration status, body position, and electrode placement. These factors can introduce variability in the measurements and affect the accuracy of the results. It is crucial to standardize measurement protocols and consider individual

characteristics when interpreting BIA data (LUKASKI, 2013).

Ultrasound is a non-invasive imaging technique used in body composition assessment that utilizes high-frequency sound waves to analyze various body tissues. By creating detailed images, it can measure subcutaneous fat thickness, muscle thickness, and other tissue characteristics. Ultrasound is valuable in estimating body fat percentage, muscle mass, and provides real-time imaging with portability and cost-effectiveness (KASPER *et al.*, 2021). However, limitations include operator-dependent results, variation in image quality and interpretation, and inability to directly measure internal organs or differentiate between similar tissues. However, according to (KASPER *et al.*, 2021), currently there are no universally accepted guidelines for measuring subcutaneous adipose tissue with ultrasound.

2.1.2.3 Computer vision-based method to estimate BFP

The study of Majmudar *et al.* (2022) evaluates an automated computer vision method called Visual Body Composition (VBC), which uses two-dimensional smartphone photographs to estimate percentage total body fat (%BF). The VBC algorithm, based on a convolutional neural network, aims to provide accurate and accessible body fat measurements compared to other consumer-grade devices. The study includes 134 healthy adults who underwent %BF measurements with VBC, consumer and professional bioimpedance analysis (BIA) systems, and DXA as the reference. Results show that VBC had the lowest mean absolute error and standard deviation compared to other methods, with good concordance and tight limits of agreement with DXA. The study concludes that VBC provides accurate and unbiased %BF estimates, surpassing other BIA and DXA methods, and its accessibility through smartphones makes it a valuable tool for quantifying adiposity.

The results reported by Majmudar *et al.* (2022) demonstrated the lowest mean absolute error and standard deviation ($2.16 \pm 1.54\%$) when measuring body fat percentage (%BF) in relation to DXA. VBC also showed good concordance with DXA, as indicated by high correlation coefficients: all participants (0.96), women (0.93), and men (0.94). In contrast, BMI exhibited poor concordance with DXA, with CCC values of all participants (0.45), women (0.40), and men (0.74).

One of the main drawbacks of the work of Majmudar *et al.* (2022) that the evaluation was conducted on a relatively small sample size of 134 healthy adults. This limited sample size may not fully represent the diverse range of individuals and populations, and it could affect the

generalizability of the findings.

The purpose of the study of Tian *et al.* (2020) was to develop an algorithm for predicting total and regional body composition from a single 2-Dimensions (2D) frontal photograph using consumer-grade equipment. The algorithm was trained on 3D optical scans and DXA whole-body scans of a subset of participants from the Shape Up! Adults Study. A linear regression model was used to establish the relationship between the 3-Dimensions (3D) shape and DXA-derived body composition. The algorithm accurately predicted body composition from the silhouette obtained from the 2D photograph, with high precision and no significant differences compared to DXA measurements. This approach has the potential to enable widespread and cost-effective measurement of body composition, allowing for early screening and monitoring of metabolic disease indicators in settings where medical imagery or clinical assessment is not readily available.

However, there are also some limitations or drawbacks to consider on the work of Tian *et al.* (2020). Firstly, the study was conducted using a specific dataset from the Shape Up! Adults Study, which may limit the generalizability of the algorithm to other populations. Moreover, there have been evaluated 416 subjects, so further validation and testing in diverse populations would be necessary to confirm its reliability across different demographic groups. The authors did not reported mean absolute errors or mean percentage error, in fact they report R^2 and $RMSE$ regarding the DXA values.

Moreover, it is important to highlight that using a 2D system to estimate body compositions can be useful, especially when combined with a 3D model, such as a three-component (3C) model, for example.

2.2 Convolutional Neural Networks

2.2.1 Multilayer Perceptron

The Multilayer Perceptron is a non-parametric supervised machine learning method. It is an extension of the Perceptron classifier designed to handle non-linearly separable datasets (HAYKIN, 2008).

The Perceptron algorithm, introduced by Rosenblatt (1958), computes a weighted sum of the inputs and applies it to an activation function. In essence, each weight represents the relevance of a particular feature in predicting a specific class. More formally, the Perceptron

prediction can be expressed as shown in Equation 2.1.

$$\hat{y} = \varphi \left(\sum_{i=1}^p x_i w_i + b \right), \quad (2.1)$$

where φ is an activation function (e.g., logistic/softmax), x is an input vector, p is the number of attributes in a given input, w is the weight vector, and b is a bias term. Both w and b are learnable parameters of the Perceptron. Additionally, the training process of this classifier is performed iteratively, given by:

$$w_i(t+1) = w_i(t) + \eta \cdot (y - \hat{y}(t)) \cdot x_i \quad \forall i = 1, \dots, p, \quad (2.2)$$

where t is a particular training time step, y is the true label, $\hat{y}$ is the predicted label, and η is the learning rate. However, despite being a simple and powerful algorithm, the Perceptron is not capable of performing nonlinear classification.

To overcome this limitation, the MLP was proposed. This classifier consists of a computational graph composed of a stack of perceptron layers. Each of these layers, except for the last one, has its logistic activation functions, which are used to create a linear classifier, replaced by another non-linear function, such as the sigmoid function, which is used to perform a non-linear transformation in the feature space. More formally, the decision rule of the MLP can be mathematically formulated as:

$$\hat{y}_{l_k} = \begin{cases} \varphi_{0_k}(\sum_{i=1}^p x_i \cdot w_{0_{k_i}} + b_{0_k}) & \text{if } l = 0 \\ \varphi_{l_k}(\sum_{i=1}^D \hat{y}_{l-1_i} \cdot w_{l_{k_i}} + b_{l_k}) & \text{otherwise} \end{cases}, \quad (2.3)$$

where x is the input vector of p dimensions, l is a specific layer index, D is the number of neurons in layer $l - 1$, and φ is the non-linear function. It is worth noting that, in general, the learnable parameters of the MLP (weights and biases) are initialized with random values close to zero.

Furthermore, to improve its performance, the MLP typically employs an iterative optimization process, expressed by the equation:

$$\hat{w}_{l_{k_i}}(t+1) = \begin{cases} w_{0_{k_i}}(t) + \eta \cdot \delta_{1_i}(t) \cdot x_k(t) & \text{if } l = 0 \\ w_{l_{k_i}}(t) + \eta \cdot \delta_{l+1_i}(t) \cdot \hat{y}_{l_k}(t) & \text{otherwise} \end{cases}, \quad (2.4)$$

where η is the learning rate and $\delta_{l_k}(t)$ is the local gradient of the k -th neuron in the l -th layer. This local gradient is given by:

$$\delta_{l_k}(t) = \begin{cases} \phi'_{l_k}[u_{l_k}(t)] \cdot \frac{1}{2M} \sum_{o=1}^M [y_{l_o}(t) - \hat{y}_{l_o}(t)]^2 & \text{if } l = z \\ \phi'_{l_k}[u_{l_k}(t)] \cdot \sum_{j=1}^J w_{l_{k_j}}(t) \cdot \delta_{l+1_j}(t) & \text{otherwise} \end{cases}, \quad (2.5)$$

where z is the last layer, M is the number of neurons in the last layer (z), and J is the total number of neurons in the l -th layer.

Note that the derivative $\phi' i[u_{l_i}(t)]$ has different values depending on the chosen activation function, for example:

$$\phi'_{l_k}[u_{l_k}(t)] = \begin{cases} y_{l_k}(t)[1 - y_{l_k}(t)], & \text{if logistic/sigmoid} \\ \frac{1}{2}[1 - (y_{l_k}(t))^2], & \text{if hyperbolic tangent} \end{cases}. \quad (2.6)$$

2.2.2 A brief background about CNNs

Convolutional Neural Networks are a class of Deep Learning algorithms designed to process matrices such as images. Unlike classical computer vision algorithms, CNNs can operate directly on raw data with less preprocessing. Another characteristic of these networks, compared to primitive computer vision methods, is their ability to learn relevant features for each problem (LECUN *et al.*, 1998).

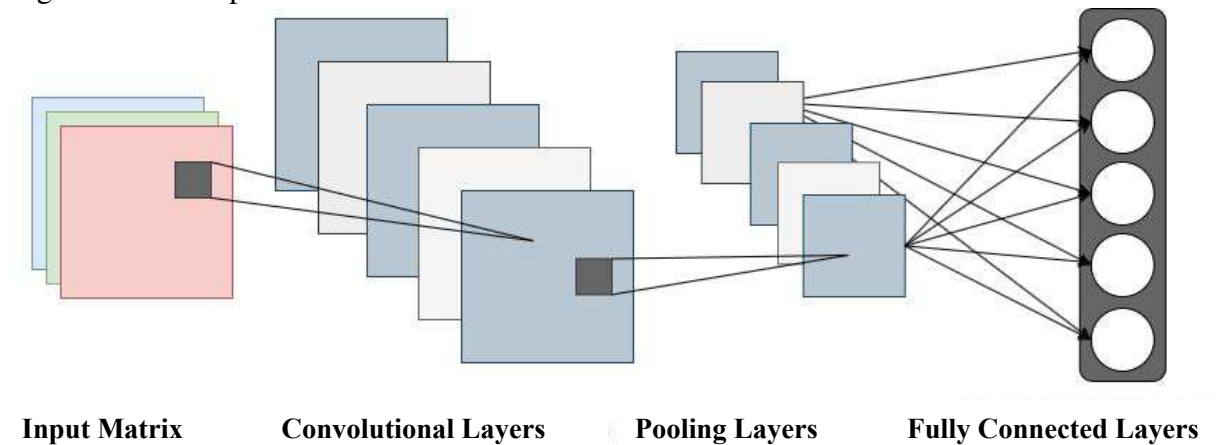
The convolutional layers in a CNN resemble the structure of the human visual cortex, where a series of layers process an input image and identify progressively more complex features. Therefore, each of these stacked convolutional layers is capable of learning more sophisticated shapes (BISHOP, 2006).

ANNs do not adapt well to full images. In the Canadian Institute For Advanced Research (CIFAR-10) dataset, for instance, even with image sizes of $32 \times 32 \times 3$ (32 width channels, 32 height channels, and 3 color channels), a single fully connected neuron in the first hidden layer of an ANN would have $32 \times 32 \times 3 = 3072$ weights. This quantity still seems manageable, but clearly, this fully connected structure does not scale well to larger images. On the other hand, a larger input, e.g., $200 \times 200 \times 3$, would require neurons with $200 \times 200 \times 3 = 120,000$ weights. Notably, such connectivity is wasteful, and the large number of parameters would quickly lead to overfitting.

On the other hand, CNNs exploit the fact that the input is a structured data matrix to optimize their architecture. Unlike an ANN, the neurons in CNN layers are organized in width, height, and depth. It's important to note that the term "depth" does not refer to the total number of layers in a network. In CIFAR-10, for example, the volume would have dimensions of $32 \times 32 \times 3$ due to having 3 color channels. Additionally, neurons in a layer are only connected to a small region of the previous layer, rather than all neurons.

A CNN consists of various types of layers, which are described in the following sections: convolutional layer, pooling layer, and fully connected layer. Stacking these layers forms a CNN architecture (GOODFELLOW *et al.*, 2016). Figure 1 shows an example of a CNN architecture with these layers.

Figure 4 – Example architecture of a CNN.



Source: Elaborated by the author.

The Convolutional Layer

The convolutional layer is the central building block of a CNN and is responsible for the majority of the computational work. Its parameters consist of a set of filters that can be learned. Each filter is small spatially (along the width and height) but extends across the entire depth of the input volume. As the filter is convolved with this input volume, a two-dimensional activation map that provides the filter's responses at each spatial position is produced. Intuitively, the network will learn filters that activate in the presence of visual patterns. These activation maps are stacked along the depth dimension and produce the output volume (GOODFELLOW *et al.*, 2016). The discrete convolution between two functions f and g is defined as follows:

$$(f * g)(x) = \sum_t f(t)g(x+t). \quad (2.7)$$

For two-dimensional signals, we consider 2D convolutions.

$$(\mathbf{W} * \mathbf{I})(i, j) = \sum_{m, n} \mathbf{W}(m, n) \mathbf{I}(i+n, j+m), \quad (2.8)$$

where $\mathbf{W}$ is a convolution kernel (or filter) applied to a two-dimensional signal $\mathbf{I}$.

2D convolution involves moving a kernel over the structured data matrix. At each position, the convolution between the kernel and the region of the matrix is computed. The kernel is then moved by a number s of cells, where s is called the stride. When the stride is small, redundant information is obtained. Sometimes, zero-padding is also added, which is a margin of size p containing zero values around the matrix to control the output size. Suppose C_{i+1} kernels are used, each with size $w \times w$. Also, consider that the size of the input matrix is $L_i \times A_i \times C_i$, where L_i is the width, A_i is the height, and C_i is the number of channels. In this case, the output volume is $L_{i+1} \times A_{i+1} \times C_{i+1}$, where C_{i+1} corresponds to the number of kernels considered, and

$$\begin{aligned} L_{i+1} &= \frac{L_i - w + 2p}{s} + 1, \\ A_{i+1} &= \frac{A_i - w + 2p}{s} + 1. \end{aligned} \quad (2.9)$$

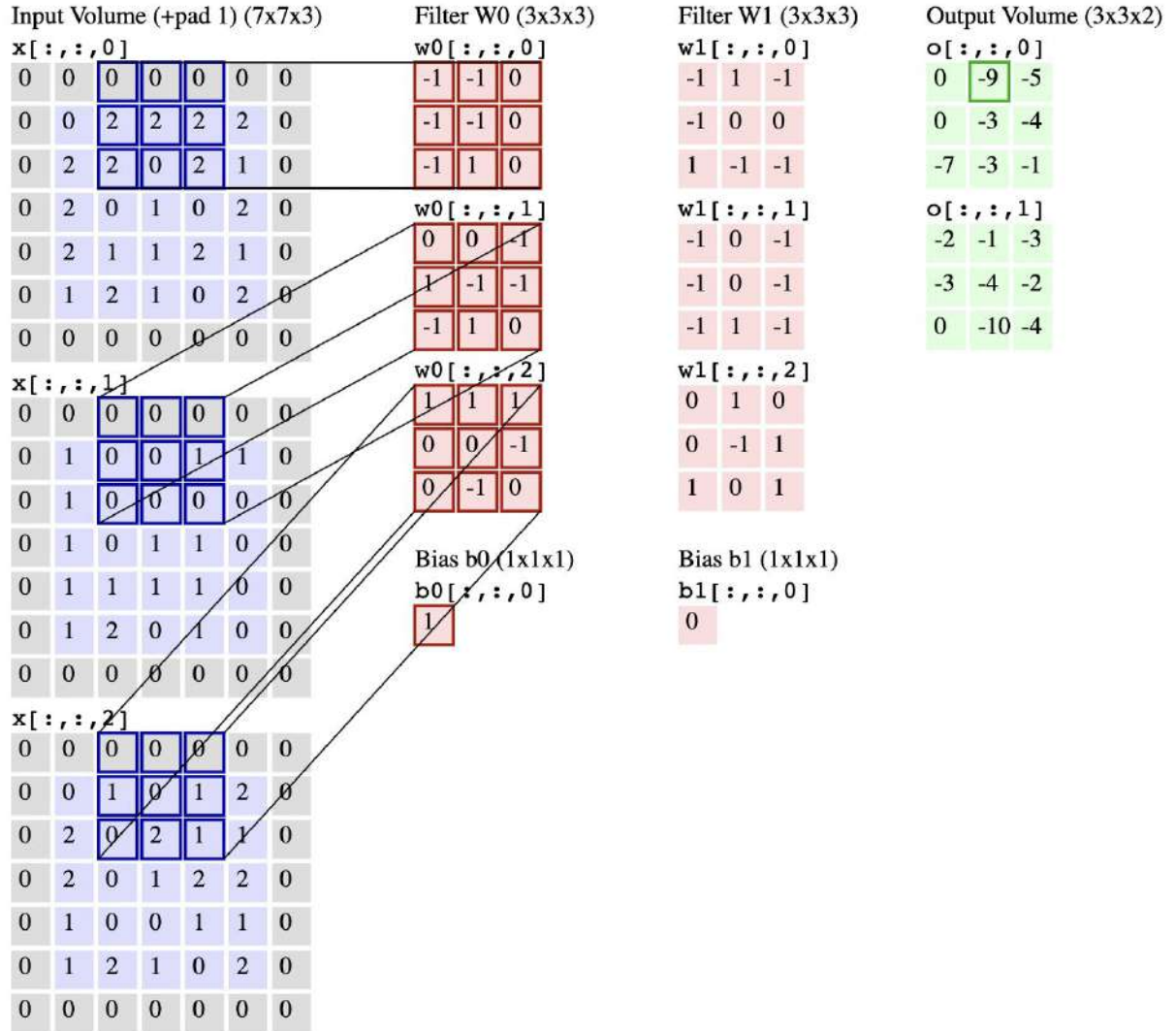
Figure 5 exemplifies the execution of the convolution operation. To facilitate understanding, the depth of the volumes is presented in slices. The input matrix has dimensions of $(5 \times 5 \times 3)$ and is highlighted in blue. Two filters of size 3×3 (red matrix) are used in the convolutional layer. Additionally, the stride and padding values are $s = 2$ and $p = 1$, respectively. Therefore, the size of the output volume is $(5 - 3 + 2)/2 + 1 = 3$. The example shows that each element is computed by element-wise multiplication of the highlighted input with the filter, summing them up, and then offsetting the result by the bias.

If the matrix has 3 channels and $\mathbf{W}_l$ ($l = 1, \dots, C_{i+1}$) denotes $5 \times 5 \times 3$ kernels, the convolution of matrix $\mathbf{I}$ with kernel $\mathbf{W}_l$ corresponds to Equation 1:

$$\mathbf{W}_l * \mathbf{I}(i, j) = \sum_{c=0}^2 \sum_{n=0}^4 \sum_{m=0}^4 \mathbf{W}_l(n, m, c) \mathbf{I}(i+n-2, i+m-2, c). \quad (2.10)$$

⁰ Available at: <<https://cs231n.github.io/convolutional-networks/>>. Acesso em 12 de maio de 2023.

Figure 5 – Illustration of the convolution operation. The input matrix is highlighted in blue, the filters of the convolutional layer are shown in red, and the resulting output volume from the convolution is depicted in green.



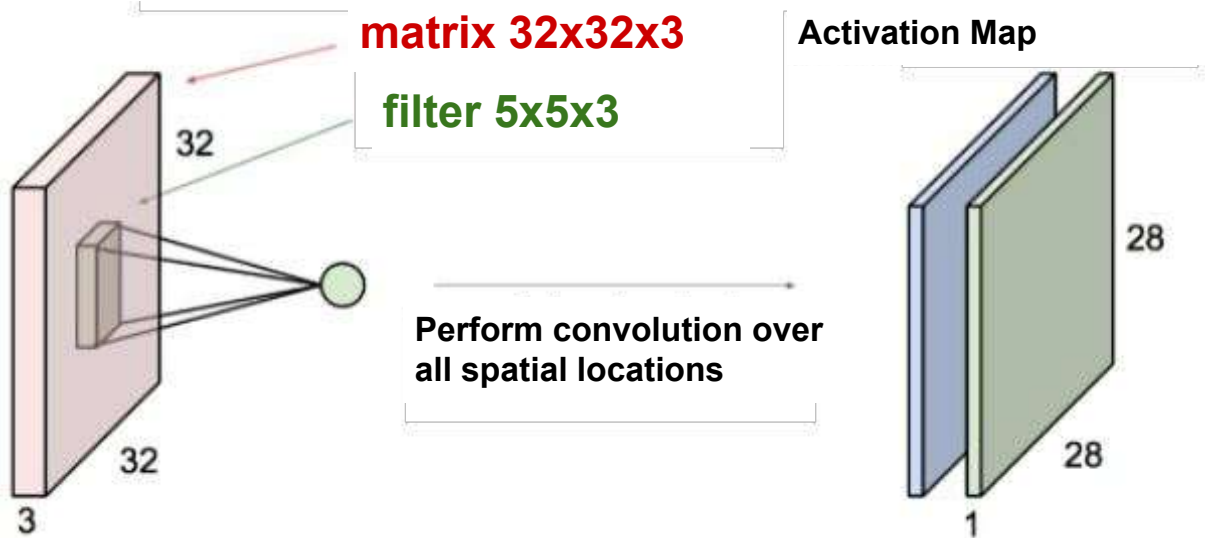
Source: Adapted from CS231n *Convolutional Neural Networks for Visual Recognition*.

In a more general sense, for images with C_i channels, the kernel is of size (w, w, C^i, C^{i+1}) , where C^{i+1} is the depth of the output, i.e., the number of considered kernels. In Figure 3, for example, the kernel is $(5, 5, 3, 2)$. The number of parameters associated with a kernel of shape (w, w, C^i, C^{i+1}) is $(w \times w \times C^i + 1) \times C^{i+1}$.

The convolution operations are combined with an activation function $\phi(\cdot)$, commonly the ReLU activation function. For a kernel $\mathbf{W}$ of size $w \times w$ and a region $\mathbf{x}$ of the same size as the image, the activation is obtained by sliding the $w \times w$ window and computing $z(\mathbf{x}) = \phi(\mathbf{W} * \mathbf{x} + b)$, where b is the bias.

It is in the convolutional layer where the actual learning takes place through the

Figure 6 – Activation map generated by a convolutional layer with 2 filters.



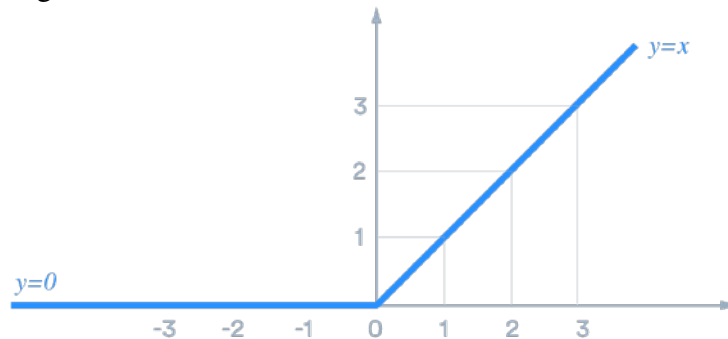
Source: Elaborated by the author.

adaptation of filters for each specific task. Another advantage is that multiple convolutional layers can be considered, so the output of one convolution becomes the input for the next one.

The non-linearity ReLU

In order for a neural network to be effective, it needs to incorporate non-linear activation functions. In the case of convolutional neural networks (CNNs), the output of the convolution operation is passed through the Rectified Linear Unit (ReLU) activation function. This introduces non-linearity by transforming the output values to be positive or zero, enhancing the network's ability to capture complex patterns and features in the data.

Figure 7 – The ReLU activation function.



Source: Elaborated by the author.

The Figure 7 presents a graphical representation of the ReLU function. The adoption of this function was mainly motivated by the experiments conducted by Nair and Hinton (NAIR;

HINTON, 2010), which indicated that the use of ReLUs significantly accelerates the convergence of training compared to other non-linear activation functions such as logistic sigmoid or hyperbolic tangent. The ReLU function can be mathematically defined as shown in Equation 2.11.

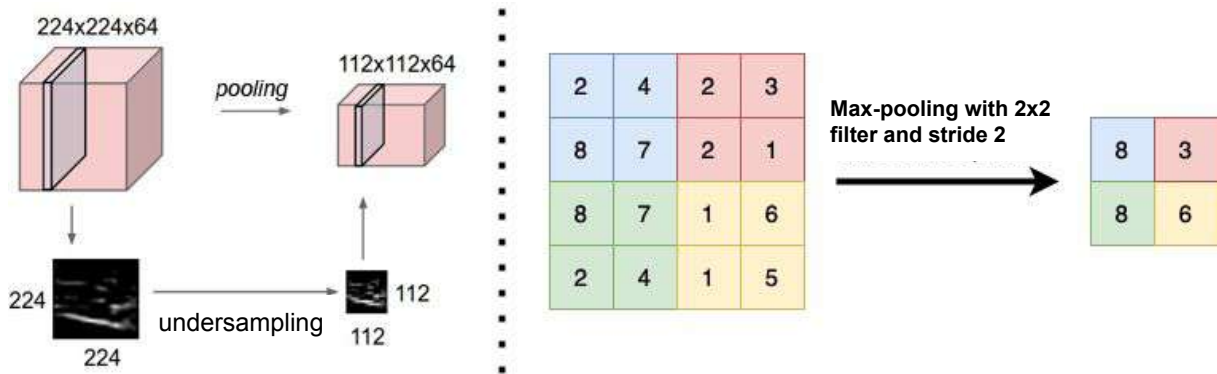
$$g(z) = \max(0, z) \quad (2.11)$$

The Pooling layer

Another important layer in CNNs is the pooling layer, also known as subsampling. This layer allows for dimension reduction by calculating the average or maximum value within regions of the image.

Like the convolutional layers, the pooling layer operates on small patches of the image and can have strides. By calculating the maximum value within 2×2 regions and considering a stride of $s = 2$, the width and height of the image are divided by 2, as shown in Figure 3. It is worth noting that it is also possible to reduce the dimension of the convolutional layer with a larger stride and without zero-padding. Another advantage of the pooling layer is that it makes the network less sensitive to small translations of the input images.

Figure 8 – Illustration of the effect of the pooling layer with the maximum operator.



Source: Elaborated by the author.

In a more general sense, the pooling layer takes a volume of size $L_i \times A_i \times C_i$, requires two hyperparameters (its spatial extent and stride), and produces a volume of size $L_{i+1} \times A_{i+1} \times C_{i+1}$, where:

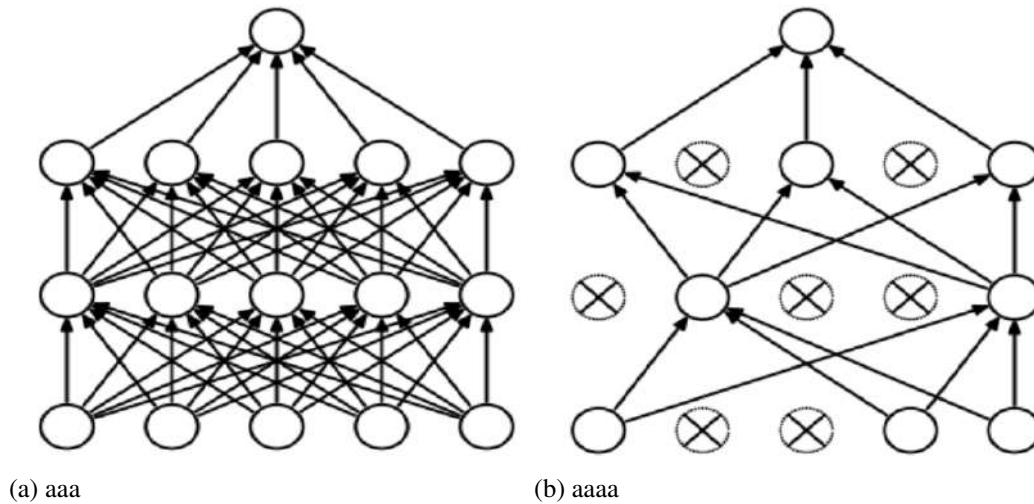
$$\begin{aligned}
L_{i+1} &= \frac{L_i - w}{s} + 1, \\
A_{i+1} &= \frac{A_i - w}{s} + 1, \\
C_{i+1} &= C_i.
\end{aligned} \tag{2.12}$$

The Dropout Layer

It is quite common for a network to learn to classify images from its training set very well. However, when presented with new images, its performance may decline. To prevent this issue, known as overfitting, the technique of Dropout is commonly used.

The Dropout method involves randomly deactivating some units in the neural network Srivastava *et al.* (2014), as shown in Figure 9. According to this idea, these deactivated neurons (set to zero) do not contribute to the forward propagation of the network and, consequently, their weights are not updated during the backward gradient propagation. A probability p is set, and each unit has this probability of its signal being propagated in the network.

Figure 9 – Dropout illustration. in (a), conventional flow of information during a forward pass. In (b), this flow is interrupted (set to zero, or Dropped out) in some neurons.



Source: Source: Srivastava *et al.* (2014).

The Batch-normalization layer

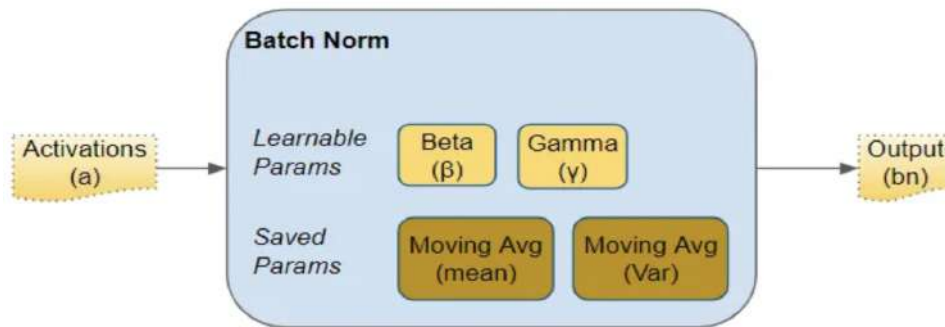
One of the most challenging problems in optimizing deep networks is Internal Covariate Shift (IOFFE; SZEGEDY, 2015), which is essentially the change in the distribution

of network activations due to the change in network parameters during training. This occurs because the inputs to a particular layer are affected by all the weights of the preceding layers, which typically change during optimization. Furthermore, as the layer weights need to adapt to a new input distribution with each weight update, a detrimental effect arises, slowing down the model's convergence.

To address the issue of Internal Covariate Shift, (IOFFE; SZEGEDY, 2015) proposed batch normalization, a new mechanism that incorporates normalization activations into the architecture itself.

Normalization is a well-known concept in machine learning, and the same principle applies here. The main difference is that batch normalization can be applied before feeding inputs into each layer, and it is not limited to just the first layer. In fact, it is common to use batch normalization between layers in deep neural networks. Let us exhibit the process in Figure 10 and Figure 11.

Figure 10 – Parameters of a Batch Norm layer.

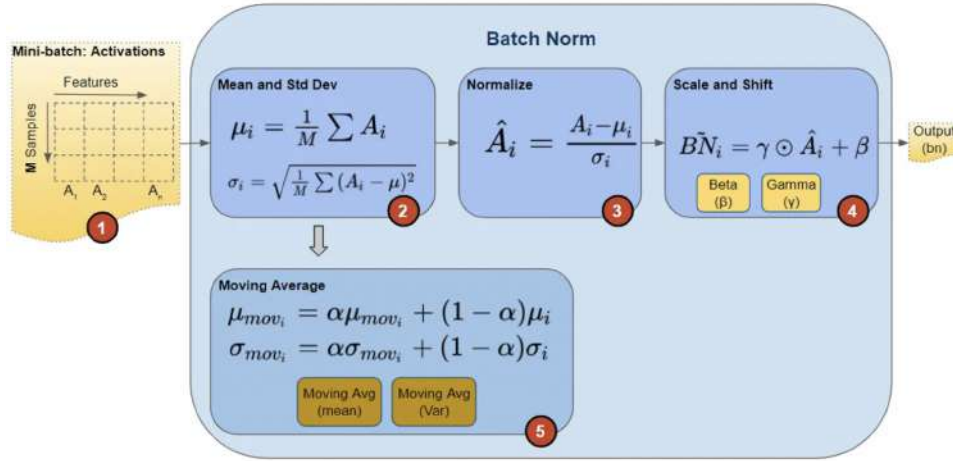


Source: Adapted from (ANDRES, 2019).

With just two additional parameters per activation, batch normalization normalizes the outputs of the dot product between weights and inputs of a given layer based on their mean and standard deviation. This is equivalent to "cleaning" the data by making it centered around zero and normalizing its range across the batch. Additionally, among the benefits of batch normalization are: the layer following a batch normalization layer focuses on how to best associate inputs; batch normalization has a powerful regularization effect and reduces or eliminates the need for dropout.

Considering γ and β as the parameters to be learned during training, the output is calculated according to Equation 2.13.

Figure 11 – Calculations performed by Batch Norm layer.



Source: Adapted from (ANDRES, 2019).

$$y = \gamma \hat{x} + \beta, \quad (2.13)$$

where $\hat{x}$ represents the normalized input. To calculate $\hat{x}$, based on Equation 2.16, for a mini-batch $M = x_1, x_2, \dots, x_b$, where b is the batch size, it is necessary to first compute the mean and variance of the mini-batch using Equations 2.14 and 2.15, respectively.

$$\mu_M \leftarrow \frac{1}{b} \sum_{i=1}^b x_i, \quad (2.14)$$

$$\sigma_M^2 \leftarrow \frac{1}{b} \sum_{i=1}^b (x_i - \mu_M)^2, \quad (2.15)$$

$$\hat{x}_i = \frac{x_i - \mu_M}{\sqrt{\sigma_M^2 + \lambda}}, \quad (2.16)$$

where the constant λ is introduced to provide numerical stability.

The Fully Connected Layer

The neurons in a fully connected layer have connections to all activations in the previous layer, as seen in ANNs. Their activations can therefore be calculated using a matrix

multiplication along with the bias term. This layer is typically a Multilayer Perceptron with a single hidden layer. Nowadays this type of network is referred as *vanilla neural networks*.

The only difference between fully connected layers and convolutional layers is that neurons in the latter are connected only to a local region in the input and many of them share parameters. However, neurons in both layers still compute scalar products, so their functional form is identical. Therefore, conversion between these layers is possible (GOODFELLOW *et al.*, 2016).

2.3 State of The Art Architectures

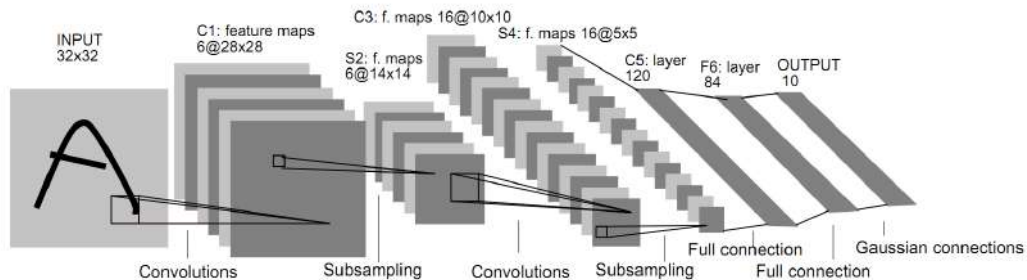
This section contains a review describing the most important and recent CNNs architectures supported by the scientific community.

Lenet-5

Inspired by Artificial Neural Networks (HAYKIN, 2008), the Neocognitron architecture (FUKUSHIMA; MIYAKE,), the back-propagation optimization strategy (BENGIO *et al.*, 2015), and the concept of shared weights (also known as local connectivity (RUMELHART *et al.*, 1986)), and driven by the need to solve real-world image recognition problems without computationally expensive pre-processing stages, LeCun *et al.* (1998) developed a pioneering 7-level convolutional network for classifying handwritten numbers in check images (LECUN *et al.*, 1990).

The architecture presented in Figure 12 was named LeNet-5, and its main contributions to the field of Deep Learning are the introduction of convolutional layers and subsampling layers.

Figure 12 – The architecture of the LeNet-5 Convolutional Neural Network.



Source: Adapted from LeCun *et al.* (1998).

AlexNet

Despite being developed in the late 1980s, CNNs only gained popularity in 2012 when Krizhevsky et al. proposed a CNN model called AlexNet (KRIZHEVSKY *et al.*, 2012). This model overwhelmingly won the Large-Scale Visual Recognition Challenge (ILSVRC) (RUSSAKOVSKY *et al.*, 2015) for object detection, becoming a significant reference for computer vision techniques (DENG *et al.*, 2009; KRIZHEVSKY *et al.*, 2012). It's worth noting that the attention given to the ILSVRC competition is mainly due to the size and complexity of the dataset. The dataset, known as ImageNet, consists of 1000 categories and over 1.2 million images (DENG *et al.*, 2009).

Since the discovery of AlexNet, every subsequent winner of the ILSVRC has used CNN-based models. However, despite this excitement, the AlexNet model has a similar architecture to its predecessor, LeNet-5 (LECUN *et al.*, 1998), consisting mainly of convolutional layers, pooling layers, and fully connected layers.

In addition, among the most notable contributions of the AlexNet paper to the field of CNN research are the Rectified Linear Unit (ReLU) and Dropout techniques.

VGG architectures

The experiments conducted in the Oxford Visual Geometry Group (VGG) (SIMONYAN; ZISSERMAN, 2014), including the winning model for the ILSVRC 2014 detection and localization tasks (RUSSAKOVSKY *et al.*, 2015), demonstrated that for all evaluated architectures (11, 13, 16, or 19 layers), there is a positive correlation between increasing model depth (number of layers) and accuracy on the ImageNet Dataset. This increase, without overfitting the model, was made possible through a regularization strategy called Factorized Convolutions.

Factorized Convolutions

Inspired by the article by Zeiler and Fergus (ZEILER; FERGUS,), which used smaller receptive field sizes (kernels/filters) compared to the AlexNet, the VGG network adopted a different approach. For instance, in the first convolutional layer, Zeiler and Fergus used 7x7 kernels, while the AlexNet used 11x11 kernels. However, the VGG network employed very small kernels of size 3x3 in all of its convolutional layers.

Using smaller kernels allows the VGG network to capture more detailed local

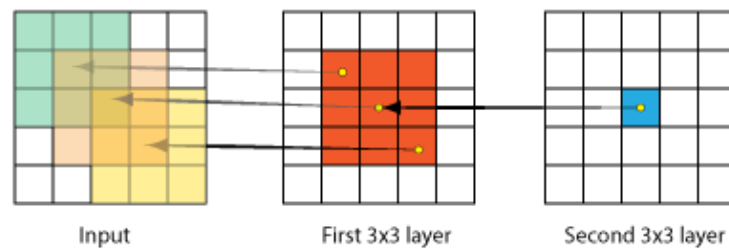
information in the input images. By stacking multiple layers with 3×3 kernels, the network is able to learn complex hierarchical representations. This approach contributes to the network's ability to achieve impressive performance on various visual recognition tasks

There are three important aspects that need to be clarified to justify why small kernels allowed the VGG network to outperform most previous models. First, in a convolution operation, a specific point in the activation map is influenced only by a connected sub-region of the input. For example, a 3×3 convolution is influenced by a 3×3 input window.

Additionally, if a second 3×3 convolution is performed on the first activation map, each point in the new activation map will be influenced by a 3×3 window of the first map. However, since each point in the first activation map combines information from a 3×3 window of the input, a combination of 3×3 points in the first map is influenced by a 5×5 information from the input.

Therefore, if this process is repeated one more time, each output point of the three-layer stack of 3×3 convolutions will be influenced by a sub-region of 7×7 from the input. In other words, a stack of three 3×3 convolutions has the same effective receptive field as a single 7×7 convolution.

Figure 13 – The effective receptive field of a stack of two 3×3 convolutions.



Source: Adapted from Simonyan e Zisserman (2014).

Secondly, a 3×3 convolutional stack with three layers has fewer parameters than a single 7×7 convolutional layer. Thirdly, since each layer has its non-linearity, a stack of convolutional layers has more discriminative power than just a single layer. Therefore, the VGG networks have become quite popular not only because of their good results in the ILSVRC 2014 (RUSSAKOVSKY *et al.*, 2015), but also due to their uniform architecture.

Inception architectures

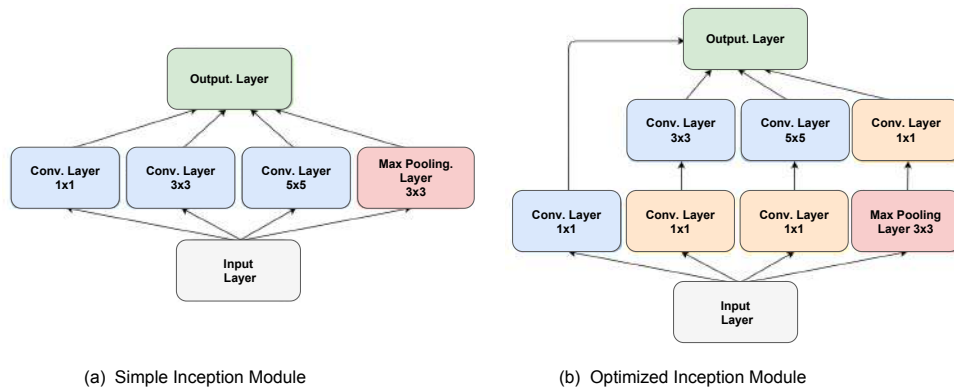
Since LeNet-5, the basic elements of deep neural architectures remain the same (convolutional layers, pooling layers, and fully connected layers). However, this changes when the *Google Brain* team introduces *GoogLeNet* (SZEGEDY *et al.*, a). Inception, also known as *GoogLeNet*, sets state-of-the-art results in the ILSVRC 2014 detection task (RUSSAKOVSKY *et al.*, 2015), achieving a top-5 error rate of 6.67

Moreover, among the contributions of the *Inception* models to the field of deep learning research are the *Inception* module, *Batch Normalization*, and Asymmetric Convolutions.

Inception module

Designing a CNN architecture typically involves choosing which operations to use (1x1 convolution, 3x3 convolution, 5x5 convolution, or pooling) in each layer. So, why not perform all operations at once and concatenate their outputs, leveraging feature extraction at multiple levels? An illustration of this parallel architecture is shown in Figure 14(a).

Figure 14 – Two illustrations of initial architecture: (a) Simple Inception module and (b) Inception module with dimensionality reduction.



Source: Adapted from (SIMONYAN; ZISSERMAN, 2014).

The main drawback of this simple Inception module (the parallel approach shown in Figure 14) is that, after a few layers like this, the number of parameters added to the network would likely make it excessively large during training.

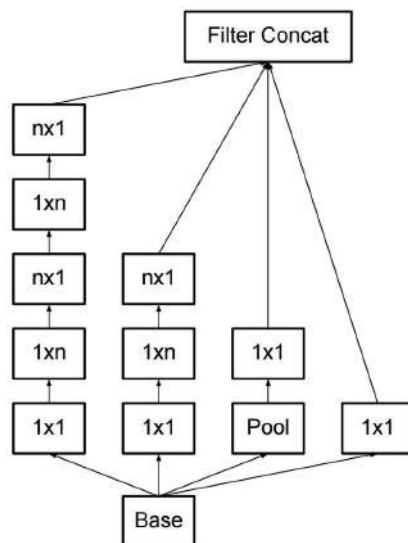
However, to address the parameter inefficiency, *Google Brain* performed a 1x1 convolution to compress the inputs of the convolutional layers and the output of the pooling layers. An illustration of the Inception module is shown in Figure 14(b). This simple adaptation

preserved most of the accuracy of the parallel approach while significantly reducing the number of free parameters.

Assymetrical Convolutions

Extending the concept of Factorized Convolutions, previously applied in the VGG article (SIMONYAN; ZISSERMAN, 2014), and also based on Flattened CNNs (JIN *et al.*, 2014), the Google Brain team employs Asymmetric Convolutions to propose a new version of Inception modules aimed at improving the parameter efficiency. They observe that instead of performing an $n \times n$ convolution, it is better to apply an $n \times 1$ convolution followed by a $1 \times n$ convolution, thus factorizing it into two asymmetric convolutions. A generic architecture of an Inception module with asymmetric convolutions is shown in Figure 16.

Figure 15 – Here is an illustration of the architecture of the Inception module with asymmetric convolutions.



Source: Adapted from (SZEGEDY *et al.*, b).

ResNet Architectures

Despite the indications from VGG, which suggest a positive correlation between network depth and accuracy, other experiments conducted in a Residual Neural Network (ResNet) article (WU *et al.*, 2017) do not support this view, at least for classical CNNs (networks composed solely of convolutional layers, pooling layers, and fully connected layers).

The authors of ResNet argue that very deep classical CNNs are more difficult to

optimize, primarily due to the problem of vanishing gradients (BENGIO *et al.*, 1994). In practice, an 18-layer network exhibits higher test and training errors compared to a 34-layer network, even though these additional 16 layers perform only an identity mapping.

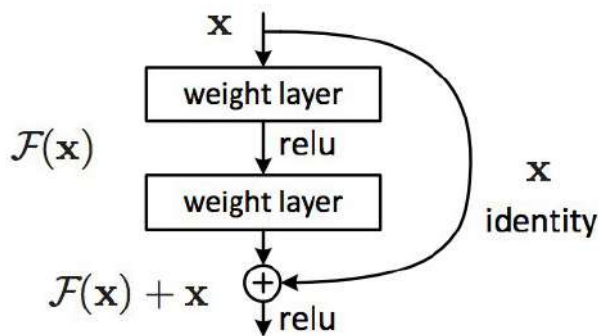
To overcome this problem and inspired by the concept of "skipping connections" of Recurrent Neural Networks, the authors of ResNet introduced a new building block of deep neural networks, the Residual Block.

Residual blocks

In a residual block, the input flows into two branches: one connects the input to the merge stage, and the other performs a stack of convolutional layers (typically 2 or 3) while preserving the dimensions of the input (e.g., if the input has C channels, then C filters are applied) and sends the results to the merge stage.

Finally, the outputs from both branches are summed together. It's important to note that the last convolutional layer in the second branch does not have any non-linearities after the convolution operations. These missing activation functions are applied to the output of the merge stage, composing the output of the residual block. An illustration of a generic version of this residual block is shown in Figure ??.

Figure 16 – An example of the architecture of a Residual Block.



Source: Adapted from Wu *et al.* (2017).

With this simple strategy, ResNet achieved state-of-the-art results in object detection, localization, and segmentation tasks in ILSVRC (RUSSAKOVSKY *et al.*, 2015) and COCO 2015 (LIN *et al.*, 2014), achieving a top-5 classification error rate of 3.57% (ILSVRC) and even surpassing human-level performance on the ImageNet dataset.

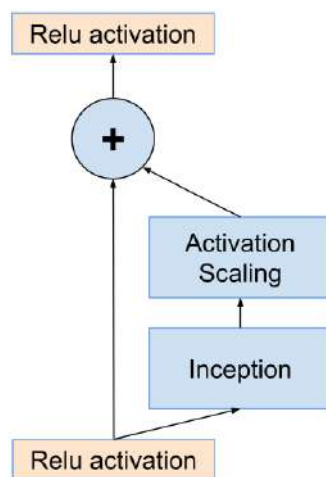
Inception-ResNet architectures

In 2016, less than a year after ResNet, the Google Brain team conducted architecture design experiments by combining the ideas of Residual Connections (WU *et al.*, 2017) with the Inception Module (SZEGEDY *et al.*, a). Among these experiments is the Inception-ResNet-V2 model (LäNGKVIST *et al.*, 2014), which achieved a top-1 classification error rate of 16.5% and a top-5 error rate of 3.1%, surpassing the performance of both ResNet and Inception models.

Inception Module with Residual Connections

The only difference between a direct combination of an Inception Module with Residual Connections is a scaling layer in the output of the Inception Module. Without this scaling stage, the network starts producing only zeros after a few thousand iterations. The authors conducted several experiments to address this issue, but only reducing the initial output by a factor between 0.1 and 0.3 before adding it to the inputs stabilized the optimization process.

Figure 17 – A generic illustration of the overall architecture when combining Residual Connections with the Inception Module.



Source: Adapted from (LäNGKVIST *et al.*, 2014).

Xception architecture

In 2016, François Chollet released the paper "Extreme Inception" (Xception) (CHOLLET,), investigating the hypothesis that channel correlations can be completely decoupled from spatial correlations. To achieve this, the author proposes the Depthwise Separable Convolution

layer.

It is worth noting that with 36 layers of Depthwise Separable Convolutions structured into 14 modules (most of them with Residual Connections), the Xception architecture achieved impressive results on the JFT dataset (HINTON *et al.*, 2015), a database of 350 million images with 17,000 possible labels, surpassing the performance of both ResNet and ImageNet architectures.

Depthwise Separable Convolutions

In a 2D convolution layer, the idea is to learn features directly from a 3D space (width, height, and channels). Therefore, a convolution maps correlations between channels and spatial locations simultaneously.

Furthermore, in an Inception Module, a convolution is factorized into a set of 1x1 convolutions, which are specifically influenced by channel correlations, followed by regular convolutions (e.g., 3x3 or 5x5 convolutions) that map spatial correlations and cross-channels.

However, in Depthwise Separable Convolution, first a depthwise convolution is performed, which maps only spatial correlations. Then, a set of pointwise convolutions is calculated, mapping channel correlations, and finally, all the convolution outputs are concatenated in the depth dimension. Thus, spatial correlation is completely dissociated from channel correlations.

Among the benefits of Depthwise Separable Convolution over the classic Inception Module are a reduction in the number of parameters and more efficient computation. These two characteristics are valuable for mobile or even embedded applications.

MobileNet Architectures

Em 2017, os pesquisadores do Google projetaram e lançaram uma arquitetura *mobile-first* chamada MobileNet (HOWARD *et al.*, 2017). Essa rede de 28 camadas é essencialmente (exceto a primeira camada) construída em convoluções separáveis por profundidade, que foram introduzidas por (HOWARD *et al.*, 2017) em 2016 e alcançaram, em sua configuração de linha de base, 99,1% da acurácia de detecção do InceptionV3 (SZEGEDY *et al.*, a) (no banco de dados *ImageNet*) com 87,7% menos de parâmetros e 88,6 % menos de operações.

A abordagem MobileNet distingue-se do modelo Xception pela introdução de dois hiperparâmetros (multiplicador de largura, multiplicador de resolução), que permitem que o

modelo corresponda, de forma eficiente, às restrições de *hardware* dos sistemas móveis e embarcados.

Width Multipliers

To create flexible and efficient architectures, the authors of MobileNets propose a hyperparameter α , also known as the "width multiplier." In the MobileNets paper, this parameter α is a reduction factor applied to the number of channels in depthwise separable convolution layers. Additionally, α affects the number of parameters and computational cost of the model almost quadratically. Note that $\alpha \in (0, 1]$, where 1 represents the baseline configuration of MobileNet.

Resolution Multiplier

The second proposed hyperparameter is the resolution multiplier, denoted as ρ . In practice, this parameter ρ adapts the network architecture to process different input resolutions (width and height). For example, if ρ is set to 1, the network input will be a 224×224 image. If ρ is set to 0.75, the input will be a 168×168 image, and so on. Similar to α , $\rho \in (0, 1]$ (with $\rho = 1$ being the baseline), and it also has a quadratic effect on the computational cost.

Therefore, the width multiplier and resolution hyperparameters offer a useful trade-off between size, accuracy, and latency, allowing for customization according to the hardware limitations of a system.

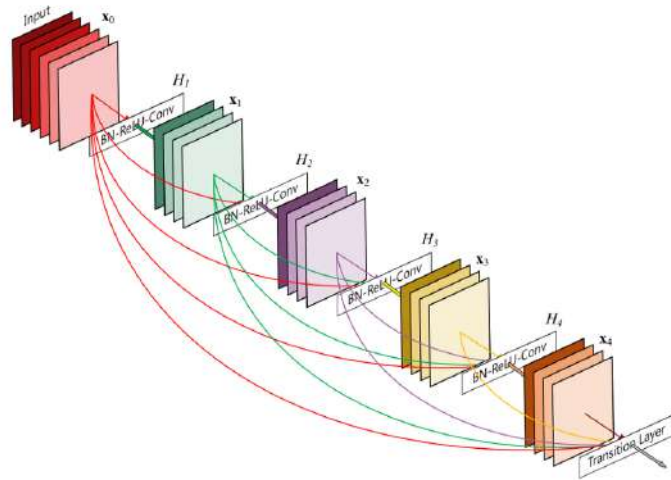
DenseNet Architectures

The DenseNet, proposed by Huang *et al.* (), extends the concept of short connections previously explored by Szegedy *et al.* (a), composing what is called a Dense Block.

Dense Blocks

There are two main differences between DenseNet and ResNet architectures. First, instead of performing short connections between adjacent layers, DenseNet connects each layer to all other layers with the same input resolution (width and height). This set of interconnected layers is called a Dense Block. An illustration of a dense block is shown in Figure 18. Second, every DenseNet layer performs a deep concatenation of previous outputs instead of summation.

Figure 18 – Illustration of the architecture of a Dense Block.



Source: Adapted from (HUANG *et al.*,).

As a result of this strategy, DenseNet slightly outperforms ResNet in terms of accuracy on the ImageNet detection task, while requiring approximately 55% fewer parameters and 50% fewer operations than ResNet.

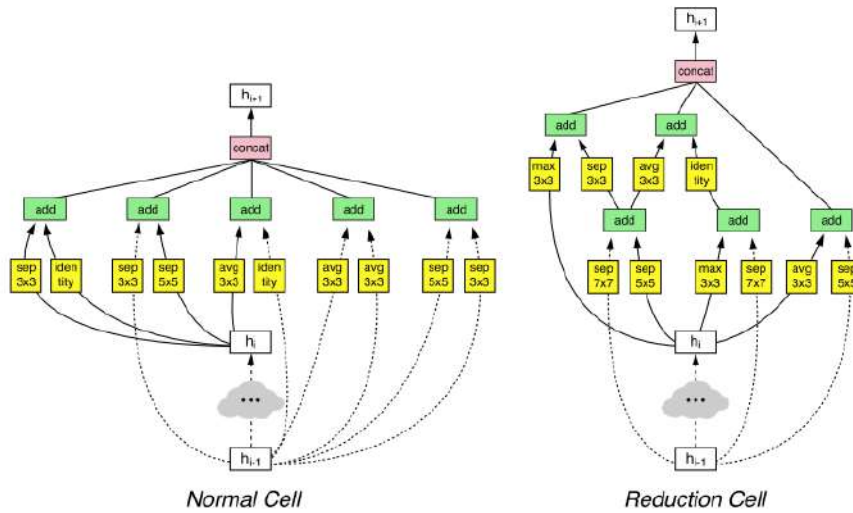
NASNet architectures

Based on the assumption that designing accurate and efficient image classification architectures is often a complex process of architectural engineering, the article "Neural Architecture Search" (NAS) (ZOPH; LE, 2016) explores the possibility of efficiently learning a model architecture directly from a dataset of interest. Inspired by the NAS experiments, the Google Brain team proposed a new paradigm for designing convolutional architectures, the NASNet search space.

NASNet Search Space

This paradigm restricts the vast search space by combining various building blocks (e.g., identity, 1×3 and 3×1 convolution, 1×7 and 7×1 convolution, 5×5 max pooling, 1×1 convolution, 3×3 convolution, depthwise separable convolution 3×3 , etc.) to form a complete network, seeking only two architectural modules (normal cell and reduction cell). An illustration of the discovered accuracy modules (on ImageNet) in the NASNet search space is shown in Figure 19. These modules are alternated and repeated multiple times to compose the final architecture.

Figure 19 – Illustration of the NASNet modules.



Source: Adapted from (ZOPH; LE, 2016).

Furthermore, to accelerate this search process, NASNet relies on the assumption that a module learned on a small dataset, such as CIFAR-10, can be transferred to a larger dataset, such as ImageNet.

As a result, NASNet achieved state-of-the-art detection accuracy on the ImageNet dataset, specifically 82.7% top-1 accuracy and 96.2% top-5 accuracy. Additionally, in its "lite" version, NASNet matches the parameter efficiency of MobileNet while being more accurate and faster. However, it is worth noting that even with the use of 500 Nvidia K40 GPUs, the fastest architecture search took 4 days.

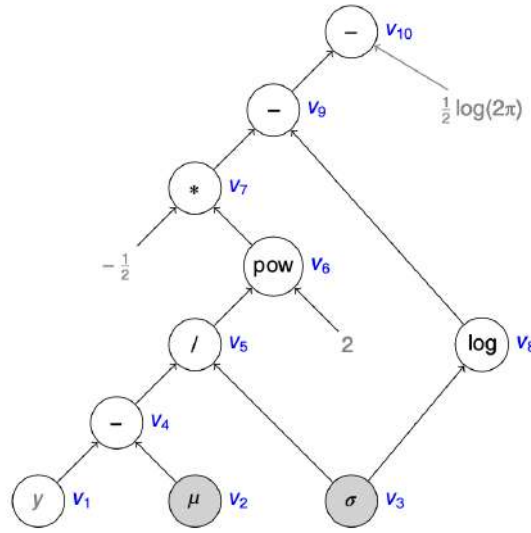
2.4 Automatic differentiation in Deep Learning

Automatic Differentiation (AD) has emerged as a significant advancement in training deep learning models, particularly in the domain of multi-task learning where the loss function equations can become complex. This method enables the accurate evaluation of partial derivatives, surpassing the limitations of symbolic and numerical differencing approaches (GRIEWANK *et al.*, 1989).

Furthermore, AD addresses the challenges and numerical approximations associated with traditional methods of symbolic or numerical differentiation. By decomposing functions into elementary operations and applying the chain rule of calculus, AD provides an exact and precise approach to computing derivatives (BAYDIN *et al.*, 2018).

Moreover, the concept of AD aligns well with the theory of computational graphs, as any mathematical function can be represented as a computational graph, as illustrated in Figure 1. Modern deep learning models are typically formulated as computational graphs (PASZKE *et al.*,), leveraging the principles of automatic differentiation to compute partial derivatives of intricate loss functions.

Figure 20 – Illustration of a computational graph



Source: Adapted from Margossian (2019).

Moreover, the key aspects of AD that make it preferable over symbolic differentiation are:

- Efficiency: AD offers computational efficiency by directly computing the derivatives numerically without explicitly constructing and manipulating symbolic expressions. This makes AD particularly well-suited for functions with high-dimensional inputs and complex structures, such as deep neural networks;
- Accuracy: AD provides high numerical accuracy since it computes derivatives by evaluating the function and its derivatives at specific input values. It avoids the accumulation of errors that can occur in symbolic differentiation, especially when dealing with complex expressions;
- Flexibility: AD can handle functions that are defined by computer programs or numerical simulations. It can differentiate through loops, conditional statements, and other control structures, allowing for differentiation of complex algorithms;
- Computational Graph Flexibility: AD allows for dynamic computation graphs, where the graph structure can change during runtime. This flexibility is particularly beneficial

in machine learning frameworks, where the computational graph depends on the model architecture and input data. Most common libraries have AD in their frameworks;

- Handling Non-Differentiable Functions: AD can handle functions with non-differentiable components by providing subgradients or directional derivatives, allowing for optimization of functions with certain types of non-differentiability;
- Integration with Optimization Algorithms: AD seamlessly integrates with optimization algorithms, such as gradient-based methods like stochastic gradient descent (SGD). The ability to efficiently compute gradients using AD makes it a natural choice for training deep neural networks and optimizing complex objective functions.

While symbolic differentiation has its own advantages, such as providing analytical expressions for derivatives, AD has become the method of choice for differentiating functions numerically in many practical scenarios.

3 BODY FAT ESTIMATION WITH SHAPEDNET

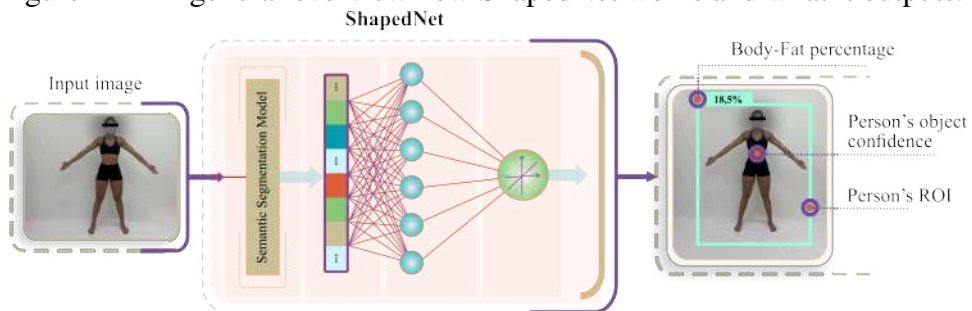
The methodology chapter presents the approaches and techniques employed in this study for body composition analysis. The chapter begins by introducing ShapedNet, a multi-task learning deep neural network developed for the purpose of this research. ShapedNet aims to address the challenges of accurately estimating body fat percentage and fat distribution. The next section discusses DBF-SCAN, which is a visual representation of the distribution of body fat in the human body. This innovative approach provides insights into the spatial distribution of body fat and aids in analyzing fat patterns. The chapter then delves into the database and experiments conducted in the study. It highlights the characteristics of the BCDB23 database used, including the demographics and specific attributes of the dataset. The experimental setting is also described, providing details on the protocols, measurements, and procedures employed during the experiments. By outlining these methodologies, this chapter lays the foundation for the subsequent analysis and results presented in the study.

3.1 ShapedNet: a multi-task learning deep neural network

The goal of our network is not limited to predict human's body fat percentage, but also localize and identify the person's body. The whole process is summarized in Figure 21. The core idea is that the feature representations used to identify objects in inner layers empowers of meaningful representation to estimate body fat percentage as well.

An input image containing a person is given to the model and after a series of operations, such as convolutions, the network provides the following outputs: (1) the body fat percentage; (2) the person's ROI information, *i.e.* x, y, w, h ; (3) person's object confidence and (4) the object class, in this case is set to 1, since the network is only trained to identify humans.

Figure 21 – A general overview how ShapedNet works and what it outputs.

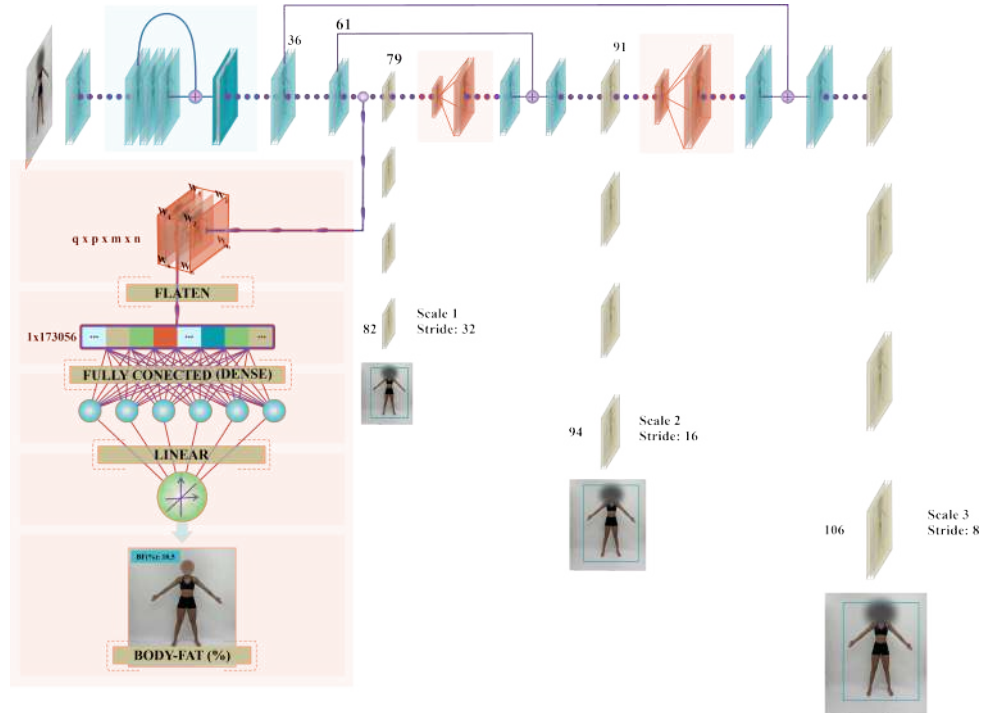


Source: Elaborated by the author.

We inspired ourselves in Redmon e Farhadi (2018) to design the network's architecture. The higher level architecture of the ShapedNet is summarized in Figure 22. The upper part of the figure are a series of convolutional layers, residual, concatenation and stride, but most of it is omitted in this figure and will be explained later. The layer's numbers are the top part of the figure and the first 53 layers are built based on Darknet-53. And, right after the last layer of the network we created a branch, which is highlighted in the figure. The input of the branch is modeled as a 4-D tensor, with dimensions $q \times p \times m \times n$, wherein $q = 1024$, $p = 13$, $m = 13$, $n = 1$. Thus, a flatten layer vectorizes everything into a 1-D tensor of size 1×173056 , then the tensor flows to a fully connected layer, *i.e.* Dense with 173056 neurons and 1 output neuron using a linear function to estimate body fat percentage.

Another output are created from layers 79, 91 and 103, these were originally developed by Redmon e Farhadi (2018) to make the model able to perform a multi scale detection, which means detect objects at different scales and sizes.

Figure 22 – The higher-level architecture of ShapedNet.



Source: Elaborated by the author.

Thus, a lower level architecture of ShapedNet is exhibited in Figure 23. The convolutional layers work as feature extractors, and according to Redmon e Farhadi (2018), this configuration is able to get both meaningful semantic information and finer-grained understand-

ing, due to the residual blocks and features concatenation from a set of layers. It is important to note that operations such as batch normalization, dropout, activation function are still present, they were only omitted for simplicity.

The layers of for body fat regression are highlighted in yellow in Figure 23, and as abovementioned there are three layers here, one flatten, one dense with a linear activation function, which is represented as another layer in this illustration just to the reader should keep in mind that others functions commonly used in regression could be employed as well.

Therefore, there are three main parts that composes the multi-task learning of Shaped-Net they are: (1) the object detection, (2) the classification outputs and the (3) regression output. They are $\hat{\mathbf{Y}}_{loc}$, the one that will be later used to estimate coordinates of the given person, which is an tensor of size $[bs \times cs \times cs \times 3 \times 2]$. Plus, $\hat{\mathbf{Y}}_{conf}$, that estimates the confidence of the prediction bounding box, this is a tensor of size $[bs \times cs \times cs \times 3 \times 2]$ as well. And, $\hat{\mathbf{P}}_c$, which gives the probability distribution of the given classes, in a tensor size $[bs \times cs \times cs \times 3 \times nc]$. So, these ouputs composes a multi dimensional tensor, represented as follows,

$$\hat{\mathbf{Y}} = \left[\hat{\mathbf{Y}}_{loc} \quad \hat{\mathbf{Y}}_{conf} \quad \hat{\mathbf{P}}_c, \right]. \quad (3.1)$$

Additionally, the output used to estimate body fat percentage is $\hat{\mathbf{Y}}_{bf}$ of ShapedNet is,

$$\hat{\mathbf{Y}}_{bf} = f(\mathbf{y}_{fl}^T \mathbf{w}_{bf}), \quad (3.2)$$

being, $\mathbf{w}_{bf}$ the vector of parameters that ought to be optimize later. Its size is $[1 \times 173056]$. And,

$$\mathbf{y}_{fl} = vec(\mathbf{y}_{conv-53}), \quad (3.3)$$

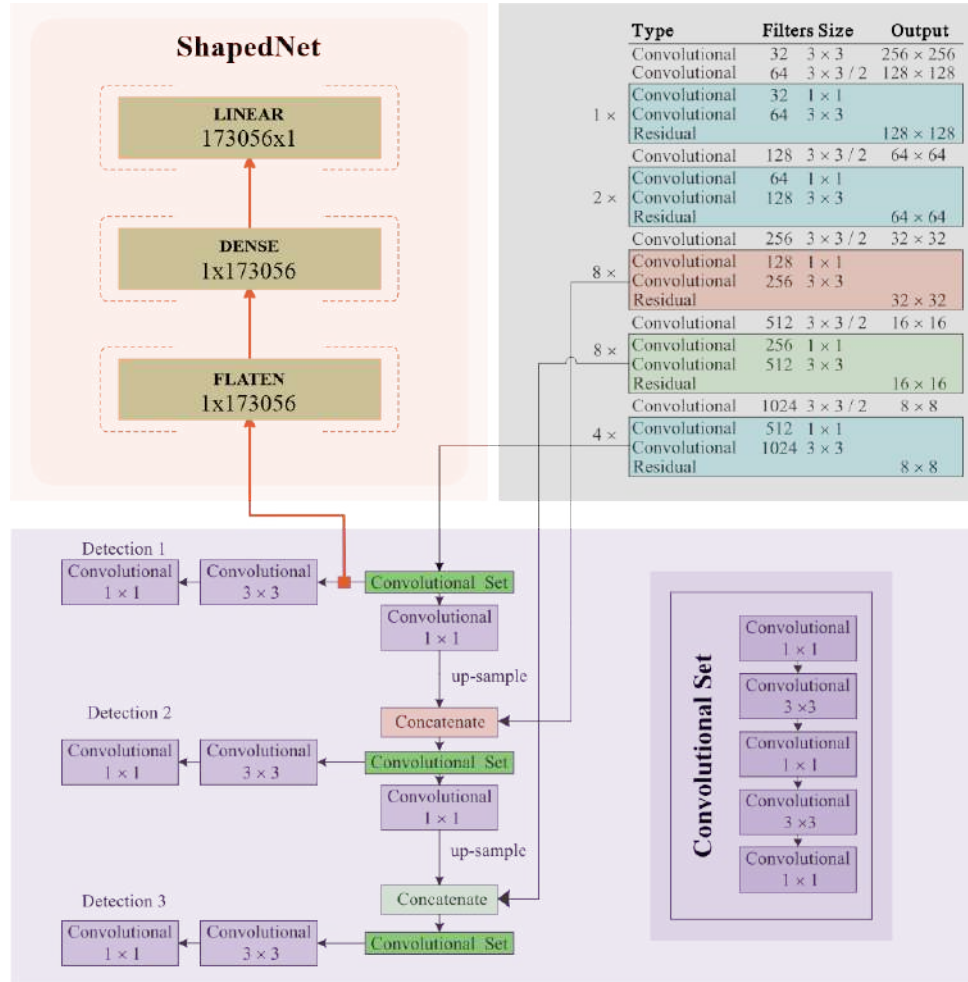
wherein $vec(\cdot)$ is the vectorization function, that transforms a tensor of any size into a 1-D tensor. And f is the identify function, $f(X) = X$.

Let the output tensor of Equation 3.1 be expanded to

$$\hat{\mathbf{Y}}_{loc} = \left[\hat{\mathbf{Y}}_{loc} \quad \hat{\mathbf{Y}}_{conf} \quad \hat{\mathbf{P}}_c \mid \hat{\mathbf{Y}}_{bf}, \right], \quad (3.4)$$

and now we have the full ouput of the ShapedNet.

Figure 23 – The lower-level architecture of ShapedNet.



Source: Elaborated by the author.

The addition of the term $\hat{\mathbf{Y}}_{bf}$ to $\mathbf{Y}$ requires that a loss term is considered to optimize the parameters in $\mathbf{w}_{bf}$. So, a new term, L_f , was added to the original loss function. So, we consider the following multi-part loss function:

$$\mathbf{L} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \quad (1)$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \quad (2)$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \quad (3)$$

(3.5)

$$+ \lambda_{\text{no-obj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{no-obj}} (C_i - \hat{C}_i)^2 \quad (4)$$

$$+ \sum_{i=0}^{S^2} \mathbb{I}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \quad (5)$$

$$+ L_f, \quad (6)$$

The L_f is the loss term to measure body fat percentage error, and is expressed as,

$$L_f = \lambda_{\text{f}} \frac{(b - \hat{b})}{b}. \quad (3.6)$$

Note that b is the true value of body fat percentage and $\hat{b}$ is the predicted value. λ_{f} is a value used to provide a weight, usually equal to 1, to the total loss. Finally, the total loss function is

$$\mathbf{L} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \quad (1)$$

$$+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \quad (2)$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 \quad (3)$$

(3.7)

$$+ \lambda_{\text{no-obj}} \sum_{i=0}^{S^2} \sum_{j=0}^B \mathbb{I}_{ij}^{\text{no-obj}} (C_i - \hat{C}_i)^2 \quad (4)$$

$$+ \sum_{i=0}^{S^2} \mathbb{I}_i^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2 \quad (5)$$

$$+ \lambda_{\text{f}} \frac{(b - \hat{b})}{b}. \quad (6)$$

The parts (1) and (2) are the terms to estimate person's localizatin. Parts (3) and (4) to estimate confidence in a given object. Part (4) is the classification loss to estimate the probability of a

given class. The last part, (5), in Equation 3.7 is used to estimate BFP. We have incorporate the idea of **finetuning**, and used the original weights of darknet as a starting point during training. The loss is computed every batch iteration to update all the weights.

The $\mathbf{w}_{bf}$, *i.e.* the weights conected to the new regression output $\hat{\mathbf{Y}}_{bf}$ are initialized using the propositions of Glorot e Bengio ():

$$\mathbf{w}_{bf} \sim U \left[-\frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}} \right], \quad (3.8)$$

wherein $U[a, a]$ is the uniform distribution in the interval (a, a) and n is the size of the previous layer, in this case $n = 173056$. This means the samples are drawn from a uniform distribution that relies on the limits of U . Also, Glorot e Bengio () suggests that all biases are initialized to 0.

Various important concepts for efficiently training large deep neural networks are of described (HE *et al.*,). We have used **learning rate warmup** of Goyal *et al.* (2017), which consists in using a few batches at first to linearly increase the learning rate from 0 to the initial learning rate set before training. In our case, the first 2 epochs are used for warmup and the initial learning is $\eta_i = 1e^{-4}$. Therefore, during warmup the learning is $\eta_k = k\eta_i/m$, in which $m = 2 * N$ is the number of warmup steps, N is the training set size, in respect to $1 \leq k \leq m$.

And a **cosine learning rate decay** is used after warmup procedure (LOSHCHILOV; HUTTER, 2016), because it has demonstred to enhance accuracy and led to better tranfer learning performance in other applications such as semantic segmentation (HE *et al.*,). Within the t -th batch we decay the learning as follows:

$$\eta_t = \eta_f + \frac{1}{2}(\eta_i - \eta_f) \left(1 + \cos \left(\frac{t\pi}{T} \right) \right), \quad (3.9)$$

where $\eta_f = 1e^{-6}$, the final learning rate, T is the total number of bathes, ignoring warmup stage.

As optimizer we have used ADAM (KINGMA; BA, 2014), which is widely know and used in deep learning applications. The idea behind ADAM is that it combines two gradient descent methodologies: momentum and Root Mean Square Propagation (RMSP). Thus, the equation to update the tensor of parameters $\mathbf{W}$ in the t -th batch is

$$\mathbf{W}_t = \mathbf{W}_{t-1} - \eta_t \frac{\widehat{m}_t}{(\sqrt{\widehat{v}_t} + \epsilon)}. \quad (3.10)$$

Regarding Equation 3.10, $\widehat{m}_t$ and $\widehat{v}_t$ are the first and second momentum estimates respectively. For more details how these parameters are estimate, please refer to the original paper (KINGMA; BA, 2014).

The criteria used to stop training is the number of epochs. When the training reaches 1000 epochs it stops and the chosen weights are the one that reaches the lowest loss in the validation dataset. As an example the following output given by ShapedNet is shown in Figure 24. The person is localized and its body fat percentage estimated.

Moreover, a visualization of the ShapedNet in two dimensions is shown in Figure 24. We create a heatmap showing regions that are most likely to be related to the body fat percentage estimations of our method. Regions with hot color intensity have a heavy bias to calculate $\mathbf{Y}_{bf}$, *i.e.* the body fat. And, regions with cold color intensity are the counterfactual explanations for the output. This visualization not only provide a 2-D visual map of fatness likelihood, but also explains in a compressed manner what is going on under this extremely deep neural network that makes it possible to estimate human body fat percentage. This technique is named **DBF-SCAN**.

Figure 24 – Input photo in (a), ShapedNet output in (b) and output of DBF-SCAN in (c).



Source: Elaborated by the author.

3.2 DBF-SCAN: Visual representation of the distribution of body fat in human body

Creating visual constructs from computer vision AI-based models, such as CNN, is a challenging area that has been addressed by the scientific community for the last years. Although the higher challenge, we designed a method that can address the following aspects: (1) Interpretability, which can help us to build trust in the AI agent used on par with humans, wherein the goal is to establish reliability in its usage (SELVARAJU *et al.*, 2019); (2) Aid clinical

routines, by generating a visual map that could give the professional a better guiding in nutrition protocols; And last but not least, (3) higher level information compression, since we modeled a 2-D compressed signal that has a mapping to body fatness. Our work is an extension of the method presented by Selvaraju *et al.* (2019) and its basis are detailed below.

First, let us define the output $y^{(d)}$, which is the output of a neuron in the d -th convolutional layer,

$$y_{i,j,k}^{(d)} = b_k + \sum_{m=1}^M \sum_{n=1}^N \sum_{c=1}^C w_{m,n,c,k} \cdot x_{i+m-1,j+n-1,c} \quad (3.11)$$

In this equation,

- $y_{i,j,k}^{(d)}$ is the output of the neuron at position (i, j) in the k -th feature map of the convolutional layer;
- M and N are the height and width of the receptive field of the neural network and C is the number of feature maps in the previous layer (or channel C if the previous layer is the input layer);
- $x_{i+m-1,j+n-1,c}$ is the input value from the c -th feature map/channel at position $(i+m-1, j+n-1)$ in the given input;
- $w_{m,n,c,k}$ is the weight of the connection from the c -th channel at position (m, n) in the filter to the k -th feature map, and b_k is the bias term for the k -th feature map.

Let us call $y_{i,j,k}^{(d)}$ as the **feature map activations**. In our case, $d = 58$ which is the output of the 58th layer of our network. Since, this is a deep neural network the input $x = f(BN(y^{(57)}))$, wherein $f(\cdot)$ is the Leaky ReLU activation function and $BN(\cdot)$ is the batch normalization operation. Moreover, let us place the proper dimensions in all terms of Equation 3.11. $M = 3$, $N = 3$, $C = 512$. Therefore, we rewrite the equation as follows,

$$y_{i,j,k}^{(58)} = b_k + \sum_{m=1}^3 \sum_{n=1}^3 \sum_{c=1}^{512} y_{m,n,c,k}^{(57)} \quad (3.12)$$

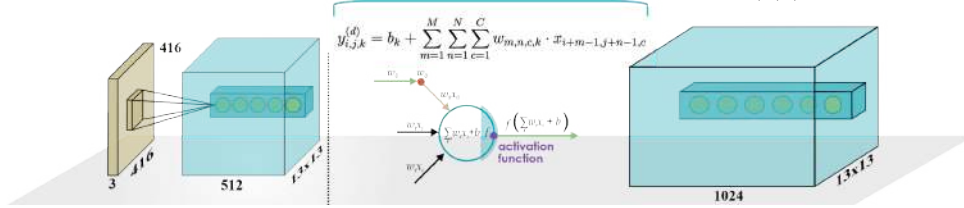
wherein $k \in \mathbb{Z}$, $1 \leq k \leq 1024$ and $(i, j) \in \mathbb{Z}^2$, being $1 \leq i \leq 13$, $1 \leq j \leq 13$ and $u = v = 13$. Therefore,

$$\mathbf{y}^{(58)} \in \mathbb{R}^3 \Rightarrow \mathbf{y}^{(58)} \in \mathbb{R}^{u \times v \times k} \Rightarrow \mathbf{y}^{(58)} \in \mathbb{R}^{13 \times 13 \times 1024}, \quad (3.13)$$

$\mathbf{y}^{(58)}$ is a 3-D tensor that contains all neuron activation $y_{i,j,k}^{(58)}$ in respect to both body fat percentage estimate, $\hat{\mathbf{Y}}_{bf}$, and person's localization coordinates, $\hat{\mathbf{Y}}_{loc}$. Note that the dimensions $13 \times 13 \times 1024$ were inherited from the semantic segmentation model used as the basis of our work.

The process described above is shown in Figure 25. In the most left part is the input image, that after passing through 56 layers it gets into the 57th layer in which we use $y_{m,n,c,k}^{(57)}$, *i.e.* the output of the left blue cuboid, to calculate $y_{i,j,k}^{(58)}$, *i.e.* the input of DBF-SCAN the right cuboid object in Figure 25. Moreover, the cuboid is modeled here as the 3-D tensor, $\mathbf{y}^{(58)}$.

Figure 25 – Cuboid transformation to create the tensor $\mathbf{y}_{m,n,c,k}^{(57)}$.



Source: Elaborated by the author.

Thus, in order to obtain the a fatness-likelihood map $L_{BF-SCAN}^r \in \mathbb{R}^{u \times v}$ of width u and height v for any regression output r , we first compute the gradient of a given regression output r , y_r , with respect to feature map activations $y_{i,j,k}^{(d)}$ of a convolutional layer. The gradients are also global averaged pooled over dimensions u and v , as suggested by Selvaraju *et al.* (2019). Thus we calculate,

$$\alpha_r^{(d)} = \overbrace{\frac{1}{Z} \sum_i \sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y_r}{\partial y_{i,j,k}^{(d)}}}_{\text{gradients via backprop}} \quad (3.14)$$

$\alpha_r^{(d)}$ are the neural importance weights of the r -th output in respect to the d -th feature map activation. Note this procedures represents a single forward and backward pass in the neural network, so that we can be able to capture the behavior of the gradients, which are backpropagated until reaches the given d -th layer.

We agree that the weights in $\alpha_r^{(d)}$ represents a *partial linearization* of the deep neural network, as told by Selvaraju *et al.* (2019) before. Since it provides values of “importances” in the feature maps the are likely to be relevant to the target regression output, *i.e.* the body fatness. The term *partial linearization* refers that due to the nature of the deep neural network

(*e.g.* fully connected neurons) is hard to isolate neuron weights and activations from the other, so it is possible that part of relevant information are propagated to other features maps/layers and neurons.

Thus, let us rewrite Equation 3.14 as,

$$\alpha_{\hat{\mathbf{Y}}_{bf}}^{(58)} = \frac{1}{Z} \sum_i \sum_j \frac{\partial y_{\hat{\mathbf{Y}}_{bf}}^{(58)}}{\partial y_{i,j,k}^{(58)}} \quad (3.15)$$

given $d = 58$ (the 58th convolutional layer) in respect to the regression output of the body fat estimate, $\hat{\mathbf{Y}}_{bf}^{(58)}$. The term Z refers to the number of spatial locations or feature maps in the final convolutional layer activations. Here, $\alpha_{\hat{\mathbf{Y}}_{bf}}^{(58)}$ represents the importance weight for each spatial location, and the summation $\sum$ is performed over Z spatial locations or feature maps. Each gradient of the target class score with respect to the activations at the i -th spatial location and k -th feature map. Dividing the sum by Z calculates the average gradient for each feature map.

By considering the average gradient across all spatial locations, the importance weights $\alpha_{\hat{\mathbf{Y}}_{bf}}^{(58)}$ capture the relative contribution of each feature map to the target class prediction. These weights are then used to weight the final convolutional layer activations in the subsequent equation.

The operation made in Equation 3.15 performs a **compression**, it reduces $i = 13 \Rightarrow i = 1$ and $j = 13 \Rightarrow j = 1$ so,

$$\alpha_{\hat{\mathbf{Y}}_{bf}}^{(58)} \in \mathbb{R}^2 \Rightarrow \alpha_{\hat{\mathbf{Y}}_{bf}}^{(58)} \in \mathbb{R}^{1 \times 1024}, \quad (3.16)$$

Then, we perform a weighted combination of feature maps,

$$L_{\text{BF-SCAN},i,j}^r = f_t \left(\sum_r y_{i,j,k}^{(d)} \alpha_{i,j,r}^{(d)} \right) \quad (3.17)$$

$$L_{\text{BF-SCAN}}^r = f_t \left(\mathbf{y}^{(d)} \alpha_{\hat{\mathbf{Y}}_{bf}}^{(d) \top} \right) \quad (3.18)$$

let us define $\mathbf{H}_r$, a **heatmap matrix**, to be the weighted combination,

$$\mathbf{H}_r^{(d)} = \mathbf{y}^{(d)} \alpha_{\hat{\mathbf{Y}}_{bf}}^{(d) \top}. \quad (3.19)$$

Finally, DBF-SCAN computes the scores as,

$$L_{\text{BF-SCAN}}^r = f_t \left(\mathbf{H}_r^{(d)} \right) = f_t \left(\mathbf{H}_r^{(58)} \right). \quad (3.20)$$

wherein $f_t(\cdot) \in [0, 1]$ is the normalizing function,

$$f_t(\mathbf{H}_r^{(d)}) = \frac{\mathbf{H}_r^{(d)} - \mathbf{H}_r^{(d)}_{\min}}{\mathbf{H}_r^{(d)}_{\max} - \mathbf{H}_r^{(d)}_{\min}}. \quad (3.21)$$

In short, the full Equation to generate the visual representation DBF-SCAN

$$L_{\text{BF-SCAN}}^r = f_t \left[\mathbf{y}^{(58)} \left(\frac{1}{Z} \sum_i \sum_j \frac{\partial y_{\hat{\mathbf{Y}}_{bf}}^{(58)}}{\partial y_{i,j,k}^{(58)}} \right)^{\top} \right] \quad (3.22)$$

$$L_{\text{BF-SCAN}}^r = f_t \left[\underbrace{\mathbf{y}^{(58)}}_{58^{\text{th}} \text{ conv. layer}} \underbrace{\left(\frac{1}{Z} \sum_i \sum_j \frac{\partial y_{\hat{\mathbf{Y}}_{bf}}^{(58)}}{\partial y_{i,j,k}^{(58)}} \right)^{\top}}_{\alpha_{\hat{\mathbf{Y}}_{bf}}^{(58)}} \right] \quad (3.23)$$

and to summarize,

$$L_{\text{BF-SCAN}}^r \in \mathbb{R}^2 \Rightarrow L_{\text{BF-SCAN}}^r \in \mathbb{R}^{u \times v} \Rightarrow L_{\text{BF-SCAN}}^r \in \mathbb{R}^{13 \times 13} \quad (3.24)$$

$$\mathbf{H}_r^{(58)} \in \mathbb{R}^2 \Rightarrow \mathbf{H}_r^{(58)} \in \mathbb{R}^{u \times v} \Rightarrow \mathbf{H}_r^{(58)} \in \mathbb{R}^{13 \times 13}, \quad (3.25)$$

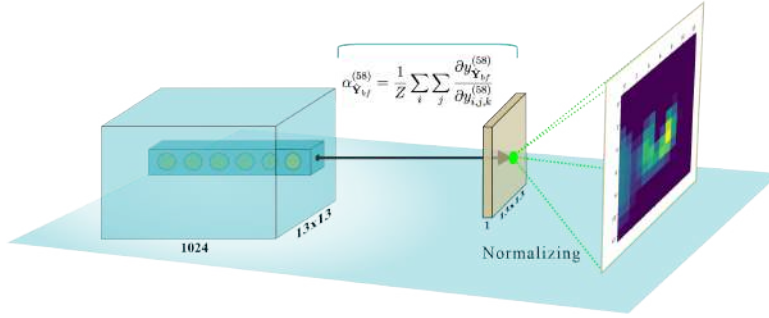
$$\alpha_{\hat{\mathbf{Y}}_{bf}}^{(58)} \in \mathbb{R}^2 \Rightarrow \alpha_{\hat{\mathbf{Y}}_{bf}}^{(58)} \in \mathbb{R}^{1 \times k} \Rightarrow \alpha_{\hat{\mathbf{Y}}_{bf}}^{(58)} \in \mathbb{R}^{1 \times 1024}, \quad (3.26)$$

$$\mathbf{y}^{(58)} \in \mathbb{R}^3 \Rightarrow \mathbf{y}^{(58)} \in \mathbb{R}^{u \times v \times k} \Rightarrow \mathbf{y}^{(58)} \in \mathbb{R}^{13 \times 13 \times 1024}. \quad (3.27)$$

The process of to calculate $L_{\text{BF-SCAN}}^r$ is summarized in Figure 26. Using $\mathbf{y}^{(58)}$, *i.e.*, the compression made by $\mathbf{H}_r^{(58)}$ is shown as the yellow cuboid in Figure 26, which is 13×13 .

Note that this after the normalizing function $f(\cdot)$ the output shown is a coarse map that shows pixels that have a higher influence to the regression output r , in our case, $\hat{\mathbf{Y}}_{bf}$. Brighter regions reflect higher influence to $\hat{\mathbf{Y}}_{bf}$ and darker regions lower influence.

Figure 26 – Visual representation of the compression made by the partial linearization of $\alpha_{\hat{\mathbf{Y}}_{bf}}^{(58)}$.



Source: Elaborated by the author.

Let us represent $L_{\text{BF-SCAN}}^r$ as a 13×13 **heatmap matrix**,

$$L_{\text{BF-SCAN}}^r = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,13} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,13} \\ \vdots & \vdots & \ddots & \vdots \\ h_{13,1} & h_{13,2} & \cdots & h_{13,13} \end{bmatrix}_{13 \times 13} \quad (3.28)$$

each value $h_{u,v}$ is **mapping** a region in the input image X . The input dimensions for ShapedNet is 416, then each value of maps a regions of size 32×32 . For that reason we say the DBF-SCAN does both compression and mapping in respect to the person's body fat.

This process is shown in Figure 27. Each pixel in the heatmap matrix is tied up to an square in the original image of 32×32 pixels. In the Figure 27 (a) 13×13 grid is overlapped in the input image and each single part of that grid is a body region, Figure 27 (b) with 32×32 pixels. Moreover, the whole visual process of DBF-SCAN is shown in Figure 28. An input image is subjected through a series of transformations and its gradient heatmap matrix are used to generate the output shown in the most right part of Figure 28. The heatmap, represented by the $L_{\text{BF-SCAN}}^r$ matrix, is overlaid on the input image. The central region of the heatmap is particularly relevant for body fat estimation. It is important to note that the body contour of the individual is used as a mask to exclude other region from the photo.

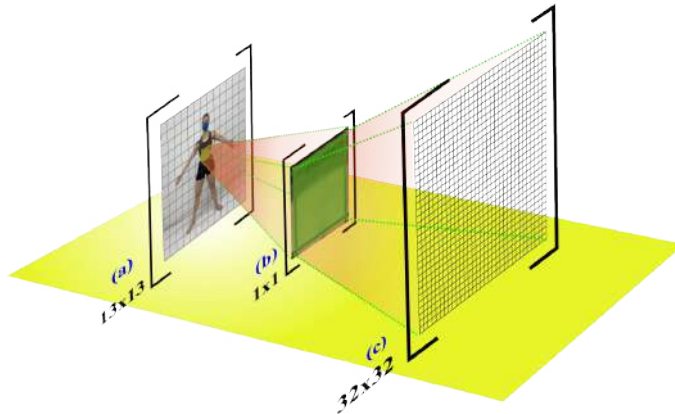
During training and optimization of the loss function, $\mathbf{L}$ in Equation 3.7, it is expected that the heatmap matrix $L_{\text{BF-SCAN}}^r$ changes over each batch and epoch. Moreover, $h_{u,v}$

values in $\mathbf{H}_r^{(d)}$ are constantly changing since it is a function of $\mathbf{y}^{(58)}$ and optimized in respect to $\hat{\mathbf{Y}}_{bf}$, as shown in Equation 3.23. Therefore, the heatmap also changes accordingly.

In Figure 29 it is shown how $L'_{\text{BF-SCAN}}$ changes during training. Brighter colors indicates pixels that are most relevant to body fat estimation. The first epochs during training the error in the neural network is much higher than during final stages. Up to 150 epochs the heatmap has not change much, Figure Subfigure 29a. At early stages of training the relevant pixels are close to each other in most left part of the map, and this might be due to random weights initialization, which are far from a local/global minima. These regions seems to have little relevance to estimate the person's body fat, *i.e.* they are not regions closes to the human body. As the training continues to middle stages, such as in Figure Subfigure 29b relevant pixels are moving towards the center of the map, this might be due to the regions where the person is located in the photo. Note that in Figure Subfigure 29c, for example, the brighter region is already close to the center, however the heatmap still has a lot of imprecise information in the left region. Due to the stochastic process of training there might be a little noise towards the gradient, therefore brighter regions vary a bit, note Figure Subfigure 29d and Figure Subfigure 29e despite being 250 epochs apart the most relevant regions is drifting towards the center.

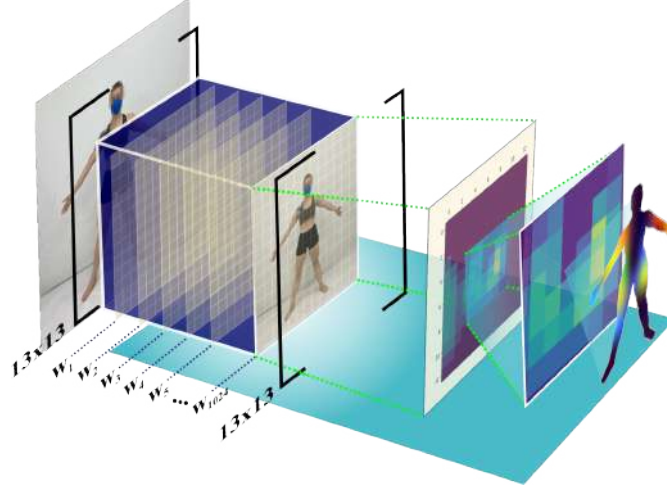
After 500 epochs the heatmap converged to highlight its center, which is the most relevant region in respect to $\hat{\mathbf{Y}}_{bf}$. See the heatmap after 700 epochs in Figure Subfigure 29e, at this stage the training was stopped since the error rate has not change more than its preset tolerance.

Figure 27 – Mapping region of input space, X and heatmap space, $L'_{\text{BF-SCAN}}$.



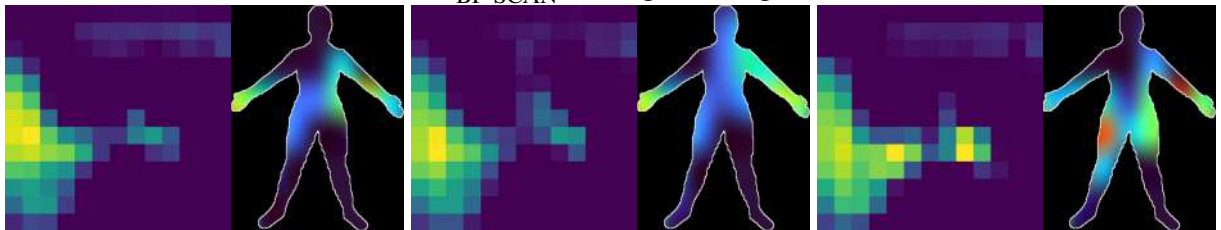
Source: Elaborated by the author.

Figure 28 – Application of the heatmap matrix in the partial linearization of an input image.

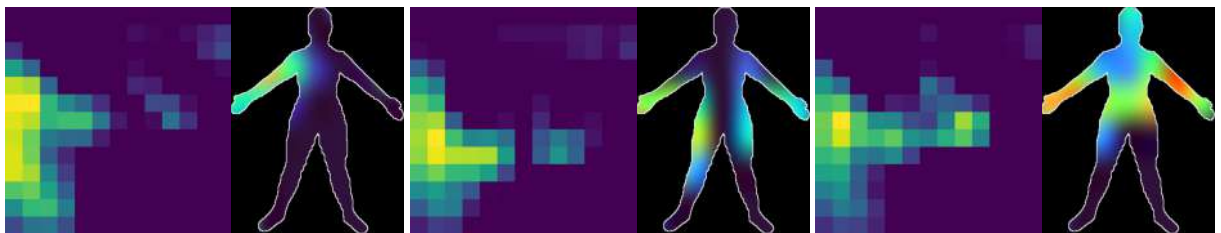


Source: Elaborated by the author.

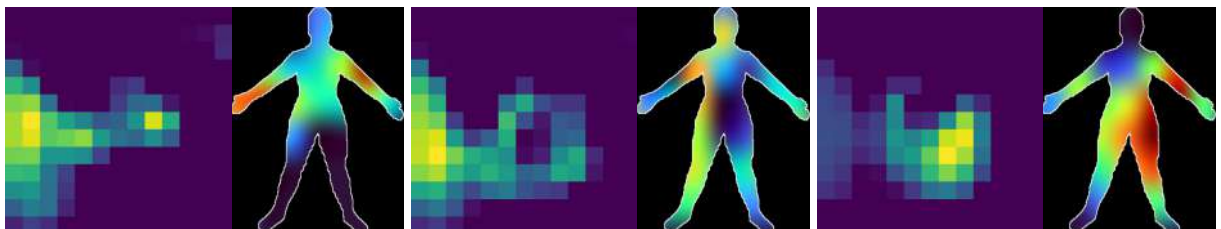
Figure 29 – Dynamic Visualization of the evolution of the DBF-SCAN and heatmap matrix, $L_{BF-SCAN}^r$, during Training.



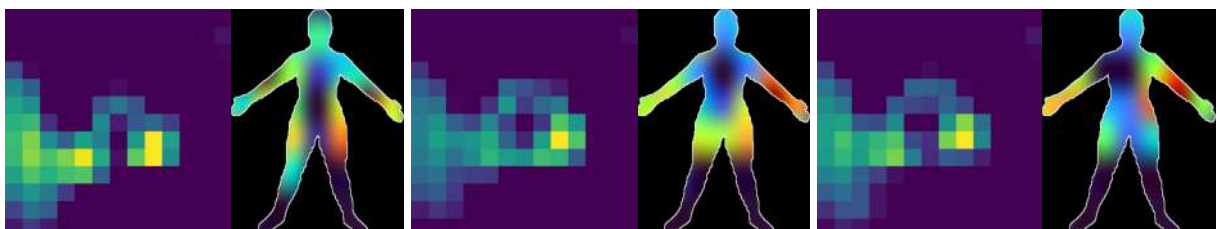
(a) ~50 to 150 epochs



(b) ~200 to 400 epochs



(c) ~450 to 550 epochs



(d) ~600 to 700 epochs

3.3 Database and Experiments

Creating databases for body composition assessments is a costly and time-consuming process. In previous studies (UÇAR *et al.*, 2021; CHIONG *et al.*, 2021; SHAO, 2014), researchers have reported the use of public databases as an alternative. One commonly used resource is the StatLib database, which contains anthropometric measurements for 252 men, including age, weight, height, and ten body circumferences. However, it should be noted that this dataset does not include data for female individuals. Furthermore, the StatLib database uses Siri's equation (SIRI,) as the reference method for body fat estimation, rather than DXA, and does not provide photographs of the subjects.

To address this gap, we conducted a comprehensive data collection effort involving anthropometric measures, photographs, and DXA exams from both male and female individuals. This collaborative effort involved research groups from the Federal University of Rio de Janeiro (UFRJ) and the Federal Institute of Ceará (IFCE). The clinical trials and examinations took place at the Josue Carlos Nutrition Institute at UFRJ, while the analysis system was implemented at IFCE. Ethical compliance was ensured by registering the dataset under the National Health Council through our university, following the necessary procedures for research involving human subjects. The study adhered to the guidelines and requirements of the Research Ethics Committee of the IFCE, protocol number 3593367. The project is registered in Plataforma Brasil under CAAE registration 16885519.1.0000.5589. Participants provided their informed consent for the use of the collected data in research, confirming compliance with the required parameters.

Our dataset is named BCDB23, it consists of 1273 samples, including 578 men and 695 women. It includes information such as age, height, weight and ten anthropometric measures. The images were obtained using a smartphone camera with a resolution of 1920×1080 pixels and 8-bit depth. Each individual's body was captured from frontal, back, right-side, and left-side perspectives. The dataset covers individuals aged between 18 and 65, with DXA values ranging from 9.30 to 57.60. This dataset is an extension of the one used by Alves (2022), which had 912 subject.

Table 1 displays a summary of the collected information from the 1273 participants. The data shows a high degree of variability, indicating a well-distributed dataset. Figure Subfigure 30 shows the distribution of our dataset, demonstrating a relatively balanced representation of males and females, with a slightly higher proportion of women. The majority of individuals in the dataset fall within the 25 to 35 age range. Interestingly, the distribution of samples across age

groups is more evenly distributed for women, whereas there is a noticeable underrepresentation of men between the ages of 40 and 50.

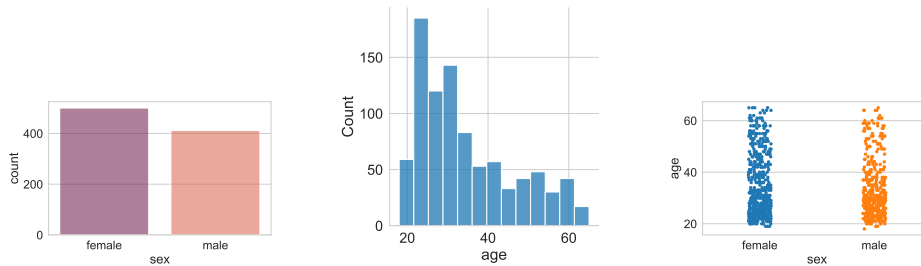
Table 1 – Dataset distribution in average, min, std, and max.

	mean	std	min	25%	50%	75%	max
Age	34.67	11.92	18.0	25.0	31.0	43.0	65.0
Height	168.3	9.27	144.0	161.1	168.0	174.8	199.0
Weight	73.43	16.62	30.0	61.0	71.7	82.93	143.5
Arm	31.94	4.61	20.65	28.4	31.92	35.0	48.55
Forearm	26.24	3.29	7.85	23.9	26.2	28.51	50.45
Waist	81.3	12.61	37.4	72.0	79.8	88.8	134.2
Hip	100.6	9.43	47.5	94.5	99.12	105.0	148.5
Thigh	54.29	6.38	28.0	50.15	53.55	57.55	98.35
Calf	37.26	3.55	24.55	34.8	36.95	39.4	53.7
DXA	31.21	9.59	9.3	24.3	31.15	38.3	57.6

Source: Elaborated by the author.

The distribution of DXA values by sex is exhibited in Figure Figure 31. It is evident that body fat distribution is relatively even within each sex. There is also a noticeable difference in the range of DXA values between males and females, indicating variations in body composition between the two genders.

Figure 30 – Dataset distribution.



(a) By sex

(b) By age

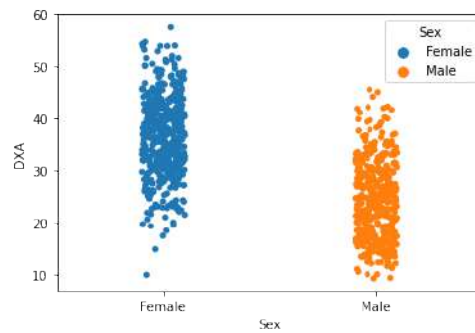
(c) By sex and age

Source: Adapted from Alves (2022).

Figure 32 presents a comparison of the dataset attributes by sex and their correlations. The analysis shows a positive correlation between weight and anthropometric measures, indicating that higher weight is associated with higher measures. There are also correlations between various measures, such as a positive correlation between thigh and arm measures.

When comparing the attributes by sex, it is observed that men generally have higher values for waist, forearm, arm, height, and weight attributes compared to women. However, the hip, calf, and thigh measures are relatively similar, with women having slightly higher hip

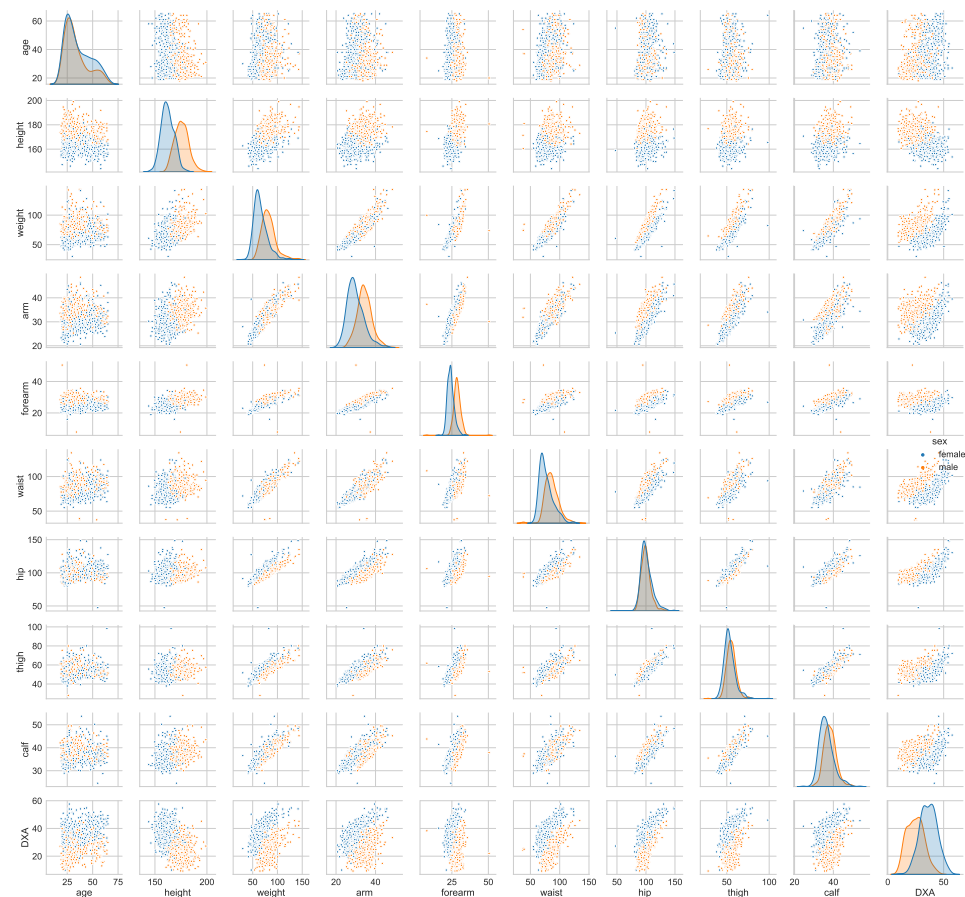
Figure 31 – Dataset DXA distribution by sex.



Source: Adapted from Alves (2022).

measures. Notably, the DXA values and their frequency are significantly higher for women compared to men, underscoring the importance of considering sex as an associated attribute for DXA analysis.

Figure 32 – Dataset attributes distribution regarding anthropometric measurements.



Source: Adapted from Alves (2022).

The dataset's correlation was determined using Pearson Correlation, and the results

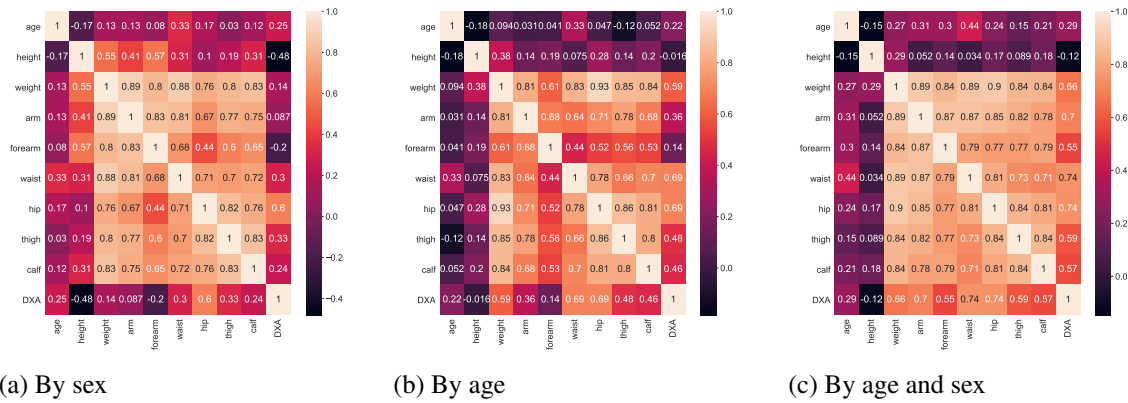
are presented in Figure 1. DXA exhibits an insignificant positive correlation ($0 \leq p < 0.2$) with arm and weight. However, DXA demonstrates a moderate positive correlation ($0.2 \leq p < 0.5$) with thigh, age, calf, and waist, as well as a strong positive correlation ($p \geq 0.5$) with hip, indicating that hip is a crucial attribute to consider. On the other hand, DXA displays a negative correlation with forearm and height, with both measures exhibiting a moderate correlation ($-0.5 < p \leq -0.2$).

The anthropometric measures show positive correlations with each other, indicating a proportional relationship within the body. However, age does not exhibit a significant correlation with the other attributes. Weight, on the other hand, demonstrates a moderate correlation with height and a strong correlation with all other anthropometric measures. However, the correlation between weight and DXA ($p = 0.14$) is found to be insignificant, suggesting that weight may not be a significant predictor of DXA. However, it could still be useful in correlating among the anthropometric measures themselves. This could be attributed to differences in body composition between men and women.

Figure 33 illustrates the correlation matrices of the dataset stratified by sex. It is important to note that the correlation between weight and body fat percentage increased significantly in both sexes compared to the mixed dataset. While weight can serve as a reliable indicator of high body fat percentage, it is worth considering that the dataset includes both muscular individuals and obese individuals. Muscular individuals may have high weight and low body fat, while obese individuals may have high weight and high body fat. This unique characteristic of the dataset may explain the lower-than-expected correlation between weight and DXA.

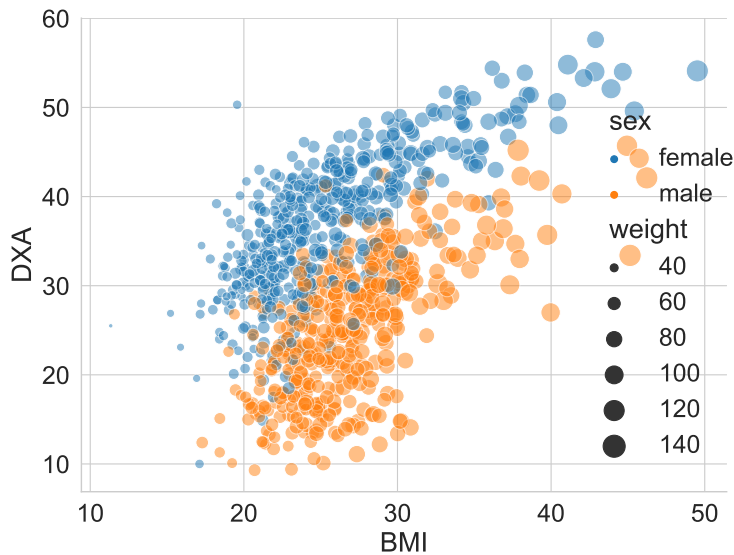
In Figure 34, a scatter plot depicting BMI versus DXA is presented. The plot reveals a general trend where higher BMI values are associated with higher body fat percentages. A similar pattern can be observed in weight. However, it is important to note that this trend is not universal. There are instances where individuals with higher BMI may have lower body fat percentages as estimated by DXA. This is especially the case when individuals have a higher proportion of fat-free mass. The blue highlighted cluster represents women, who typically have higher body fat percentages compared to men. The points that typically would be considered outliers in other domains are in fact an individual with a unique body composition, BMI and BFP.

Figure 33 – Correlation matrices divided by sex.



Source: Adapted from Alves (2022).

Figure 34 – Relation weight in DXA vs BMI.



Source: Adapted from Alves (2022).

3.3.1 Characteristic of BCDB23

This database is distinct and comprehensive compared to other databases reported in the literature. It includes four variations, allowing for analysis of different aspects in body composition assessment. The primary aspects we aim to analyze in our method are: (1) accuracy, which measures the difference between the model's estimated BFP and DXA; (2) precision or repeatability, which assesses the model's ability to produce consistent measurements under similar conditions; and (3) concordance, which examines whether the method's estimates in two different time periods align with DXA measurements.

Therefore, the BCDB23 was specifically designed to enable assessment of the

three aforementioned aspects. Figure 35 provides an overview of the database and its variations, allowing for a comprehensive understanding of its structure and purpose. The database comprises a total of 1273 human subjects, including 578 men and 695 women. The labels T_1 and T_2 represent two distinct time periods, where subjects visited the trial site to undergo DXA scans and have their photos taken. These time periods are spaced approximately three months apart. As an example, let's consider a woman who visited the trial site on May 31st for the initial exams (T_1). She would then return to the site three months later, on August 30th, to repeat all the exams and take new photos (T_2). 170 subjects returned for repeating exams (T_2), being 88 men and 82 women. The main focus is to evaluate concordance in estimating body fat percentage. It examines whether changes in BFP over time, such as fat gain or loss, are reflected consistently across different measurement methods. The second trial of exams T_2 is also referred as **follow-up assesment**.

The 1st and 2nd Sections are the round of exams that a person is subjected to it within the same day. Let us say a person arrive at the trial site, go through a DXA scan, BIA, Skilfolds and Photos, then a few minutes later the process is repeated. The goal here is to access precision, or repeatability, there should not be significantly differences among this estimates. This concept is also applied in T_2 as well.

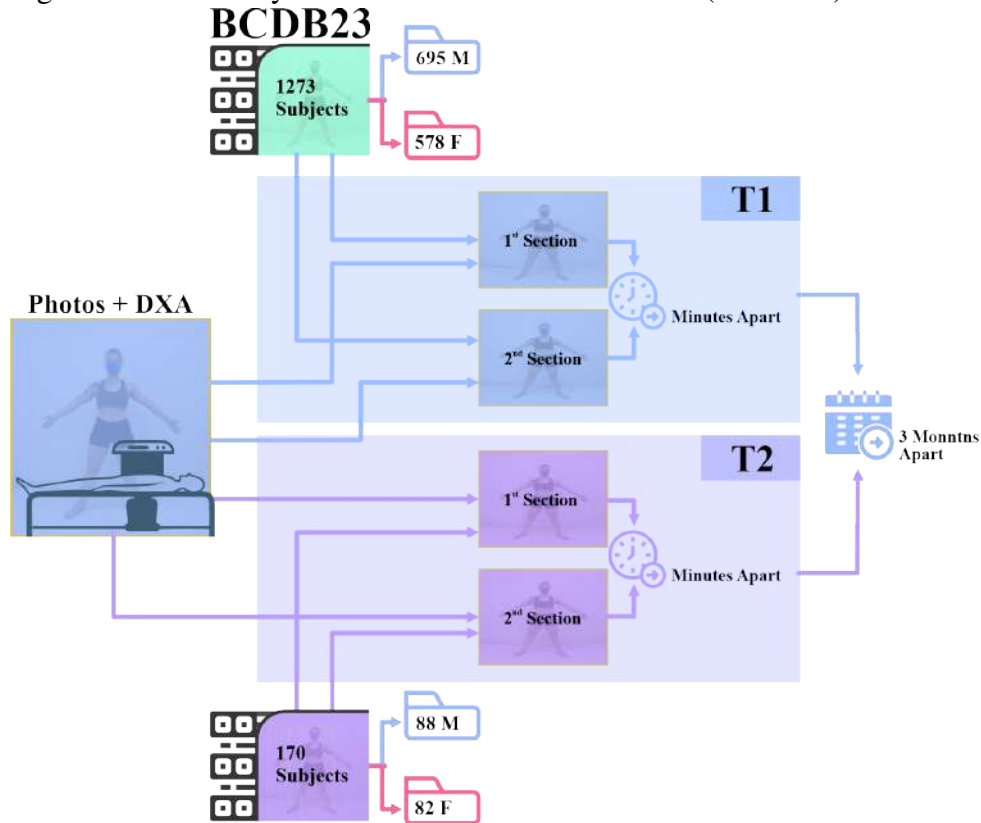
3.3.2 *Experimental setting*

In our experiments, we allocated 10% of the subjects for testing (127), 10% for validation (127), and 80% for training purposes (1019). The validation set was specifically used for early stopping during the training process. To ensure representative sampling, the subjects were stratified based on their Body Mass Index (BMI), a commonly used indicator to assess the relationship between body weight and height. The dataset was divided in a way that maintained the proportions of the six BMI categories across the data subsets.

The metrics used to compare performance throughout the experiments are the Mean Absolute Error (MAE), Mean Percentage Error (MAPE) and Mean Signed Deviation (MSD), calculated as follows:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (3.29)$$

Figure 35 – Summary of all details about the database (BCDB23).



Source: Elaborated by the author.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.30)$$

$$\text{MSD} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (3.31)$$

wherein y_i is the body fat percentage estimate from DXA and $\hat{y}_i$ from ShapedNet.

Moreover, we also report the Confidence Interval (CI) at 95% level, for each one of the metrics above. The confidence interval depends on the specific statistical test. In our case of estimating the mean of a population using a t -distribution with a given confidence level (95% or $\alpha = 0.05$), the formula for the confidence interval would be:

$$\text{CI} = \bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad (3.32)$$

where $\bar{x}$ is the sample mean, s is the sample standard deviation, n is the sample size, and $t_{\alpha/2}$ is the critical value from the t -distribution corresponding to the desired confidence level and

degrees of freedom. CI values are reported as $[CI_{min}, CI_{max}]$. We also calculate the spread of the interval, $v = CI_{max} - CI_{min}$. The smaller the spread, the better. The points estimated in the Confidence Interval are the range of values that that an error in prediction of ShapedNet is most likely to fall into. We also report the values of Person's Correlation, to compare if predictions ShapedNet are correlated to DXA. We proposed that the baseline metrics is $MAPE \leq 10\%$.

We also perform Tukey's Honestly Significant Difference (HSD) test as a *post-hoc* statistical test used to compare multiple group means. Its purpose is to identify which specific group means differ significantly from each other. The formula for Tukey's HSD test is as follows:

$$HSD = q \cdot \sqrt{\frac{MS}{n}} \quad (3.33)$$

wherein HSD is the Honestly Significant Difference, q is the critical value obtained from the Studentized range distribution table, MS is the mean square error, which is the estimate of the population variance within each group, and n is the number of observations in each group. This equation represents the minimum difference that must exist between two group means for them to be considered statistically significant.

Tukey's HSD test is a popular choice for *post-hoc* analysis as it controls the error rate, ensuring that the overall probability of making a Type I error (incorrectly rejecting a true null hypothesis) remains at the desired level. It provides a comprehensive way to compare all possible pairwise differences among group means.

The confidence is set to 95%, so for $p < .05$ there are enough evidences to reject the null hypothesis, H_0 , and accept the alternative hypothesis, H_1 . For $p \geq .05$ there are evidences to accept the null hypothesis H_0 . The following hypotheses ought to be tested:

Is a gender-based approach better than a non-gender based one?

H_0 : No significant differences are observed when using a gender-based approach or a non-gender-based approach for ShapedNet.

H_1 : Using a gender-based approach is in fact different than using a non-gender one to train the ShapedNet.

Is the method able to repeat measurements in similar conditions?

H_0 : There are no differences between ShapedNet errors in 1st and 2nd Sections, so the model is precise.

H₁: ShapedNet errors in 1st Section are in fact different than in the 2nd Sections, so the model is unprecise.

Is ShapedNet really different from a DXA exam?

H₀: There are no differences between ShapedNet and DXA. So they can be used interchangeably.

H₁: ShapedNet is different of DXA. So, they can be used respecting their constrains.

4 RESULTS AND DISCUSSIONS

This chapter presents a comprehensive analysis of the ShapedNet method, starting with a quantitative assessment of its accuracy. The first section focuses on evaluating the error against DXA as the gold standard, providing insights into the method's performance. The following section explores the precision of the first and second sections, examining the precision in measurements. The chapter then delves into the concordance evaluation of ShapedNet over time, considering follow-up assessments with DXA. Limitations of the ShapedNet approach are discussed to provide a comprehensive understanding of its capabilities. Additionally, a qualitative analysis of the DBF-SCAN method is presented, followed by an examination of its limitations. Finally, the chapter concludes with a comparison of ShapedNet to state-of-the-art (SOTA) methods, highlighting its relative performance in the field.

4.1 A quantitative analyses of ShapedNet results

We conducted three different experiments to achieve various objectives. The first experiment aimed to assess accuracy by measuring the error against DXA and determining the superior approach between gender-based and non-gender-based methods. The second experiment focused on precision (or repeatability) by comparing the error in the 1st and 2nd Sections, specifically in the first photo section (T_1), to evaluate their equivalence. The third experiment aimed to evaluate concordance against DXA by analyzing the error in T_2 , where the control group returned three months after the initial trials.

4.1.1 Accuracy Assessment and Method Comparison: Evaluating Error against DXA

We evaluate our method on training, validation and test datasets, as shown in Table 2. The reported values are mean and standard deviation of MAE, MAPE and MSD in respect to DXA. The results for the training and validation sets are provided to assess the model's performance and mitigate overfitting. Each row in the table represents a different model training. The models "Male" and "Female" are gender-based models trained using data from individuals of the corresponding gender. Conversely, the non-gender model is not designed to consider gender-specific information, making it a more general approach.

ShapedNet achieved a mean absolute error (MAE) in body fat percentage of 1.59 for males and 2.03 for females. The corresponding MAPE values indicate that, on average, the

measurement error accounted for 7.04% and 5.98% of the total body fat estimation for males and females, respectively. Notably, the non-gender model outperformed the gender-specific models, demonstrating an MAPE of 4.91% (1.42) in error.

The MAE values in the test set are comparable to those in the validation set, indicating that the model did not overfit during training. For example, the MAE for males was 1.59 in the test set and 1.73 in the validation set, while for females it was 2.03 in the test set and 2.25 in the validation set. The non-gender model exhibited a slightly larger difference between the sets, particularly in the train set. The average error during training for this model was 0.14, however it is not an indication of overfitting since errors in validation (1.51) and test (1.42) sets are similar. This is actually promising, as it suggests that a single model performs better than gender-specialized models.

However, the reported values of MSD are interesting. The female model tends to undermeasure the body fat percentage in women. For instance, the model tends to estimate the body fat percentage -0.21, on average, in women. And this behavior happens in all sets of data. This might be due to factors such as body fatness density location in women are different than in men. Moreover, it is well-known that women store more fat in the gluteal-femoral region than men and we are using only the front photo to estimate BFP, so it is possible that the model is not considering different angles of the person. And, this is an investigation that ought to be addressed in the future.

It is important to pay attention to the confidence intervals presented in Table 3. These intervals were estimated with a significance level of $\alpha = 0.05$, indicating that for 95% of the subjects, the error in body fat percentage estimation lies between 5.89% and 8.19% for the male group. Another important consideration is the spread, denoted as $v = CI_{\max} - CI_{\min}$. A narrower spread indicates less uncertainty in our estimates. For the male, female, and non-gender models, the spreads are $v_m = 2.3\%$, $v_f = 1.72\%$, and $v_{ng} = 1.8\%$, respectively. These values are considered acceptable, as they are below the threshold of 10%, as discussed in Section 3.3.

The negative mean signed difference values observed in the female group are further supported by the confidence intervals presented in Table 3. The lower bound value of -6.53% indicates that the error in the female group is predominantly negative, suggesting that the estimated body fat percentage (BFP) using ShapedNet is consistently lower than the DXA measurements for females. This reaffirms that the female BFP is being underestimated compared to DXA.

Let us remind that we have risen a few research questions in subsection 3.3.2, the first one was: *Is a gender-based approach better than a non-gender based one?* and the following hypothesis were draw.

H₀: No significant differences are observed when using a gender-based approach or a non-gender-based approach for ShapedNet.

H₁: Using a gender-based approach is in fact different than using a non-gender one to train the ShapedNet.

A non-gender-based model appears to outperform gender-based approaches, suggesting that a single model could be utilized instead of two during production, resulting in a more efficient system. This finding contrasts with other gender-based approaches that have been traditionally used in clinical procedures. However, it is crucial to determine if these approaches are indeed different. To investigate this, a Tukey's HSD paired test was conducted, and the results are presented in Table 4. A significance level of $p < 0.05$ was chosen to reject the null hypothesis, indicating the presence of statistically significant differences among the groups. The absolute error values were used as variables in the tests, enhancing their statistical power.

As a result, the null hypothesis **H₀** is rejected, indicating that the non-gender model's estimations are indeed different from those of the male or female models in terms of BFP calculations. Consequently, it is possible to conclude that the non-gender model has a lower mean absolute error compared to using a single model, with a difference of 0.17 (1.59 – 1.42) for males and 0.61 (2.03 – 1.42) for females. Additionally, the percentage error of the non-gender-based model is 2.13 points lower for males (7.04% – 4.91%) and 1.07% lower for females (5.98% – 4.91%) compared to gender-based models.

Table 2 – Overall results on test, validation, and train sets. The metrics are MAPE, MAE, and MSD, and the groups of analyses are Male, Female, and Non-gender.

Gender/Metric	MAPE (%)	MAE	MSD
Test			
Male	7.04 ± 6.43	1.59 ± 1.38	0.38 ± 2.08
Female	5.98 ± 5.05	2.03 ± 1.58	-0.21 ± 2.57
Non-gender	4.91 ± 6.78	1.42 ± 1.82	0.28 ± 2.29
Validation			
Male	7.23 ± 6.50	1.73 ± 1.52	0.45 ± 2.26
Female	6.56 ± 6.46	2.25 ± 1.92	-0.08 ± 2.96
Non-gender	5.15 ± 6.01	1.51 ± 1.67	0.05 ± 0.06
Train			
Male	4.64 ± 4.21	1.10 ± 0.97	0.28 ± 1.44
Female	3.90 ± 3.36	1.35 ± 1.06	-0.38 ± 1.68
Non-gender	0.52 ± 2.01	0.14 ± 0.48	0.03 ± 0.50

Source: Elaborated by the author.

Table 3 – Error Analysis with 95% Confidence Intervals: MAPE, MAE, and MSD.

	MAPE (%)	MAE	MSD
Male	[5.89, 8.19]	[1.34, 1.84]	[0.01, 0.75]
Female	[5.12, 6.84]	[1.76, 2.30]	[-0.65, 0.22]
Non-gender	[4.01, 5.81]	[1.18, 1.66]	[-0.01, 0.59]

Source: Elaborated by the author.

Table 4 – Results of the Tukey's HSD hypothesis test. MD denotes Mean Differences and the values of p highlighted in bold represent the groups are significantly different at the 95% confidence level

Variable	Groups		MD	p
MAE	Non-gender	Male	-1.0273	.001
	Non-gender	Female	-1.4612	.001

Source: Elaborated by the author.

4.1.2 Precision Evaluation of the First and Second Sections: Examining Equivalence in Error Measurements

The experiments conducted aim to assess the precision and repeatability of the model. The findings presented in Table 5 compare the measurements obtained from the first and second

photo sessions. It is important to note that these sessions took place on the same day, nearly at the same time, with a 30-minute interval between them. Therefore, any significant differences in the measurements are not expected. The second research question posed in subsection 3.3.2 was: "Is the method able to repeat measurements under similar conditions?" and the corresponding hypotheses were formulated.

H₀: There are no differences between ShapedNet errors in 1st and 2nd Sections, so the model is precise.

H₁: ShapedNet errors in 1st Section are in fact different than in the 2nd Sections, so the model is unprecise.

The reported errors in Table 5 are categorized based on the metrics of MAPE, MAE, and MSD. The rationale behind this categorization is that the MAPE values for the first photo session are equivalent to those of the second session, and the same applies to the other metrics. As discussed earlier, the non-gender model consistently exhibits lower MAPE and MAE values compared to the other models, and this pattern holds true for both the first and second photo sessions. Interestingly, the MAPE values are observed to be higher in males than in females in both sessions. It is worth noting that the MAE values in females are higher than those in males, despite the lack of this trend in the MAPE metrics. This suggests that women tend to have a higher body fat percentage than men, as mentioned previously. Additionally, the female model slightly underestimates the BFP in both photo sessions based on the MSD metric.

As shown in Table 5 the mean differences of MAPE and MAE between the first and second photo sessions, are 0.17% (-.08) for males, 0.38% (.13) for females, and 0.62% (.20) for the non-gender model. Although these differences are below 1% and may not be considered relevant, we have employed a statistical test and the results are presented in Table 6. For each metric, the groups being tested are the 1st and 2nd photo sessions, and there is no evidence to suggest that these groups are different from each other. Therefore, we accept the null hypothesis **H₀**. They are equivalent and the model is precise.

As shown in Table 5, the mean differences of MAPE and MAE between the first and second photo sessions are 0.17% (-0.08) for males, 0.38% (0.13) for females, and 0.62% (0.20) for the non-gender model. Although these differences are below 1% and may not be considered significant, a statistical test was conducted, and the results are presented in Table 6. The test compared the groups of the first and second photo sessions for each metric, and there is no evidence to suggest that these groups differ from each other. Therefore, the null hypothesis **H₀** is

accepted, indicating that the photo sessions are equivalent and the model demonstrates precision.

In Table 7, the confidence intervals for the first and second photo sessions are reported. Additionally, the spread values, v , for each model are presented. It is observed that all models meet the acceptable threshold of $v \leq 5\%$. Notably, the spread values in both sessions are quite similar, with v_{1st} being equivalent to v_{2nd} . Furthermore, it is observed that the lower bounds in the first session are similar to those in the second session, and the same trend is observed for the upper bounds as well.

The results obtained from the gender and non-gender designed models of ShapedNet indicate that the model is precise and does not exhibit any bias or overfitting towards any specific section. This conclusion is further supported by the consistent and equivalent results obtained from different photo sessions.

Table 5 – Comparison of Errors MAE, MAPE and MSD for the first and second Sections in T1.

Gender/metric	MAPE (%)	MAPE (%)	MAE	MAE	MSD	MSD
	1 st Session	2 nd Session	1 st Session	2 nd Session	1 st Session	2 nd Session
Male	7.04 ± 6.43	6.87 ± 6.14	1.59 ± 1.38	1.67 ± 1.47	0.38 ± 2.08	0.37 ± 2.20
Female	5.98 ± 5.05	6.36 ± 5.68	2.03 ± 1.58	2.13 ± 1.71	-0.21 ± 2.57	-0.04 ± 2.74
Non-gender	4.91 ± 6.78	5.53 ± 6.12	1.42 ± 1.82	1.62 ± 1.68	0.28 ± 2.29	0.32 ± 2.32

Source: Elaborated by the author.

Table 6 – Results of the Tukey's HSD hypothesis test in first and second session. p is highlighted to represent the groups that are significantly different at the 95% confidence level.

Approach	Groups		MD	p
	1 st Session	2 nd Session		
Male	MAPE		-.0017	.835
	MAE		.0842	.650
	MSD		-.0121	.965
Female	MAPE		.0039	.558
	MAE		.0992	.625
	MSD		.1684	.606
Non-gender	MAPE		.0063	.313
	MAE		.2011	.233
	MSD		.0359	.871

Source: Elaborated by the author.

Table 7 – Comparison of Error Analysis for the first and second Sections in T1, with 95% Confidence Intervals: MAPE, MAE, and MSD.

	CI MAPE (%) 1 st Session	CI MAPE (%) 2 nd Session	v_{1st}	v_{2nd}
Male	[5.89, 8.19]	[5.75, 7.98]	2.30	2.23
Female	[5.12, 6.84]	[5.41, 7.33]	1.72	1.92
Non-gender	[4.01, 5.81]	[4.72, 6.35]	1.80	1.62
	CI MAE 1 st Session	CI MAE 2 nd Session	v_{1st}	v_{2nd}
Male	[1.34, 1.84]	[1.41, 1.94]	0.50	0.53
Female	[1.76, 2.30]	[1.83, 2.42]	0.54	0.59
Non-gender	[1.18, 1.66]	[1.39, 1.84]	0.48	0.45
	CI MSD 1 st Session	CI MSD 2 nd Session	v_{1st}	v_{2nd}
Male	[0.01, 0.75]	[-0.02, 0.77]	0.74	0.79
Female	[-0.65, 0.22]	[-0.51, 0.41]	0.87	0.92
Non-gender	[-0.01, 0.59]	[0.01, 0.63]	0.60	0.62

Source: Elaborated by the author.

4.1.3 Concordance Evaluation with DXA over Time with Follow-up Assessments

The experiments in this subsection aim to evaluate the correlation and direction of ShapedNet in comparison to DXA. A control group was selected, and their examinations were repeated three months after the initial ones. The correlation analysis focuses on assessing how closely ShapedNet aligns with DXA measurements. Meanwhile, the direction analysis examines the ability of ShapedNet to capture changes in body fatness over time. By considering both correlation and direction, we can gain insights into the performance and consistency of ShapedNet in estimating body fat percentage.

Let us recall some terms. T_1 and T_2 refer to the first and second exams, respectively, conducted at different time periods, specifically three months apart. Δ_{DXA} and Δ_{SpNet} represent the differences in body fat estimates between these two time points for DXA and SpNet, respectively. A positive Δ indicates an increase in body fat, while a negative Δ indicates a decrease. Additionally, we define δ as the product of Δ_{DXA} and Δ_{SpNet} , where a positive value indicates that DXA and SpNet exhibit the same direction of change, which is desirable for accurate predictions.

A summary of the metrics MAPE, MAE, and MSD comparing T_1 and T_2 is provided in Table 8. It is important to note that these groups consist of fewer subjects compared to other experiments due to the repetition of exams in a limited number of subjects. The MAPE error is

noticeably higher in all T_2 experiments. Specifically, the MAPE error is 4.47% for males, 2.93% for females, and 5.57% for the non-gender group. From a broader perspective, the MAPE metric holds greater significance. However, the differences become less apparent when considering only the MAE values, which are 0.58 for males, 0.94 for females, and 1.17 for the non-gender group. In terms of MSD, it is interesting to observe that in T_2 , the average metric is negatively centered across all groups, indicating potential underestimation of body fat by the model in all subjects. Overall, our method suggests that changes in body fat may not be as pronounced as indicated by DXA. The difference in MSD between the exams for female subjects (-0.2) is smaller compared to that of males (-0.86) and the non-gender group (-0.76).

It is indeed correct that statistical tests should not be used to directly compare the T_1 and T_2 groups since they represent different time periods with potential changes. However, it is appropriate to conduct a hypothesis test to determine if there is a significant difference between the two sections used in the study. In both T_1 and T_2 , the same protocol was followed, where a set of two photos was taken for each subject. The results of this test are presented in Table 9. In all groups, the p -value is greater than 0.05, indicating no significant difference between the two sections. Therefore, it is valid to conclude that the two sections are equivalent in terms of the measured variables.

Regarding the confidence intervals, comparisons were made between T_1 and T_2 , as shown in Table 10. It can be observed that for the metrics MAPE, MAE, and MSD, the spread v_{2nd} increases in all groups. This supports the findings mentioned earlier in Table 8. Specifically, the values of v_{2nd} more than double for the male, female, and non-gender models, indicating that the error in group T_2 is wider compared to T_1 . Additionally, the lower bound value for the non-gender model is 4.01 in T_1 , whereas it increases to 6.07 in T_2 . Similarly, the upper bound value for the non-gender model is 5.81 for T_1 , but it increases significantly to 12.66 for T_2 . It is important to note that the model was trained only using subjects from the T_1 experiment.

With respect to the MSD metric, it can be observed that the error in T_2 has shifted towards negative values compared to T_1 . For instance, the lower bound values of MSD in the female model are -1.92 in T_2 compared to 0.01 in T_1 . Similarly, in the male model, the lower bound values are -2.54 in T_2 compared to -0.65 in T_1 . For the non-gender model, the lower bound value in T_2 is -1.96. This suggests that ShapedNet tends to underestimate body fat percentage in T_2 experiments.

Table 8 – Comparison of MAPE, MAE, and MSD Errors over Time, with Follow-up Assessments (T_2).

Gender/metric	MAPE (%)	MAPE (%)	MAE		MSD	
	T_1	T_2	T_1	T_2	T_1	T_2
Male	5.75 ± 5.65	10.12 ± 8.69	1.22 ± 1.12	1.80 ± 1.50	0.32 ± 1.63	-0.54 ± 2.29
Female	4.37 ± 4.11	7.30 ± 6.82	1.45 ± 1.27	2.39 ± 1.78	-0.56 ± 1.85	-0.86 ± 2.87
Non-gender	2.78 ± 5.03	8.35 ± 8.35	0.74 ± 1.29	1.91 ± 1.78	0.11 ± 1.49	-0.65 ± 2.54

Table 9 – Results of the Tukey's HSD hypothesis test between the first and section of photos in the follow-up assesment. p highlighted in bold represent the groups that are significantly different at the 95% confidence level.

Approach	Groups		MD	p
	T_2 - 1 st Section	T_2 - 2 nd Section		
Male		MAPE	-.0138	.637
		MAE	.3961	.468
		MSD	-.386	.643
Female		MAPE	-.0077	.727
		MAE	-.2018	.724
		MSD	-.0976	.921
Non-gender		MAPE	.009	.7624
		MAE	-.0106	.982
		MSD	.0358	.961

Source: Elaborated by the author.

Table 10 – Comparison of Error Analysis with Follow-up Assessment, with 95% Confidence Intervals for MAPE, MAE, and MSD. $v = CI_{max} - CI_{min}$.

	CI MAPE (%) T_1	CI MAPE (%) T_2	v_{1st}	v_{2nd}
Male	[5.88, 8.19]	[6.35, 14.01]	2.31	7.66
Female	[5.12, 6.84]	[6.78, 12.79]	1.72	5.99
Non-gender	[4.01, 5.81]	[6.07, 12.66]	1.80	6.59
	CI MAE 1 st Session	CI MAE 2 nd Session	v_{1st}	v_{2nd}
Male	[1.34, 1.82]	[1.12, 2.35]	0.50	1.23
Female	[1.76, 2.30]	[2.27, 3.76]	0.54	1.49
Non-gender	[1.18, 1.66]	[1.47, 2.81]	0.48	1.34
	CI MSD 1 st Session	CI MSD 2 nd Session	v_{1st}	v_{2nd}
Male	[0.01, 0.75]	[-1.92, 0.01]	0.74	1.93
Female	[-0.65, 0.22]	[-2.54, 0.22]	0.87	2.76
Non-gender	[-0.01, 0.59]	[-1.96, 0.02]	0.60	1.98

Source: Elaborated by the author.

It is important to investigate whether there is a significant difference between ShapedNet and DXA. The third research question risen in subsection 3.3.2 was: *Is ShapedNet really different from a DXA exam?* We formulated the following hypotheses to address this question::

H₀: There are no differences between ShapedNet and DXA. So they can be used interchangeably.

H₁: ShapedNet is different of DXA. So, they can be used respecting their constrains.

We conducted statistical tests to address the research question regarding the comparability of ShapedNet and DXA in estimating body fat percentage. The results of these tests are presented in Table 11. The table includes two tests, one for the T_1 experiment and another for T_2 , comparing the BFP estimates obtained from DXA and ShapedNet. The results support accepting the null hypothesis **H₀**, indicating that ShapedNet and DXA are statistically equivalent in estimating body fat percentage. However, it is important to exercise caution in extrapolating these findings, as further analyses in larger populations are necessary to strengthen the conclusions. Additionally, it should be noted that the hypothesis test conducted here pertains to differences within a specific population, aiming to determine whether the methods originate from the same distribution, which is the distribution of BFP in a given population. Furthermore, ShapedNet is based on a fundamentally different mathematical approach compared to DXA. Although DXA

values were used to train the model and served as a reference for calculating training errors and adjusting model parameters.

We examined the correlation between ShapedNet and DXA in multiple experiments, including the T_1 and T_2 sections of photos. The correlation coefficients, $\rho_{x,y}$, and determination coefficients, R^2 , are summarized in Table 8. While the models are statistically equivalent, they also demonstrate a strong correlation, with coefficients exceeding 0.9 in all cases. This suggests that ShapedNet provides estimates of body fat percentage that are highly consistent with those obtained from DXA. Notably, the non-gender model exhibited the highest correlation at 0.982, which aligns with its lowest MAPE error. Conversely, the female model displayed the lowest correlation at 0.915, corresponding to its highest MAPE error, as indicated in Table 8.

The scatter diagrams visualizing the correlations are depicted in Figure 36. Each diagram represents a pair-plot between DXA and ShapedNet, accompanied by a regression line indicating the overall trend. The diagrams are grouped according to the experiments (T_1 and T_2), sections of photos, gender-based models, and the non-gender-based model, resulting in a total of twelve graphs. It is important to note that the T_2 experiment has a smaller sample size compared to T_1 , which explains the higher determination coefficient, R^2 , observed in the T_2 graphs compared to the T_1 graphs.

Table 11 – Hypothesis test with The Follow-up Assessment. ShapedNet vs DXA, in both photo sections. p highlighted in bold represent the groups that are significantly different.

Approach	Groups	MD	p		Groups	MD	p
Male	T_1 DXA vs ShapedNet	-0.324	0.781		T_2 DXA vs ShapedNet	0.548	0.620
Female		0.560	0.587			0.866	0.370
Non-gender		-0.213	0.853			0.668	0.542

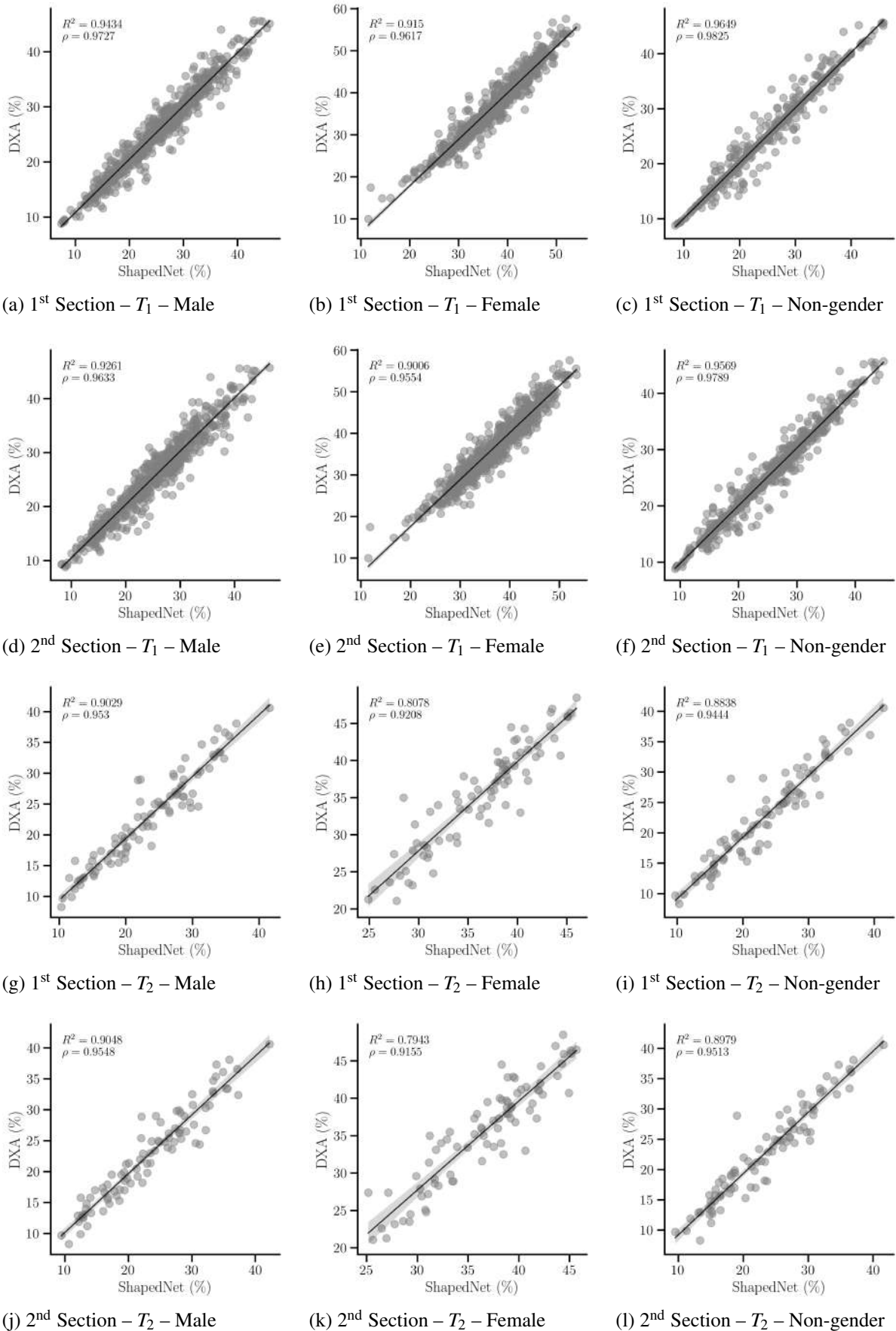
Source: Elaborated by the author.

Table 12 – Comparison of $\rho_{x,y}$ and R^2 for the first and second sections with follow-up assesment.

Approach	R^2	$\rho_{x,y}$	R^2	$\rho_{x,y}$
$T_1 - 1^{st}$ Session			$T_2 - 1^{st}$ Session	
Male	0.943	0.972	0.902	0.953
Female	0.915	0.961	0.807	0.920
Non-gender	0.964	0.982	0.897	0.951
$T_1 - 2^{nd}$ Session			$T_2 - 2^{nd}$ Session	
Male	0.926	0.963	0.904	0.954
Female	0.901	0.955	0.794	0.915
Non-gender	0.978	0.956	0.846	0.922

Source: Elaborated by the author.

Figure 36 – Correlation Analysis between DXA and ShapedNet in Experiments T_1 and T_2 in Two Exam Sections for Gender-based and Non-gender-based Approaches.



Source: Elaborated by the author.

To understand the temporal dynamics of DXA and ShapedNet outcomes, we provide a summary in Table 13. The table displays the average and standard deviation of body fat percentages for the population. In the T_1 experiment, the average BFP for males was 23.26. After three months, in the T_2 experiment, it decreased to 22.81, resulting in an average difference of Δ_{DXA} of -0.44 between T_1 and T_2 . This suggests that, on average, males experienced a reduction in body fat. Similarly, females (-0.45) and the non-gender population (-0.45) also exhibited a decrease in body fat over the three-month period. However, the pattern is different for ShapedNet. Only females showed a decrease in body fat over time (-0.15). Interestingly, for males and the non-gender population, the body fat predictions went in the opposite direction during this period.

The δ values presented in Table 13 provide an indication of the overall error direction. These values are calculated as $\Delta_{DXA} - \Delta_{\text{ShapedNet}}$, representing the difference between DXA and ShapedNet in predicting changes in body fat percentage over time for each subject. Ideally, δ would be close to 0, indicating that ShapedNet predicts body fat percentage changes in a similar manner to DXA. However, the δ values for females are the closest to 0, but with a negative sign, indicating that DXA estimates are higher than ShapedNet predictions. In all groups, $\delta \leq 0$, suggesting that ShapedNet did not capture the variability in body fatness as effectively as DXA. These findings are consistent with the previously highlighted observations, indicating that, in general, ShapedNet tends to underestimate body fat percentage compared to DXA.

Understanding the temporal dynamics of DXA and ShapedNet outcomes is essential for comprehending the accuracy of ShapedNet in capturing changes in body fat over time. We conducted an analysis to explore the relationship between DXA and ShapedNet in various experiments, including the T_1 and T_2 sections of photos. The correlation coefficients ($\rho_{x,y}$) and determination coefficients (R^2) were calculated. We emphasize again that the models are statistically equivalent and they also exhibit a strong correlation with correlation coefficients exceeding 0.9 in all cases. These findings highlight the robustness of the correlation between ShapedNet and DXA in estimating body fat percentage.

Table 13 – Comparison of ShapedNet and DXA regarding gender-based and non-gender-base approaches. Reported values are mean and standard deviation of BFP estimates.

BFP method/Gender	Male	Female	Non-gender
DXA (%) T_1	23.26 ± 7.71	35.95 ± 6.81	29.27 ± 9.66
DXA (%) T_2	22.81 ± 7.55	35.50 ± 6.84	28.82 ± 9.60
Δ_{DXA}	-0.44 ± 1.97	-0.45 ± 1.73	-0.45 ± 1.86
ShapedNet (%) T_1	22.93 ± 7.82	36.51 ± 6.21	29.36 ± 9.82
ShapedNet (%) T_2	23.36 ± 7.18	36.36 ± 5.24	29.52 ± 9.07
$\Delta_{ShapedNet}$	0.42 ± 2.83	-0.15 ± 2.56	0.15 ± 2.71
$\delta = \Delta_{DXA} - \Delta_{ShapedNet}$	-0.87 ± 2.23	-0.30 ± 2.62	-0.60 ± 2.43

Source: Elaborated by the author.

4.1.4 Limitations of ShapedNet

One limitation of ShapedNet is related to the input photos used during data acquisition. The photos were taken in a controlled environment with a white background, and the network was trained using these specific background conditions. However, in clinical practice, photos can be captured in various backgrounds and scenes, which may negatively impact the performance of the method. Although some preliminary tests were conducted using individuals outside of the original study in different scenarios, the results were inconclusive. While the estimation of body fat percentage did not seem to be significantly affected, there were challenges associated with person identification and localization. Therefore, further investigations are required to gain a better understanding of this issue and develop strategies to address it effectively.

We did not perform a dedicated investigation or quantification of how camera angle and scene illumination specifically impact the estimation of body fat percentage. However, during the data collection process, we made slight adjustments to the camera position by varying the angles from up and down. This was done in order to capture potential variations caused by the camera's angle and incorporate them into the dataset. While this approach aimed to account for some variations, further studies focusing specifically on the impact of camera angle and scene illumination on BFP estimation would provide valuable insights and help enhance the robustness of the method.

4.2 A qualitative analyses of DBF-SCAN

In this section, we explore the quantitative analysis of the visual representations generated by the DBF-SCAN tool. By capturing gradients from the last convolutional layer, the one closest to the body fat estimation output, we create visual representations to provide valuable insights into the inner calculations and feature extraction performed by ShapedNet. Through this analysis, we gain a deeper understanding of how ShapedNet processes the input data and produces accurate body fat estimations.

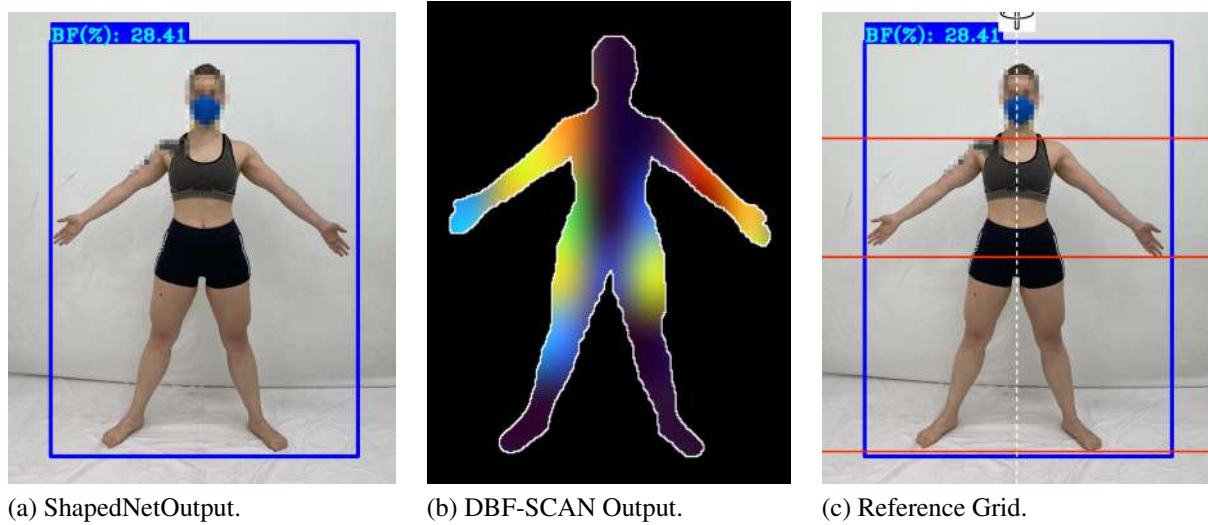
The results of ShapedNet and DBF-SCAN are presented in Table Subfigure 37. The predictions were generated using the non-gender-based model, which demonstrated the lowest error rate among the models. The output of ShapedNet is shown in Figure 37a. The person is detected, and their coordinates are plotted within the figure. The estimated body fat percentage is displayed in the upper left corner of the bounding box. Figure 37b illustrates the output of DBF-SCAN. The red and yellow regions represent pixels that are highly associated with the body fat percentage regression output, while the green regions indicate a moderate level of association. Conversely, the darker and blue regions indicate less relevance to the distribution of body fatness.

The estimate BFP using DXA in subject E04 is 28.5% and ShapedNet estimated 28.41%, an error of 0.11 points. It is noteworthy that the primary objectives of DBF-SCAN are to offer a visual representation that demonstrates the regions that are most likely to impact body fatness estimation and to assist in the development of ShapedNet for a better understanding of the underlying processes. However, the findings raise questions such as: why do the right and left sides of the body have different influences on BFP? Why do certain regions highlight more than others?

Let us examine the visualization presented in Figure Subfigure 39 in detail. Our focus is to understand the areas highlighted by DBF-SCAN and assess if the output aligns with our domain knowledge. In Figure 37c, we can observe a slight clockwise rotation of the subject. The red lines act as reference points to distinguish the right side of the body from the left side. The left foot and left hand are slightly shifted forward, resulting in the left shoulder appearing lower than the right shoulder. Moving on to Figure 39a, we can identify the main regions highlighted by our method, which include: (1) the area of the right chest and right arm, (2) the waist and thigh regions, and (3) the area of the left arm.

In Figure Subfigure 39b, a side-by-side visualization allows us to observe the subject's physique and the absence of body regions with higher fat density. The arms and chest

Figure 37 – Results of DBF-SCAN in a female subject (E04). In (a) it the output of ShapedNet, (b) the body fatness likelihood and (c) a reference grid for analyses.



Source: Elaborated by the author.

region are prominently highlighted as highly relevant. It is crucial to understand that DBF-SCAN is subject-specific, and the red regions indicate the areas that are considered more relevant to BFP for a specific subject. Therefore, it is not appropriate to assume that a subject has more fat in a certain area solely based on the color difference in another subject's visualization. In this case, we can observe that the inner parts of the subject's arms are larger, resulting in strong highlighting in the DBF-SCAN visualization. Both thighs are also highlighted as relevant regions, which aligns with the subject's hypertrophied legs. Additionally, the abdominal region, particularly the suprailiac, is shown to have medium relevance to BFP. On the other hand, the lower part of the body, such as the calves, is not deemed relevant to body fatness estimation, as evident from their absence of highlighting in the visualization.

It is known that abdominal regions tend to accumulate more fat than other body regions. This behavior is also captured by DBF-SCAN, as seen in the results for the male subject in Figure 38. The DXA exam showed a body fat percentage of 45.7%, and ShapedNet estimated 45.62%, with an error of 0.26%. This subject is considered obese, so it is expected to have a higher fat density in the abdominal region. The method clearly shows relevant regions that are consistent with the subject's photo. For example, in Figure 40a, the most relevant regions are shown, including the upper chest area that captures both the right and left arms, and the abdominal and waist areas.

The comparison presented in Figure Subfigure 40b highlights the abdominal and waist areas as the most significant contributors to BFP estimation, which is visually confirmed by

the subject's photo. Additionally, the lower parts of both arms exhibit higher relevance to body fatness. By considering the umbilical point as the center of density determined by DBF-SCAN, a closer examination reveals that the neighboring regions also play a role in body fat estimation. It is noteworthy that the left side of the body, particularly a region near the suprailiac skinfold, demonstrates a much higher relevance in terms of BFP estimation, as indicated by the red colors. However, it is important to note that the head, hands, and feet do not significantly contribute to BFP estimation.

Figure 38 – Results of DBF-SCAN in a male subject (E27). In (a) it the ouput of ShapedNet and (b) the body fatness likelihood



(a) ShapedNet Output.

(b) DBF-SCAN Output.

Source: Elaborated by the author.

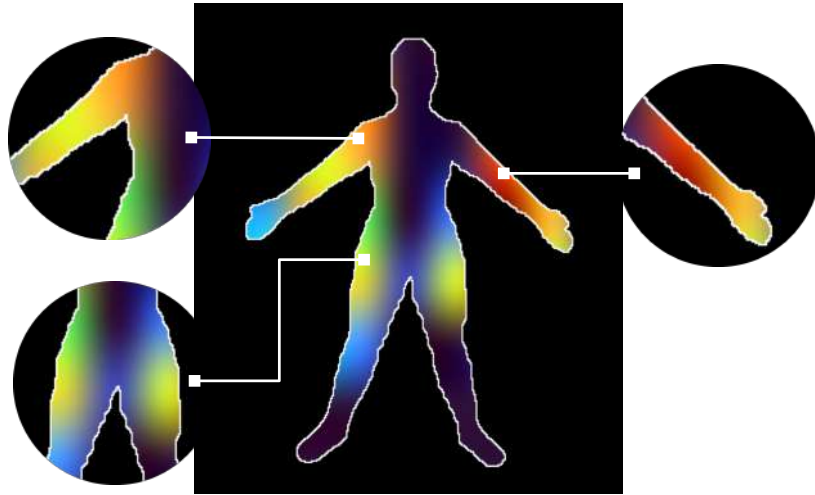
DBF-SCAN focuses on identifying and highlighting body regions that have a higher importance for the individual's BFP, rather than providing a general representation that applies to a wide range of individuals. This personalized approach allows for a more accurate estimation of body fat percentage by considering the unique characteristics and distribution of body fat in each individual.

In other words, DBF-SCAN identifies and visualizes the regions of the body that are most informative for estimating body fat percentage on an individual basis. These regions may vary from person to person, depending on their unique body composition and fat distribution patterns. Therefore, the visualization results obtained through DBF-SCAN are more tailored to the specific characteristics of each individual, making it a subject-oriented approach.

This distinction is important because it means that the visualization provided by

DBF-SCAN may not necessarily represent a universally applicable pattern or general trend of body fat distribution across a diverse population. Instead, it aims to provide insights into the specific regions that are particularly influential in determining body fat percentage for each individual. Healthcare professionals and researchers should consider this subject-oriented nature of DBF-SCAN when interpreting and utilizing its results in clinical or research settings.

Figure 39 – In-depth Analysis of DBF-SCAN Results for Subject E04.



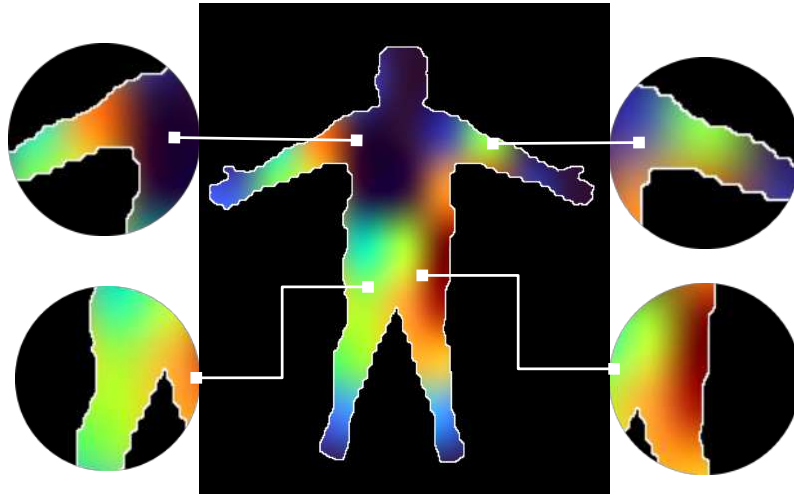
(a) Relevant Regions for BFP estimate.



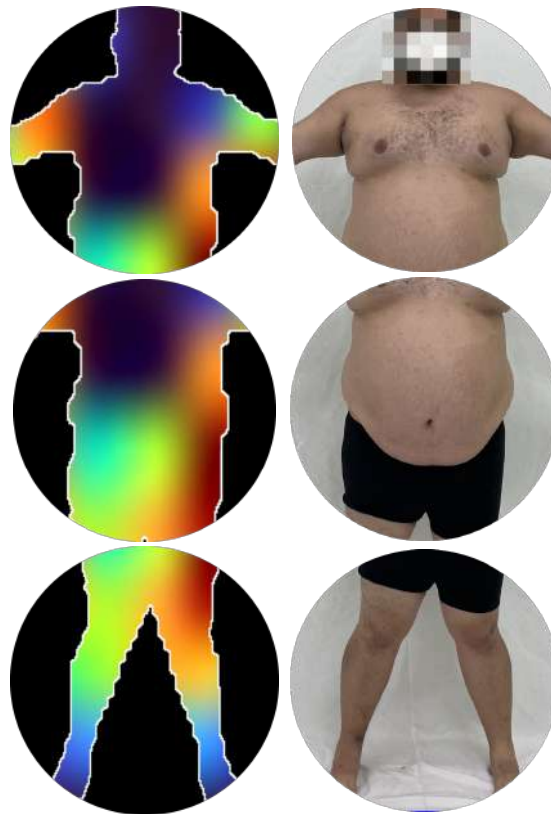
(b) Side by Side of DBF-SCAN and Body regions.

Source: Elaborated by the author.

Figure 40 – In-depth Analysis of DBF-SCAN Results for Subject E27.



(a) Relevant Regions for BFP estimate.



(b) Side by Side of DBF-SCAN and Body regions.

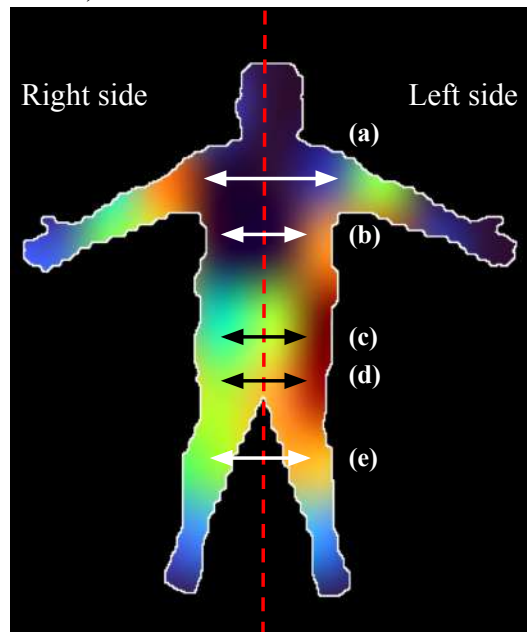
Source: Elaborated by the author.

4.2.1 Limitations of DBF-SCAN

One current limitation of DBF-SCAN is its potential to assign different levels of relevance to anatomically equivalent regions on opposite sides of the body. This can be observed in Figure 41, where five regions are labeled: (a) the upper chest area, influenced by the chest skinfold, biceps, and triceps; (b) the lower chest region, influenced by the front midaxillary skinfold and back subscapular skinfold; (c) the umbilical region, which includes both abdominal and suprailiac skinfolds; (d) the lower hip region, which includes the anterior thigh and gluteal muscles; and (e) the upper calf region.

A more detailed analysis of Figure 41 reveals several differences in the relevance assigned to equivalent regions on the right and left sides of the body. Specifically, the right side upper chest, biceps, and forearm were given greater relevance compared to the left side, as indicated by the label (a) in the figure. Conversely, the midaxillary region on the left side was more relevant than the right side, as shown by the label (b). Labels (c) and (d) demonstrate that both the abdominal and hip regions on the left side have a stronger influence on BFP compared to the right side, as evidenced by the contrasting shades of green and red. Finally, the left upper calf was found to be more relevant than the right one, as highlighted by label (e).

Figure 41 – Analyses of the right and left side of the body in DBF-SCAN (Subject E27).

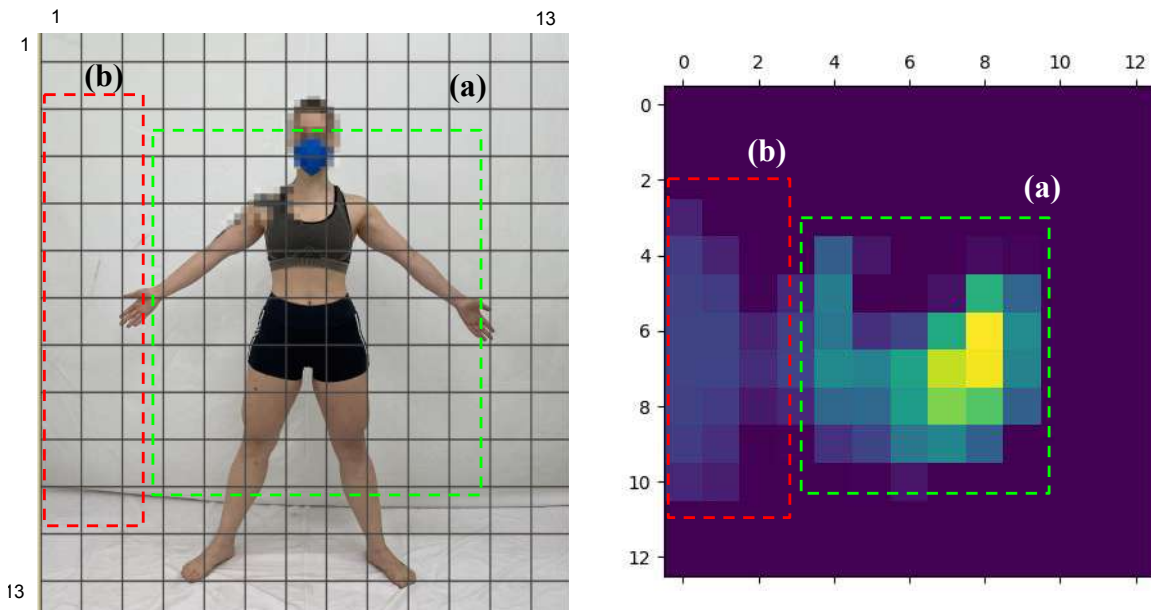


Source: Elaborated by the author.

Another limitation of DBF-SCAN is that it relies on ShapedNet which does not

segment the person in the photo, it rather estimate its localization. It is important to remember that DBF-SCAN uses ShapedNet as the basis for its calculations. Specifically, the last convolutional layer before the output to estimate body fat percentage is used to compute the matrices $\mathbf{H}_r^{(58)}$ and $L_{\text{BF-SCAN}}^r$. The tensors are then reduced to a 13×13 matrix. To illustrate this point, Figure 42 shows a side-by-side comparison of the subject's input photo and $L_{\text{BF-SCAN}}^r$. In the $L_{\text{BF-SCAN}}^r$ matrix shown in Figure 42, the region marked as (a) is the correct region relative to the person in the input photo. This region in Figure 42 pertains to the portion of the image that includes the subject's body and is marked as (a). Conversely, the region labeled as (b) does not contain any part of the subject's body and is therefore inconsequential for estimating body fat percentage. However, the area has regions that contain relevance for BFP estimation, which should not exist since the person is outside the bounds of region (b). This discrepancy is a clear error that affects both the accuracy of body fat percentage calculation and the visualization of DBF-SCAN.

Figure 42 – Side by Side Comparison of Subject E04 and Her Heatmap Matrix $L_{\text{BF-SCAN}}^r$.



Source: Elaborated by the author.

4.3 Using Transfer Learning to estimate Body Fat Percentage

Using transfer learning to estimate body fat percentage on photos involves leveraging pre-trained deep learning models trained on large-scale image datasets and adapting them to predict body fat percentage from input photos. Transfer learning is a technique that allows us to utilize knowledge learned from one task (source task) and apply it to a different but related task

(target task). In this case, we can take advantage of pre-trained models that have learned rich representations from general image data, such as ImageNet.

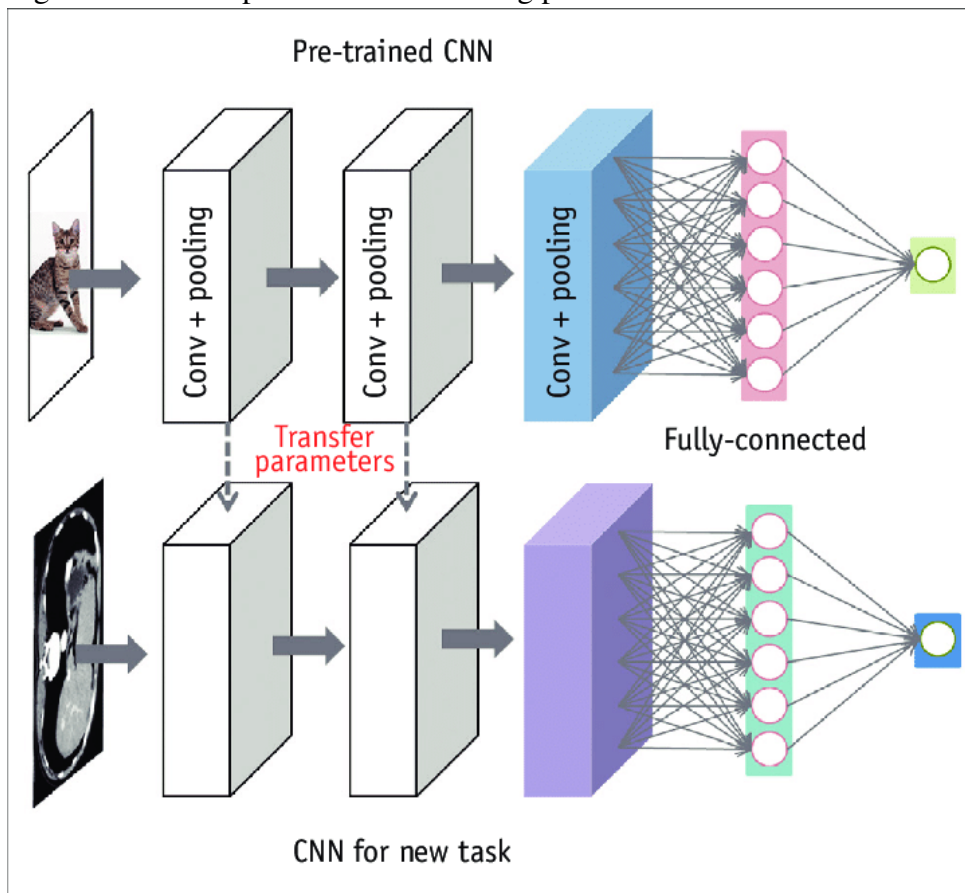
To estimate body fat percentage, we fine-tune the pre-trained model by retraining the last layer specifically designed for our target task. An example for the transfer-learning procedure is exhibited in Figure 43. In our case, we used 10 widely known architectures to train a fully-connected layer to estimate body fat percentage using a single photo. The models are pre-loaded and its weights are kept frozen until the last layer. We added a Flatten layer, for vectorization, then a Dense layer with one neuron with a linear activation function. Moreover, as loss function we used the Mean Absolute Error.

The dataset utilized in this study was the BCDB23, which is described in detail in Section 3.3.1. The experimental settings employed for training the deep neural networks were consistent with those outlined in Section 3.3.2. Specifically, each deep neural network model was trained for 1000 epochs using the ADAM optimizer (KINGMA; BA, 2014), with early stopping implemented. To facilitate efficient training on a large scale, Tensorflow 2.4 (ABADI *et al.*, 2015) was utilized, taking advantage of its implementation details and the utilization of Graphical Processing Units (GPUs).

A general overview of the proposed methodology is depicted in Figure 44, which also highlights the pre-trained models utilized in the study. The results reported in this study adhere to the metrics described in Section 3.3.2. Tables 14, 15, and 16 present the results for the non-gender, male-gender, and female-gender-based models, respectively. The results presented in the table concern the test data. The best-performing results are highlighted in bold.

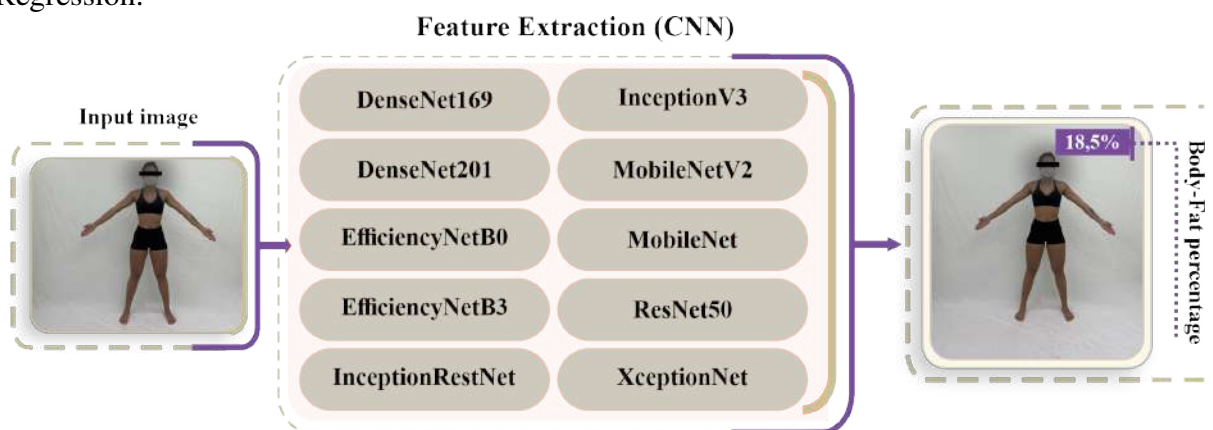
Among the non-gender-based models the network that used MobileNet achieved 16/53% of MAPE, performing better than other architectures. The male-based network reached its lowest error in DenseNet169 with 13.33% of error and the female-based model 14.35% of error. It is important to note that the output of these models differs significantly from that of ShapedNet. Unlike ShapedNet, the models discussed in this study do not operate within a multi-task learning framework. Instead, their primary objective is to estimate body fat percentage (BFP) and they do not provide information regarding localizations or identifications.

Figure 43 – Example of transfer learning procedure.



Source: Adapted from Do *et al.* (2020).

Figure 44 – Proposed methodology to incorporate Transfer Learning for Deep Learning Regression.



Source: Elaborated by the author.

Table 14 – Reported metrics MAE and MAPE for the non-gender based model. The 95% confidence intervals are reported as well.

CNN	MAE	MAE - CI	MAPE (%)	MAPE (%) - CI
DenseNet169	4.34	[3.77, 4.90]	16.56	[13.16, 19.96]
DenseNet201	4.48	[3.88, 5.07]	17.69	[13.87, 21.51]
EfficiencyNetB0	4.70	[4.03, 5.37]	18.57	[14.47, 22.68]
EfficiencyNetB3	4.92	[4.25, 5.59]	19.80	[15.12, 24.47]
InceptionResNet	6.63	[1.72, 11.54]	29.42	[2.74, 56.11]
InceptionV3	6.22	[2.52, 9.92]	21.47	[13.78, 29.17]
MobileNetV2	5.14	[4.45, 5.83]	19.78	[15.61, 23.95]
MobileNet	4.21	[3.62, 4.79]	16.53	[12.83, 20.23]
ResNet50	5.72	[4.84, 6.61]	20.22	[16.29, 24.15]
XceptionNet	5.09	[4.35, 5.83]	19.37	[15.34, 23.40]

Source: Elaborated by the author.

Table 15 – Reported metrics MAE and MAPE for the male-gender based model. The 95% confidence intervals are reported as well.

CNN	MAE	MAE - CI	MAPE (%)	MAPE (%) - CI
DenseNet169	2.77	[2.08, 3.47]	13.33	[9.00, 17.57]
DenseNet201	2.93	[2.27, 3.58]	14.64	[9.81, 19.46]
EfficiencyNetB0	4.03	[3.19, 4.87]	19.19	[13.44, 24.95]
EfficiencyNetB3	3.87	[2.97, 4.77]	18.45	[12.92, 23.98]
InceptionResNet	3.11	[2.42, 3.80]	15.35	[10.19, 20.57]
InceptionV3	3.23	[2.61, 3.85]	15.10	[10.61, 19.59]
MobileNetV2	3.95	[3.25, 4.64]	17.18	[13.06, 21.30]
MobileNet	2.98	[2.38, 3.58]	14.83	[10.38, 19.27]
ResNet50	3.31	[2.53, 4.10]	16.83	[11.13, 22.03]
XceptionNet	3.72	[2.91, 4.53]	16.91	[12.57, 21.26]

Source: Elaborated by the author.

Table 16 – Reported metrics MAE and MAPE for the female-gender based model. The 95% confidence intervals are reported as well.

CNN	MAE	MAE - CI	MAPE (%)	MAPE (%) - CI
DenseNet169	3.95	[3.95, 0.14]	14.35	[9.23, 19.47]
DenseNet201	4.15	[3.31, 4.98]	14.03	[9.56, 18.49]
EfficiencyNetB0	4.54	[3.55, 5.53]	16.34	[10.58, 22.10]
EfficiencyNetB3	4.36	[3.50, 5.22]	14.02	[10.18, 17.86]
InceptionResNet	6.48	[1.80, 11.15]	18.76	[9.00, 28.52]
InceptionV3	6.59	[1.66, 11.51]	20.96	[7.72, 34.19]
MobileNetV2	4.70	[3.78, 5.62]	15.52	[10.89, 20.14]
MobileNet	4.29	[3.38, 5.19]	14.70	[9.91, 19.48]
ResNet50	6.38	[5.34, 7.42]	19.15	[15.16, 23.15]
XceptionNet	4.37	[3.36, 5.38]	15.07	[9.76, 20.39]

Source: Elaborated by the author.

4.4 Comparing with State-of-the-Art (SOTA) Methods

A comparison with state-of-the-art (SOTA) approaches for estimating body fat percentage is shown in Table 17. The table is presented in descending order, with the most recent approach from the literature at the top. "N" represents the number of subjects used in the study, while "RM" refers to the reference method used for body fat estimation. The values of MAE and MAPE for males, females, and individuals of non-gender are reported for comparison. TN stands for Transfer Learning.

Table 17 – Comparative Analysis: ShapedNet vs. SOTA Approaches for BFP estimate. Metrics: MAE, MAPE for Male, Female, Non-Gender. N: Number of Subjects. RM: Reference Method. Dash indicates missing information.

Approach	Author	Input	N	RM	MAE _m / MAPE _m	MAE _f / MAPE _f	MAE _{ng} / MAPE _{ng}
ShapedNet	Our approach	Front photo	1273	DXA	1.59 / 7.04%	2.03 / 5.98%	1.42 / 4.91%
TL	Our approach	Front photo	1273	DXA	2.77 / 13.33%	3.95 / 14.35%	4.21 / 16.53%
VBC	Majmudar <i>et al.</i> (2022)	Front photo + Back photo	134	DXA	1.88 / 6.8%	2.43 / 6.1%	—
L-SVM	Alves (2022)	6 CV-AM + IV	912	DXA	3.24 / 15%	3.44 / 11%	3.47 / 13%
LSSVR	Alves <i>et al.</i> ()	10 AM + IV	163	DXA	2.75 / —	4.01 / —	3.56 / —
DT-SVM	Uçar <i>et al.</i> (2021)	13 AM	252	Siri-BD	3.53 / 30.06%	—	—
IRE-SVM	Chiong <i>et al.</i> (2021)	13 AM	252	Siri-BD	3.62 / —	—	—
MR-MARS	Shao (2014)	13 AM	252	Siri-BD	3.69 / 24.28%	—	—

Source: Elaborated by the author.

Regarding the results of a TL-based approach, it achieved its lowest error for males, with 13.33%. While using Transfer Learning outperform Shao (2014), Chiong *et al.* (2021), Uçar *et al.* (2021) and Alves (2022). of the state-of-the-art methods, none of them achieved the desired criteria of $\leq 10\%$ error or was close to reach results of ShapedNet. Nevertheless, this demonstrates the potential of transfer learning modeling in this domain. It is worth noting that there is still room for improvement in this approach, and with further effort, it is possible to attain the desired criteria.

Our approach, ShapedNet, utilizes front photos as input and uses DXA as the reference method, similar to the work of Majmudar *et al.* (2022). The method developed by these researchers, called VBC, is the most similar to ours as it also employs CNN-based models. The results achieved by Majmudar *et al.* (2022) are considered relevant, as they achieved a MAPE of less than or equal to 10%. However, the number of subjects in their study represents only 10.5%

of our dataset. Our study is more robust in terms of generalization, compared to Majmudar *et al.* (2022), due to the inclusion of a significantly larger sample size, with 9.5 times more human subjects. Additionally, a key advantage of ShapedNet over VBC is that our approach only requires one photo, whereas VBC requires two photos (front and back views). At last but not least, Majmudar *et al.* (2022) requires preprocessing in the input images before feeding VBC, such as background removal using another CNN model, while ShapedNet does not require preprocessing.

We achieved an MAE of 1.59 and MAPE of 7.04% for male subjects, and an MAE of 2.03 and MAPE of 5.98% for female subjects. In comparison, Majmudar *et al.* (2022) achieved similar results, but in a much smaller sample size. The notable difference lies in the non-gender based model, where we achieved an MAE of 1.42 and MAPE of 4.91%. This represents an improvement of 1.19 to 1.89 points in MAPE for male and female subjects, respectively, when compared to Majmudar *et al.* (2022). Moreover, the results reported by the authors are not clear if they are in the test dataset or in the all subjects.

The proposed method demonstrates a significantly lower error of 1.65 compared to Alves (2022) for male subjects. Furthermore, the error is 1.16, 1.94, 2.03, and 2.10 lower than the methods proposed by Alves *et al.* (), Uçar *et al.* (2021), Chiong *et al.* (2021), and Shao (2014), respectively. It is important to note that these methods are restricted to the male gender. In contrast, our study encompasses a broader analysis, with a dataset of 1273 individuals, including 578 men and 695 women. This significantly larger study size allows for more robust conclusions, particularly for the male population, which is approximately twice the size of the populations in the previously mentioned studies.

The only authors that studied female subjects were Majmudar *et al.* (2022), Alves (2022) and Alves *et al.* (). Regarding their results, the error in our method is 0.43 lower than Majmudar *et al.* (2022), and 1.41 and 1.98 for Alves (2022) and Alves *et al.* () respectively. Among the authors presented in Table 17 only Majmudar *et al.* (2022), Alves (2022), and Alves *et al.* () studied female subjects. In terms of results, our method achieves an error in BFP that is 0.43 lower than Majmudar *et al.* (2022), and 1.41 and 1.98 lower than Alves (2022) and Alves *et al.* (), respectively.

According to the baseline metrics mentioned in Section 3.3.2, models that have MAPE errors above 10% are considered inappropriate for use. Therefore, the errors reported by (SHAO, 2014) and (CHIONG *et al.*, 2021), which are 30.06% and 25.28%, respectively, are

significantly higher compared to our method, which achieves an error of 4.91%. In these studies, the errors are 2-3 times larger than our method.

In terms of simplicity in clinical usage, ShapedNet is the most advantageous among these approaches. Our method only requires a single photo without any additional user input to estimate the body fat percentage (BFP). In contrast, Majmudar *et al.* (2022) requires two photos and user input for gender. Additionally, Alves (2022) utilizes a computer vision model to estimate anthropometric measurements and relies on user input for age, height, weight, and gender. The method proposed by Alves *et al.* () involves the measurement of 10 anthropometric circumferences by a trained professional, along with user input for age, height, weight, and gender. Uçar *et al.* (2021), Chiong *et al.* (2021), Shao (2014) also involve the measurement of 13 anthropometric circumferences by a professional, along with user input for weight and gender. Moreover, the last two studies rely on Siri's equations (SIRI,) as the reference method for training and evaluation, that despite being vastly used in clinical routines it is considered limited in current clinical practices and research.

5 CONCLUSIONS

Our research has demonstrated that the design of multi-task learning, as applied in our study, can have significant contributions not only to body composition estimation but also to other domains that require models capable of performing identification, localization, and regression tasks simultaneously. We have developed a novel method, ShapedNet, based on CNNs that simultaneously estimates body fat percentage, identifies and localize individuals.

We outperformed previous state-of-the-art machine learning methods in terms of generalization and accuracy for body fat estimation. By utilizing a substantially larger sample size of 1273 individuals, which is 9.5 times larger than previous studies, our method provides more robust results. Notably, our non-gender-based model outperformed the best reported results for computer vision-based body fat estimation using photos by a margin of 1.19 to 1.89 points.

The accuracy of ShapedNet was validated against the gold standard reference method, Dual-Energy X-ray Absorptiometry (DXA). ShapedNet achieved the lowest error, with a mean absolute percentage error (MAPE) of 4.91% and mean absolute error (MAE) of 1.42. Hypothesis testing demonstrated that the non-gender-based model was 1.07 points more accurate than the female-based model and 2.12 points more accurate than the male-based model.

With 95% confidence, ShapedNet estimates body fat percentage within an error range of 4.01% to 5.81% compared to DXA, or a difference of 1.18 to 1.66 in DXA estimates. The female-based model slightly underestimated body fat percentage by -0.21 points, while the male-based model and non-gender-based model overestimated by 0.38 and 0.28 points, respectively.

The precision analysis showed that both the gender-based and non-gender-based models were considered acceptable, as the MAE, MAPE, and MSD results in the first and second sessions of photos were statistically equivalent. ShapedNet exhibited a high correlation with DXA across various variants, with correlation coefficients greater than 0.9 in 100% of experiments, and determination coefficients R^2 greater than 0.9 in 75% of experiments.

In terms of temporal concordance assessment, ShapedNet predicted body fat percentage within a small margin of deviation. On average, it estimated a 0.30% increase in females, a -0.86% decrease in males, and a -0.6% decrease using the non-gender-based model. However, further investigations are required to explore the differences between DXA and ShapedNet when individuals gain or lose fat.

The visualization method, DBF-SCAN, demonstrated correlation with relevant areas

that accumulate fat. However, it indicates regions more relevant to body fat percentage within individuals, making it subject-oriented rather than general-oriented.

Overall, our research advances the field of body composition estimation by providing a robust and accurate method, ShapedNet, supported by rigorous validation, correlation analysis, and visualization techniques. Regarding body composition assessment theory, the ShapedNet method is an innovative indirect approach classified as a two-compartment method. These findings have the potential to improve personalized health assessments, contribute to the development of explainable AI models, and broaden the applications of multi-task learning in various domains. By developing a CNN-based algorithm that incorporates multi-regression, classification, and localization tasks, we have opened up possibilities for various applications beyond body composition estimation.

5.1 Regarding the goals

Based on the specific goals outlined in Section 1.3.1, the following conclusions have been drawn:

- Firstly, we have successfully designed a CNN-based algorithm, named **ShapedNet**, to tackle multi-task learning by simultaneously addressing multi-regression, classification, and localization tasks. By harnessing the power of deep learning, this algorithm effectively extracts relevant features from input photos and accurately estimates body fat percentage. The integration of Convolutional Neural Networks within the architecture enables precise estimation of human body fat percentage based on a single input photo, while also identifying the person's coordinates.
- Furthermore, a visualization technique has been developed to highlight and quantify the relative importance of different body regions in determining body fatness, called **DBF-SCAN**. This technique provides valuable visual insights into the distribution and composition of body fat, contributing to a better understanding of the factors influencing an individual's body fat percentage. The integration of this visualization technique enhances the interpretability and usability of the proposed method.
- The proposed method has been thoroughly validated using a comprehensive and heterogeneous dataset consisting of 1273 subjects. This dataset encompasses various variants, including two photo sections and samples collected three months apart. The validation process involved comparing the estimated body fat percentage with the gold standard

reference method, Dual-Energy X-ray Absorptiometry, and assessing the method's performance in terms of accuracy, precision, and reliability. The results demonstrate the effectiveness and accuracy of the developed method, further validating its potential for practical application in reliable and precise body fat estimation.

- Additionally, to foster inclusivity and cater to diverse individuals, the developed method has undergone validation with a particular focus on gender diversity. This validation approach incorporates both gender-based and non-gender-based considerations, ensuring that the method remains applicable to a wide range of individuals and avoids biases associated with conventional body fat estimation techniques. By placing emphasis on inclusivity, we enhance the applicability and fairness of the proposed method.

In conclusion, the designed CNN-based algorithm, supported by the integration of Convolutional Neural Networks, offers a robust solution for simultaneous multi-regression, classification, and localization tasks in the estimation of body fat percentage. The visualization technique enhances the understanding of body fat distribution, and the thorough validation process ensures accuracy and reliability. By considering gender diversity, the method promotes inclusivity and avoids biases. Overall, these achievements contribute to the advancement of body fat estimation techniques, with potential applications in various fields related to health, fitness, and well-being.

5.2 Future works

This research serves as a stepping stone for future studies, and we acknowledge that there are numerous topics that extend beyond the scope of this thesis. We propose the following areas for further investigation:

- Overcome the limitations of ShapedNet by introducing strategies to mitigate impact of different backgrounds, scenes and illumination during modeling;
- Investigate a multi-input Deep Neural Network that uses two or more photos as inputs;
- Investigate the multi-task learning regression in a semantic segmentation model.
- Investigate DBF-SCAN deeply to extrapolate findings for practical application of this method;
- Incorporate the DBF-SCAN during training as a recurrent input, creating a new architecture that feeds itself using XAI to guide learning.

REFERENCES

- ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; GOODFELLOW, I.; HARP, A.; IRVING, G.; ISARD, M.; JIA, Y.; JOZEFOWICZ, R.; KAISER, L.; KUDLUR, M.; LEVENBERG, J.; MANÉ, D.; MONGA, R.; MOORE, S.; MURRAY, D.; OLAH, C.; SCHUSTER, M.; SHLENS, J.; STEINER, B.; SUTSKEVER, I.; TALWAR, K.; TUCKER, P.; VANHOUCKE, V.; VASUDEVAN, V.; VIÉGAS, F.; VINYALS, O.; WARDEN, P.; WATTENBERG, M.; WICKE, M.; YU, Y.; ZHENG, X. **TensorFlow**: Large-scale machine learning on heterogeneous systems. 2015. Software available from tensorflow.org. Disponível em: <https://www.tensorflow.org/>. Acesso em: 10 de mar. de 2023.
- ALVES, S. S.; OHATA, E. F.; NASCIMENTO, N. M.; SOUZA, J. W. D.; HOLANDA, G. B.; LOUREIRO, L. L.; FILHO, P. P. R. Gender-based approach to estimate the human body fat percentage using machine learning. *In*: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN), 2021, 'on-line'. Anais eletrônicos [...]. [S. l.]: IEEE Computational Intelligence Society, 2021. p. 1–8.
- ALVES, S. S. A. **Sex-based approach to estimate human body fat percentage from 2D camera images with Deep Learning and Machine Learning**. 2022. Tese (Doutorado) – Universidade Federal do Ceara, Fortaleza, 2022.
- ANDRES, E. **Batch Norm Explained Visually**: How it works and why neural networks need it. 2019. Disponível em: <https://towardsdatascience.com/batch-norm-explained-visually-how-it-works-and-why-neural-networks-need-it-b18919692739>. Acesso em: 10 de mar. de 2023.
- BAYDIN, A. G.; PEARLMUTTER, B. A.; RADUL, A. A.; SISKIND, J. M. Automatic differentiation in machine learning: a survey. **Journal of Machine Learning Research**, Microtome Publishing, v. 18, p. 1–43, 2018. Disponível em: <https://www.jmlr.org/papers/volume18/17-468/17-468.pdf>. Acesso em: 10 de dez. de 2022.
- BEHNKE, A. R.; FEEN, B.; WELHAM, W. The specific gravity of healthy men: body weight ÷ volume as an index of obesity. **Journal of the American Medical Association**, American Medical Association, v. 118, n. 7, p. 495–498, 1942. Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/j.1550-8528.1995.tb00152.x> Acesso em: 10 de out. de 2022.
- BENGIO, Y.; GOODFELLOW, I. J.; COURVILLE, A. Deep learning. **Nature**, Citeseer, v. 521, n. 7553, p. 436–444, 2015.
- BENGIO, Y.; SIMARD, P.; FRASCONI, P. Learning long-term dependencies with gradient descent is difficult. **IEEE Transactions on Neural Networks**, Institute of Electrical and Electronics Engineers (IEEE), v. 5, n. 2, p. 157–166, mar 1994. Disponível em: <https://doi.org/10.1109/72.279181>. Acesso em: 09 de jan de 2023.
- BISHOP, C. M. **Pattern recognition and machine learning**. Secaucus, NJ: Springer, 2006. ISBN 0387310738.
- BOISVERT, J.; SHU, C.; WUHRER, S.; XI, P. Three-dimensional human shape inference from silhouettes: Reconstruction and validation. **Machine Vision and Applications**, v. 24, n. 1, p. 145–157, January 2013. ISSN 1432-1769.

BRODIE, D.; MSC, H. V. M.; HUTCHEON, R. Body composition measurement: a review of hydrodensitometry, anthropometry, and impedance methods. **Nutrition**, Elsevier, v. 14, n. 3, p. 296–310, 1998.

CAROLINA ARTHRITIS CENTER. **What a DXA Screening Can Tell You About Your Bones**. 2016. Disponível em: <https://www.carolinaarthritis.com/what-a-dxa-screening-can-tell-you-about-your-bones/>. Acesso em: 12 de dez. de 2022.

CASADEI, K.; KIEL, D. Anthropometry in clinical nutrition. **Nutrition in Clinical Practice**, Sage Publications Sage CA, v. 37, n. 1, p. 48–58, 2022.

CHIONG, R.; FAN, Z.; HU, Z.; CHIONG, F. Using an improved relative error support vector machine for body fat prediction. **Computer Methods and Programs in Biomedicine**, Elsevier, Amsterdam, Netherlands, v. 198, p. 105749, 2021.

CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2017, Honolulu. Anais [...]. [S. l.]: IEEE computational intelligence society, 2017. p. 1251–1258. Disponível em: <https://doi.org/10.1109/cvpr.2017.195> . Acesso em 12 de fev. de 2023.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: IEEE. **Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on**. [S. l.], 2009. p. 248–255.

DO, S.; SONG, K. D.; CHUNG, J. W. Basics of deep learning: a radiologist's guide to understanding published radiology articles on deep learning. **Korean journal of radiology**, The Korean Society of Radiology, v. 21, n. 1, p. 33–41, 2020. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6960318/pdf/kjr-21-33.pdf>. Acesso em: 17 de jan. de 2023.

DURNIN, J.; RAHAMAN, M. M. The assessment of the amount of fat in the human body from measurements of skinfold thickness. **British journal of Nutrition**, Cambridge University Press, Cambridge, England, v. 21, n. 3, p. 681–689, 1967.

EDMONDS, C.; JASANI, B.; SMITH, T. Total body potassium and body fat estimation in relationship to height, sex, age, malnutrition and obesity. **Clinical science and molecular medicine**, Portland Press Ltd., London, United Kingdom, v. 48, n. 5, p. 431–440, 1975.

FERENCI, T.; KOVACS, L. Predicting body fat percentage from anthropometric and laboratory measurements using artificial neural networks. **Applied Soft Computing**, Elsevier, Amsterdam, Netherlands, v. 67, p. 834–839, 2018.

FUKUSHIMA, K.; MIYAKE, S. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In: Competition and cooperation in neural nets. [S. l.]: Springer, 1982, p. 267–285.

GALLAGHER, D.; HEYMSFIELD, S. B.; HEO, M.; JEBB, S. A.; MURGATROYD, P. R.; SAKAMOTO, Y. Healthy percentage body fat ranges: an approach for developing guidelines based on body mass index. **The American Journal of Clinical Nutrition**, v. 72, n. 3, p. 694–701, 2000.

GENERAL ASSEMBLY OF THE UNITED NATIONS. **The universal declaration of human rights (Resolution 217 A)**. [S. l.], 1948. Disponível em: <https://www.un.org/en/universal-declaration-human-rights/> . Acesso em: 10 de nov. de 2023.

GLOROT, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. *In: PROCEEDINGS OF THE THIRTEENTH INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND STATISTICS*, 2010, Sardinia. Anais[...]. Sardinia: PMLR, 2010. p. 249-256. Disponível em: <https://proceedings.mlr.press/v9/glorot10a.html/>. Acesso em: 10 de nov. de 2022.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, Massachusetts: MIT Press, 2016.

GOYAL, P.; DOLLÁR, P.; GIRSHICK, R.; NOORDHUIS, P.; WESOLOWSKI, L.; KYROLA, A.; TULLOCH, A.; JIA, Y.; HE, K. Accurate, large minibatch sgd: Training imagenet in 1 hour. **arXiv preprint arXiv:1706.02677**, 2017.

GRIEWANK, A. *et al.* On automatic differentiation. **Mathematical Programming: recent developments and applications**, CiteSeer, v. 6, n. 6, p. 83–107, 1989.

GUERRA, R.; AMARAL, T. F.; MARQUES, E.; MOTA, J.; RESTIVO, M. Accuracy of siri and brozek equations in the percent body fat estimation in older adults. **The journal of nutrition, health & aging**, Springer, v. 14, n. 9, p. 744–748, 2010. Disponível em: <https://doi.org/10.1007/s12603-010-0112-z> . Acesso em: 07 de out. de 2022.

HAAPALA, I.; HIRVONEN, A.; NISKANEN, L.; UUSITUPA, M.; KRÖGER, H.; ALHAVA, E.; NISSINEN, A. Anthropometry, bioelectrical impedance and dual-energy x-ray absorptiometry in the assessment of body composition in elderly finnish women. **Clinical physiology and functional imaging**, Wiley Online Library, v. 22, n. 6, p. 383–391, 2002. Disponível em: <https://doi.org/10.1046/j.1475-097X.2002.00447.x> . Acesso em: 15 de jan. de 2023.

HARTY, P. S.; SIEGLINGER, B.; HEYMSFIELD, S. B.; SHEPHERD, J. A.; BRUNER, D.; STRATTON, M. T.; TINSLEY, G. M. Novel body fat estimation using machine learning and 3-dimensional optical imaging. **European Journal of Clinical Nutrition**, Nature Publishing Group, p. 1–4, 2020. Disponível em: <https://www.nature.com/articles/s41430-020-0603-x>. Acesso em: 14 de jul. de 2022.

HAYKIN, S. **Neural Networks and Learning Machines**. McMaster University, Canada: Prentice Hall, 2008.

HE, T.; ZHANG, Z.; ZHANG, H.; ZHANG, Z.; XIE, J.; LI, M. Bag of tricks for image classification with convolutional neural networks. *In: PROCEEDINGS OF THE IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION*, 2019, "on line". Anais eletrônicos[...]. [S. l.]: IEEE computational intelligence society, 2019. p. 558–567. Disponível em : <https://arxiv.org/abs/1812.01187>. Acesso em: 16 de mar. de 2023.

HEYMSFIELD, S. B.; WANG, Z.; BAUMGARTNER, R. N.; ROSS, R. Human body composition: advances in models and methods. **Annual review of nutrition**, Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA, v. 17, n. 1, p. 527–558, 1997.

HEYWARD, V. H. Evaluation of body composition: current issues. **Sports Medicine**, Springer, Berlin, Germany, v. 22, p. 146–156, 1996.

HEYWARD, V. H. Practical body composition assessment for children, adults, and older adults. **International Journal of Sport Nutrition and Exercise Metabolism**, Human Kinetics, Inc., v. 8, n. 3, p. 285–307, 1998. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/9738136/>. Acesso em: 11 de ago. de 2022.

HINTON, G.; VINYALS, O.; DEAN, J. Distilling the Knowledge in a Neural Network. **ArXiv e-prints**, mar. 2015.

HOLMES, C. J.; RACETTE, S. B. The utility of body composition assessment in nutrition and clinical practice: an overview of current methodology. **Nutrients**, MDPI, v. 13, n. 8, p. 2493, 2021.

HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. **CoRR**, abs/1704.04861, 2017.

HUANG, G.; LIU, Z.; MAATEN, L. van der; WEINBERGER, K. Q. Densely connected convolutional networks. *In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION CVPR, 2017, Honolu. anais[...]*. Honolu: IEEE, 2017. Disponível em: <https://doi.org/10.1109/cvpr.2017.243> . Acesso em: 10 de set de 2022.

IOFFE, S.; SZEGEDY, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. **ArXiv e-prints**, fev. 2015.

JACKSON, A. S.; POLLOCK, M.; WARD, A. Generalized equations for predicting body density of men. **British Journal of Nutrition**, Cambridge University Press, Cambridge, England, v. 43, n. 1, p. 1–10, 1980.

JACKSON, A. S.; POLLOCK, M. L. Generalized equations for predicting body density of men. **British Journal of Nutrition**, Cambridge University Press, Cambridge, England, v. 40, n. 3, p. 497–504, 1978.

JANSSEN, I.; HEYMSFIELD, S. B.; ROSS, R. Low relative skeletal muscle mass (sarcopenia) in older persons is associated with functional impairment and physical disability. **Journal of the american geriatrics society**, Wiley Online Library, v. 50, n. 5, p. 889–896, 2002. Disponível em: <https://agsjournals.onlinelibrary.wiley.com/doi/10.1046/j.1532-5415.2002.50216.x>. Acesso em: 14 de out. de 2022.

JANSSEN, I.; HEYMSFIELD, S. B.; WANG, Z.; ROSS, R. Skeletal muscle mass and distribution in 468 men and women aged 18–88 yr. **Journal of applied physiology**, American Physiological Society Bethesda, MD, 2000. Disponível em: <https://journals.physiology.org/doi/full/10.1152/jappl.2000.89.1.81>. Acesso em: 16 de jan. de 2023.

JIN, J.; DUNDAR, A.; CULURCIELLO, E. Flattened convolutional neural networks for feedforward acceleration. **CoRR**, abs/1412.5474, 2014. Software available from tensorflow.org. Disponível em: <http://arxiv.org/abs/1412.5474/> . Acesso em: 08 de ago. de 2022.

KASPER, A. M.; LANGAN-EVANS, C.; HUDSON, J. F.; BROWNLEE, T. E.; HARPER, L. D.; NAUGHTON, R. J.; MORTON, J. P.; CLOSE, G. L. Come back skinfolds, all is forgiven: a narrative review of the efficacy of common body composition methods in applied sports practice. **Nutrients**, MDPI, v. 13, n. 4, p. 1075, 2021.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.

KNAPP, K. Dual energy x-ray absorptiometry: beyond bone mineral density. **Imaging & Therapy Practice**, Society & College of Radiographers, p. 5, 2015.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *In: Advances in neural information processing systems*. p. 1097–1105, 2012.

KULLBERG, J.; BRANDBERG, J.; ANGELHED, J.-E.; FRIMMEL, H.; BERGELIN, E.; STRID, L.; AHLSTROM, H.; JOHANSSON, L.; LONN, L. Whole-body adipose tissue analysis: comparison of mri, ct and dual energy x-ray absorptiometry. **The British journal of radiology**, British Institute of Radiology, London, United Kingdom, v. 82, n. 974, p. 123–130, 2009.

KURIYAN, R. Body composition techniques. **The Indian journal of medical research**, Wolters Kluwer–Medknow Publications, v. 148, n. 5, p. 648, 2018.

KYLE, U. G.; BOSAEUS, I.; LORENZO, A. D. D.; DEURENBERG, P.; ELIA, M.; GOMEZ, J. M.; HEITMANN, B. L.; KENT-SMITH, L.; MELCHIOR, J.-C.; PIRLICH, M. Bioelectrical impedance analysis–part i: review of principles and methods. **Clinical Nutrition**, Elsevier, Amsterdam, Netherlands, v. 23, n. 5, p. 1226–1243, 2004.

LÄNGKVIST, M.; KARLSSON, L.; LOUTFI, A. A review of unsupervised feature learning and deep learning for time-series modeling. **Pattern Recognition Letters**, Elsevier BV, v. 42, p. 11–24, jun 2014. Disponível em: <https://doi.org/10.1016/j.patrec.2014.01.008/> . Acesso em: 16 de dez. de 2022.

LECUN, Y.; BOSER, B. E.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W. E.; JACKEL, L. D. Handwritten digit recognition with a back-propagation. *In: Advances in neural information processing systems*. p. 396–404, 1990.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, IEEE, v. 86, n. 11, p. 2278–2324, 1998.

LIN, T.; MAIRE, M.; BELONGIE, S. J.; BOURDEV, L. D.; GIRSHICK, R. B.; HAYS, J.; PERONA, P.; RAMANAN, D.; DOLLÁR, P.; ZITNICK, C. L. Microsoft COCO: common objects in context. **CoRR**, abs/1405.0312, 2014.

LIU, C.; FAN, X.; GUO, Z.; MO, Z.; CHANG, E. I.-C.; XU, Y. Wound area measurement with 3d transformation and smartphone images. **BMC Bioinformatics**, v. 20, n. 724, 2019. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3308-1>. Acesso em: 11 de mar. de 2023.

LIU, L.; ZHANG, X.; LI, Z.; LI, X.; DONG, X.; LIU, S.; HU, P. Segmentation of liver, muscle, and fat using deep convolutional neural network and 3d-deformable models. **Journal of medical systems**, Springer, Berlin, Germany, v. 45, n. 6, p. 1–11, 2021.

LOHMAN, T. G.; ROCHE, A. F.; MARTORELL, R. *et al.* **Anthropometric standardization reference manual**. [S. l.]: Human kinetics books, 1988.

LOSHCHILOV, I.; HUTTER, F. Sgdr: Stochastic gradient descent with warm restarts. **arXiv preprint arXiv:1608.03983**, 2016.

- LUKASKI, H. C. Evolution of bioimpedance: a circuitous journey from estimation of physiological function to assessment of body composition and a return to clinical research. **European Journal of Clinical Nutrition**, Nature Publishing Group, Berlin, Germany, v. 67, n. Suppl 1, p. S2–S9, 2013.
- MAJMUDAR, M. D.; CHANDRA, S.; YAKKALA, K.; KENNEDY, S.; AGRAWAL, A.; SIPPEL, M.; RAMU, P.; CHAUDHRI, A.; SMITH, B.; CRIMINISI, A. *et al.* Smartphone camera based assessment of adiposity: a validation study. **NPJ Digital Medicine**, Nature Publishing Group UK London, London, United Kingdom, v. 5, n. 1, p. 79, 2022.
- MARGOSSIAN, C. C. A review of automatic differentiation and its efficient implementation. **Wiley interdisciplinary reviews: data mining and knowledge discovery**, Wiley Online Library, v. 9, n. 4, p. e1305, 2019.
- NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. *In*: FÜRKRANZ, J.; JOACHIMS, J. (ed.). ICML. Omnipress, p. 807–814, 2010. Disponível em: <http://dblp.uni-trier.de/db/conf/icml/icml2010.htmlNairH10/> . Acesso em: 20 de jun. de 2023.
- NÓBREGA, R. V. M. da; REBOUÇAS FILHO, P. P.; RODRIGUES, M. B.; SILVA, S. P. P. da; DOURADO JÚNIOR, C. M. J. M.; ALBUQUERQUE, V. H. C. de. Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks. **Neural Computing and Applications**, 2018.
- PARIS, M. T.; TANDON, P.; HEYLAND, D. K.; FURBERG, H.; PREMJI, T.; LOW, G.; MOURTZAKIS, M. Automated body composition analysis of clinically acquired computed tomography scans using neural networks. **Clinical Nutrition**, Elsevier, Amsterdam, Netherlands, v. 39, n. 10, p. 3049–3055, 2020.
- PASZKE, A.; GROSS, S.; CHINTALA, S.; CHANAN, G.; YANG, E.; DEVITO, Z.; LIN, Z.; DESMAISON, A.; ANTIGA, L.; LERER, A. Automatic differentiation in pytorch. *In*: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS, 31, 2017, Long Beach. Anais[...]. Long Beach: NIPS, 2017. p. 1-4.
- PEREIRA, C. R.; PEREIRA, D. R.; ROSA, G. H.; ALBUQUERQUE, V. H.; WEBER, S. A.; HOOK, C.; PAPA, J. P. Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification. **Artificial Intelligence in Medicine**, v. 87, p. 67 – 77, 2018.
- POLLOCK, M.; SCHMIDT, D.; JACKSON, A. Prediction of body density in young and middle-aged women. **Journal of Applied Physiology**, American Physiological Society Bethesda, MD, v. 46, n. 6, p. 1166–1170, 1979. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/1141107/>. Acesso em: 17 de nov. de 2022.
- REDMON, J.; FARHADI, A. Yolov3: An incremental improvement. **arXiv preprint arXiv:1804.02767**, 2018.
- ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, p. 386–408, 1958.
- ROTHMAN, K. J. Bmi-related errors in the measurement of obesity. **International journal of obesity**, Nature Publishing Group, Berlin, Germany, v. 32, p. S56–S59, 2008.

- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, Springer Nature, v. 323, n. 6088, p. 533–536, oct 1986. Disponível em: <https://doi.org/10.1038/323533a0/> . Acesso em: 06 de abr. de 2023.
- RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATY, A.; KHOSLA, A.; BERNSTEIN, M.; BERG, A. C.; FEI-FEI, L. ImageNet Large Scale Visual Recognition Challenge. **International Journal of Computer Vision (IJCV)**, Springer, Berlin, Germany, v. 115, n. 3, p. 211–252, 2015.
- SELVARAJU, R. R.; COGSWELL, M.; DAS, A.; VEDANTAM, R.; PARIKH, D.; BATRA, D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. **International Journal of Computer Vision**, Springer Science and Business Media LLC, v. 128, n. 2, p. 336–359, oct 2019. Disponível em: <https://doi.org/10.1007>
- SHAO, Y. E. Body fat percentage prediction using intelligent hybrid approaches. **The Scientific World Journal**, Hindawi, London, United Kingdom, v. 2014, 2014.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **CoRR**, abs/1409.1556, 2014. Disponível em: <https://dblp.org/rec/bib/journals/corr/SimonyanZ14a>. Acesso em: 14 de set. de 2022.
- SIRI, W. E. The gross composition of the body. *In: advances in biological and medical physics*. Academic Press, 1956, v. 4, p. 239–280.
- SRIKANTHAN, P.; KARLAMANGLA, A. S. Relative muscle mass is inversely associated with insulin resistance and prediabetes. findings from the third national health and nutrition examination survey. **The Journal of Clinical Endocrinology & Metabolism**, Oxford University Press, Oxford, England, v. 96, n. 9, p. 2898–2903, 2011.
- SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. Dropout: a simple way to prevent neural networks from overfitting. **The Journal of Machine Learning Research**, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.
- STEINBERGER, J.; JACOBS, D.; RAATZ, S.; MORAN, A.; HONG, C.; SINAIKO, A. R. Comparison of body fatness measurements by bmi and skinfolds vs dual energy x-ray absorptiometry and their relation to cardiovascular risk factors in adolescents. **International journal of obesity**, Nature Publishing Group, Berlin, Germany, v. 29, n. 11, p. 1346–1352, 2005.
- SWAINSON, M. G.; BATTERHAM, A. M.; TSAKIRIDES, C.; RUTHERFORD, Z. H.; HIND, K. Prediction of whole-body fat percentage and visceral adipose tissue mass from five anthropometric variables. **PLOS ONE**, Public Library of Science, v. 12, n. 5, p. 1–12, 05 2017. Disponível em: <https://doi.org/10.1371/journal.pone.0177175> . Acesso em: 03 de mai. de 2023.
- SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCHE, V.; RABINOVICH, A. Going deeper with convolutions. *In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION CVPR*, 2015, Boston. Anais[...]. Boston: IEEE Institute of Electrical and Electronics Engineers, 2015. p. 1-9. Disponível em: <https://doi.org/10.1109/cvpr.2015.7298594> . Acesso: 11 de abr. de 2023.
- SZEGEDY, C.; VANHOUCHE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. Rethinking the inception architecture for computer vision. *In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION CVPR*, 2016, Las Vegas. Anais[...]. Las vegas:

IEEE institute of Electrical and Electronics Engineers, 2016. p. 2818-2826. Disponível em: <https://doi.org/10.1109/cvpr.2016.308>. Acesso em: 06 de nov. de 2022.

TIAN, I. Y.; NG, B. K.; WONG, M. C.; KENNEDY, S.; HWAUNG, P.; KELLY, N.; LIU, E.; GARBER, A. K.; CURLESS, B.; HEYMSFIELD, S. B. *et al.* Predicting 3d body shape and body composition from conventional 2d photography. **Medical Physics**, Wiley Online Library, v. 47, n. 12, p. 6232–6245, 2020.

TORRES, A.; UMEDA, K.; ISHIDA, T.; YASAKA, K.; TERAMOTO, A.; FUJITA, H. Body composition estimation using convolutional neural networks. **International Journal of Computer Assisted Radiology and Surgery**, Springer, Berlin, Germany, v. 14, n. 9, p. 1529–1539, 2019.

UÇAR, M. K.; UCAR, Z.; KÖKSAL, F.; DALDAL, N. Estimation of body fat percentage using hybrid machine learning algorithms. **Measurement**, Elsevier, Amsterdam, Netherlands, v. 167, p. 108173, 2021.

UÇAR, M. K.; UÇAR, Z.; KÖKSAL, F.; DALDAL, N. Estimation of body fat percentage using hybrid machine learning algorithms. **Measurement**, v. 167, p. 108173, 2021. ISSN 0263-2241. Disponível em: <https://doi.org/10.1016/j.measurement.2020.108173> . Acesso em: 02 de ago. de 2022.

WANG, Y. C.; MCPHERSON, K.; MARSH, T.; GORTMAKER, S. L.; BROWN, M. Health and economic burden of the projected obesity trends in the usa and the uk. **The Lancet**, Elsevier, Amsterdam, Netherlands, v. 378, n. 9793, p. 815–825, 2011.

WANG, Z.; HUANG, C.; LI, L.; GUO, J.; ZHAO, Q.; LI, Q.; WEI, G. Deep neural network based on magnetic resonance imaging for body composition analysis. **Journal of magnetic resonance imaging**, Wiley Online Library, v. 51, n. 1, p. 270–277, 2020.

WANG, Z.; VISSER, M.; MA, R.; BAUMGARTNER, R. N.; KOTLER, D.; GALLAGHER, D.; HEYMSFIELD, S. B. Skeletal muscle mass: evaluation of neutron activation and dual-energy x-ray absorptiometry methods. **Journal of Applied Physiology**, American Physiological Society, v. 80, n. 3, p. 824–831, 1996. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/8964743/>. Acesso em: 11 de dez. de 2022.

WANG, Z.-M.; JR, R. N. P.; HEYMSFIELD, S. B. The five-level model: a new approach to organizing body-composition research. **The American journal of clinical nutrition**, Oxford University Press, Oxford, England, v. 56, n. 1, p. 19–28, 1992.

WILSON, R. A. **Comparison of Body Composition Assessment Techniques in Older Adults**. Knoxville, Tennessee: University of Tennessee, 2005.

WORLD HEALTH ORGANIZATION. **Physical status**: the use and interpretation of anthropometry. report of a who expert committee. 1995.

WORLD HEALTH ORGANIZATION. **Obesity**: preventing and managing the global epidemic. report of a who consultation. 1998.

WORLD HEALTH ORGANIZATION. **Policy brief**: The double burden of malnutrition (who/nmh/nhd/17.2). [S. l.], 2017. Disponível em: <https://apps.who.int/iris/handle/10665/255414>.

WORLD HEALTH ORGANIZATION. **Obesity and overweight**. 2021. Disponível em: <https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight>. Acesso em: 01 de mai. de 2023.

WU, S.; ZHONG, S.; LIU, Y. Deep residual learning for image steganalysis. **Multimedia Tools and Applications**, Springer Nature, v. 77, n. 9, p. 10437–10453, feb 2017. Disponível em: <https://doi.org/10.1007/s11042-017-4440-4>. Acesso em: 01 de jul 2022.

XIANG, Y.; SCHMIDT, T.; NARAYANAN, V.; FOX, D. Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. **CoRR**, abs/1711.00199, 2017.

XIAOHUI, T.; XIAOYU, P.; LIWEN, L.; QING, X. Automatic human body feature extraction and personal size measurement. **Journal of Visual Languages & Computing**, Elsevier, Amsterdam, Netherlands, v. 47, p. 9 – 18, 2018.

ZEILER, M. D.; FERGUS, R. Visualizing and understanding convolutional networks. *In*: Computer vision – ECCV 2014. Springer International Publishing, 2014, p. 818–833. Acesso em: 04 de mar. de 2023. Disponível em: https://doi.org/10.1007/978-3-319-10590-1_53.

ZOPH, B.; LE, Q. V. Neural architecture search with reinforcement learning. **CoRR**, abs/1611.01578, 2016.