

**USO DE MODELOS ESTATÍSTICOS E TÉCNICAS DE  
APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DO  
PERCENTUAL DE GORDURA**

**SÃO LUÍS, MA  
AGOSTO – 2023**

**MICHELE BEZERRA SILVA**

**USO DE MODELOS ESTATÍSTICOS E TÉCNICAS DE APRENDIZADO DE  
MÁQUINA PARA PREDIÇÃO DO PERCENTUAL DE GORDURA**

Tese apresentada ao Programa de Pós-Graduação em Saúde Coletiva da Universidade Federal do Maranhão como requisito parcial à obtenção do título de Doutor em Saúde Coletiva.

**Orientador:** Profa. Dra. Alcione Miranda dos Santos

**Coorientador:** Profa. Dra. Ana Karina Teixeira da Cunha França

**SÃO LUÍS, MA**

**AGOSTO – 2023**

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).  
Diretoria Integrada de Bibliotecas/UFMA

Bezerra Silva, Michele.

USO DE MODELOS ESTATÍSTICOS E TÉCNICAS DE APRENDIZADO  
DE MÁQUINA PARA PREDIÇÃO DO PERCENTUAL DE GORDURA /  
Michele Bezerra Silva. - 2023.

141 f.

Coorientador(a): Ana Karina Teixeira da Cunha França.

Orientador(a): Alcione Miranda dos Santos.

Tese (Doutorado) - Programa de Pós-graduação em Saúde  
Coletiva/ccbs, Universidade Federal do Maranhão, São Luís,  
2023.

1. Modelos preditivos. 2. Percentual de gordura  
corporal. 3. Pletismografia por deslocamento de ar. I.  
Karina Teixeira da Cunha França, Ana. II. Miranda dos  
Santos, Alcione. III. Título.

**USO DE MODELOS ESTATÍSTICOS E TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DO PERCENTUAL DE GORDURA**

Michele Bezerra Silva

Tese aprovada em \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_ pela banca examinadora constituída dos seguintes membros:

**Banca Examinadora:**

---

Prof. Dra. Alcione Miranda dos Santos  
Orientadora  
Universidade Federal do Maranhão

---

Prof. Dra. Ana Karina Teixeira da Cunha França  
Coorientadora  
Universidade Federal do Maranhão

---

Prof. Dra. Maria da Conceição Chaves de Lemos  
Examinador Externo  
Universidade Federal de Pernambuco

---

Prof. Dra. Joelma Ximenes Prado Teixeira  
Examinador Externo  
Universidade Federal do Maranhão

---

Prof. Dr. Bruno Feres de Souza  
Examinador Interno  
Universidade Federal do Maranhão

---

Prof. Dra. Vanda Maria Ferreira Simões  
Examinador Interno  
Universidade Federal do Maranhão

Dedico esta tese à minha família amada

## **AGRADECIMENTOS**

A Deus, pelo dom da vida e por ter dado saúde e forças para dar seguimento à esta jornada;

Aos meus pais, Maria da Graça Pereira Bezerra e Alzemir Carvalho Silva, que constituem meu alicerce, e meus irmãos Milena Bezerra Silva e Raphael Bruno Bezerra Silva, pelo cuidado e carinho. Sem o apoio e incentivo de vocês nada disso seria possível!

À Universidade Federal do Maranhão (UFMA), por proporcionar um ambiente acadêmico estimulante, desde a graduação, e ao Programa de Pós-Graduação em Saúde Coletiva (PGSC), por proporcionar base científica, acesso a recursos e infraestrutura que possibilitaram a realização desta tese;

À minha orientadora, Alcione Miranda dos Santos, por sua orientação cuidadosa, paciência e dedicação ao longo de toda a jornada de pesquisa. Seu incentivo foi precioso em momentos onde meu cansaço queria me dominar. Suas orientações, incentivo e amizade foram fundamentais para o desenvolvimento deste trabalho, e sou grata pela oportunidade!

À minha coorientadora, Ana Karina Teixeira da Cunha França, a quem admiro desde a graduação! Agradeço por sua orientação cuidadosa e contribuições valorosas!

Ao Instituto Federal do Piauí – Campus Uruçuí, por ter concedido o afastamento por quase metade do período do Doutorado;

Aos professores do PGSC, que compartilharam seus conhecimentos e suscitaram discussões, que enriqueceram meu aprendizado e ampliando minha visão sobre a área de Saúde Coletiva.

À Secretaria do PGSC, pelo suporte administrativo e organizacional ao longo do curso.

Aos meus colegas do PGSC, pelas trocas de ideias, debates e apoio mútuo ao longo dessa jornada, em especial à Lívia Carolina Sobrinho Rudakoff, Alessandra Karla Oliveira Amorim Muniz e Renata de Sousa Gomes, cuja construção de laços de amizade e colaboração foi enriquecedora e tornou essa experiência ainda mais leve.

Ao grupo de pesquisa em Ciência de Dados em Saúde, por promover discussões, auxílio na elaboração desta tese e pela amizade! Muito obrigada!

Aos meus queridos amigos, Mabel Nery Ribeiro, Raianne Rodrigues Lima, Angélica Cristina Silva Marques, Jefferson Rodrigues de Souza, Andréia da Silva Costa, Eduarda Maria de Sousa e Luis Augusto Silva Maciel, obrigada por cada palavra de apoio, carinho e incentivo!

Por fim, meu agradecimento estende-se a todos que, de alguma forma, contribuíram para finalização deste trabalho e para o meu crescimento acadêmico e pessoal ao longo deste percurso.

“Ensinar não é transferir conhecimento, mas  
criar as possibilidades para a sua própria  
produção ou a sua construção.”

(Paulo Freire)

## LISTA DE TABELAS

### Artigo 1

<b>Table 1.</b> Demographic and anthropometric characteristics of the sample of individuals from the Consortium of Cohorts of São Luís - MA, 2016. ....	61
<b>Table 2.</b> Evaluation of the performance of the performance of the ANNs by sampling .....	62
<b>Table 3.</b> Mean and standard deviation (SD) of fat percentage estimated by neural networks and ADP for the test sample .....	63
<b>Table 4.</b> Importance of variables in artificial neural network models for body fat prediction considering generalized and sex-specific models.....	64

### Artigo 2

<b>Tabela 1.</b> Características demográficas e antropométricas de amostra de indivíduos do Consórcio de Coortes de Ribeirão Preto - SP e São Luís - MA, 2016 e 2017.....	86
<b>Tabela 2.</b> Valores médios de obesidade de acordo com o IMC e %GC, do Consórcio de Coortes de Ribeirão Preti - SP e São Luís - MA, 2016 e 2017.....	87
<b>Tabela 3.</b> Valores de $R^2$ , RMSE e NRMSE dos modelos de regressão .....	88
<b>Tabela 4.</b> Reprodutibilidade de validade de modelos e equação para predição de percentual de gordura corporal (%GC).....	89

## LISTA DE ABREVIATURAS E SIGLAS

AAD	- Média absoluta das diferenças
AUC	- <i>Area under the curve</i>
BIA	- Bioimpedância elétrica
CC	- Circunferência da cintura
CP	- Circunferência da panturrilha
CQ	- Circunferência do quadril
DEXA	- Absorciometria por Raios X de Dupla Energia
D <sub>2</sub> O	- Diluição com óxido de deutério
%GC	- Percentual de gordura corporal
%GV	- Percentual de Gordura Visceral
IC	- Intervalo de Confiança
IMC	- Índice de Massa Corpórea
LCC	- Coeficiente de Concordância de Lin
MAD	- Desvio Médio Absoluto
MAE	- Erro Médio Absoluto
MAPE	- Erro percentual absoluto médio
MPA	- Acurácia preditiva média
MRL	- Modelo de Regressão Linear
MR-MARS	- Modelo de Regressão e Regressão Adaptativa multivariada
NRME	- Raiz do Erro médio quadrático normalizado
PDA	- Pletismografia por deslocamento de ar
OMS -	Organização Mundial da Saúde
R <sup>2</sup>	- Coeficiente de determinação
RCE	- Relação cintura-estatura
RCQ	- Relação cintura-quadril
RMSE	- Raiz do Erro médio quadrático
RNA	- Rede Neural Artificial
ROC	- <i>Receiver Operator Characteristic</i>
TCLE	- Termo de Consentimento Livre e Esclarecido
UWW	- Pesagem subaquática

SILVA, Michele Bezerra, **USO DE MODELOS ESTATÍSTICOS E TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DO PERCENTUAL DE GORDURA**, 2023, Tese (Doutorado em Saúde Coletiva) - Programa de Pós-graduação em Saúde Coletiva, Universidade Federal do Maranhão, São Luís, 2023.

## RESUMO

A obesidade, caracterizada como o excesso de gordura corporal, é um problema epidêmico, com impactos na diminuição da qualidade de vida dos indivíduos e aumento dos gastos em saúde pública. Dentre as metas para combater obesidade, propostas pela Organização Mundial de Saúde, destacam-se a ênfase na prevenção e diagnóstico na atenção primária. Nesse sentido, o objetivo desta tese foi desenvolver modelos de previsão para estimar o percentual de gordura corporal (%GC) em adolescentes e adultos jovens, a partir de medidas antropométricas simples. Os resultados obtidos possibilitaram a elaboração de dois artigos: "Previsão do percentual de gordura em adolescentes a partir de medidas antropométricas por redes neurais artificiais" e "Avaliação da precisão de equações preditivas baseadas em medidas antropométricas simples no diagnóstico da obesidade". O objetivo do **primeiro artigo** foi desenvolver modelos baseados em Rede Neural Artificial (RNA) para estimar o percentual de gordura corporal (%GC) em adolescentes. Tratou-se de um estudo transversal com 2.155 adolescentes (18 e 19 anos de idade). A percentagem de gordura foi medida por pletismografia de deslocamento de ar (PDA). Análises estatísticas foram realizadas no programa R, versão 4.3.0. Foram implementados diferentes modelos de RNA *feedforward*, generalizados e específicos por sexo, com diferentes combinações de variáveis sexo, idade (anos), peso (kg), estatura (cm), circunferência da cintura (CC), índice de massa corporal (IMC) e relação cintura estatura (RCE). Os modelos de RNA generalizados apresentaram melhor performance ( $R^2 > 0,75$ ) em comparação com os modelos por sexo ( $R^2 \leq 0,72$ ). A CC foi uma variável de alta importância, especialmente entre meninos. Não foram observadas diferenças entre o percentual de gordura medido e o estimado pela rede neural artificial ( $p > 0,05$ ). Os modelos de RNA desenvolvidos a partir de medidas antropométricas simples mostraram-se eficazes na previsão de %GC em adolescentes. A incorporação da CC apresentou diferentes respostas na estimativa de desempenho do %GC para meninos e meninas. O **segundo artigo** teve como objetivo avaliar a acurácia de equações preditivas para diagnóstico de obesidade. Estudo transversal com 3.103 indivíduos (18 a 23 anos de idade). O método de referência para medir %GC foi a PDA, análises estatísticas foram realizadas no programa R, versão 4.3.0. Regressão linear múltipla (MLR) foi desenvolvida para elaboração das equações, usando idade, sexo, além das variáveis antropométricas simples, descritas no primeiro artigo. As equações desenvolvidas mostraram um bom desempenho preditivo ( $R^2 \approx 0,80$  e  $NRMSE \approx 0,09$ ). Além disso, mostraram maior capacidade de previsão do diagnóstico da obesidade (especificidade  $\approx 0,64$ ; falso positivo =  $0,36$ ;  $AUC \approx 0,9$ ) quando comparado apenas ao IMC (especificidade  $\approx 0,39$ ; falso positivo =  $0,89$ ;  $AUC \approx 0,2$ ). As equações desenvolvidas a partir de medidas antropométricas facilmente aplicáveis à prática clínica mostraram uma elevada capacidade de previsão e uma maior capacidade de identificação de pessoas com obesidade comparada ao IMC. As equações desenvolvidas neste estudo e podem ser úteis para estimar %GC, na ausência do método de referência, podendo ser utilizadas como ferramenta prática para acompanhamento do %GC e diagnóstico de obesidade na atenção primária em saúde.

**Palavras-chave:** Gordura corporal. Antropometria. Modelos preditivos

SILVA, Michele Bezerra, **USE OF STATISTICAL MODELS AND MACHINE LEARNING TECHNIQUES FOR PREDICTING BODY FAT PERCENTAGE**, 2023. 195f. Thesis (Doctorate) - Postgraduate Program in Public Health, Federal University of Maranhão, São Luís, 2023.

## ABSTRACT

Obesity, characterized as excess body fat, is an epidemic problem, with impacts on the quality of life of individuals and an increase in public health expenditure. Among the goals for combating obesity proposed by the World Health Organization is an emphasis on prevention and diagnosis in primary care. The aim of this thesis was to develop prediction models to estimate the percentage of body fat (%BF) in adolescents and young adults, based on simple anthropometric measurements. The results obtained enabled two articles to be written: "Predicting the percentage of fat in adolescents from anthropometric measurements using artificial neural networks" and "Assessing the accuracy of predictive equations based on simple anthropometric measurements in the diagnosis of obesity". The aim of the **first article** was to develop Artificial Neural Network (ANN)-based models to estimate body fat percentage (BFP) in adolescents using simple anthropometric measurements. This was a cross-sectional study with 2,155 adolescents (18 and 19 years old). Fat percentage was measured by air displacement plethysmography (ADP). Statistical analyses were performed in the R program, version 4.3.0. Different generalized and sex-specific feedforward ANN models were implemented with different combinations of sex, age (years), weight (kg), height (cm), waist circumference (WC), body mass index (BMI) and waist-to-height ratio (WHtR) variables. The generalized ANN models showed better performance ( $R^2 > 0.75$ ) compared to the models by sex ( $R^2 \leq 0.72$ ). WC was a variable of high importance, especially among boys. No differences were observed between measured and estimated fat percentage by artificial neural network ( $p > 0.05$ ). The ANN models developed from simple anthropometric measurements were effective in predicting %BF in adolescents. The incorporation of WC showed different responses in the estimation performance of %BF for boys and girls. The second article aimed to evaluate the accuracy of predictive equations for the diagnosis of obesity. Cross-sectional study with 3,103 individuals (18 to 23 years old) of both sexes. The reference method for measuring %BF and statistical program were the same as in the first article. Multiple linear regression (MLR) was developed to elaborate the equations, using age, sex, in addition to the simple anthropometric variables,

described in the first article. The developed equations showed a good predictive performance ( $R^2 \approx 0.80$  and  $NRMSE \approx 0.09$ ). In addition, they showed greater predictive ability for obesity diagnosis (specificity  $\approx 0.64$ ; false positive = 0.36;  $AUC \approx 0.9$ ) when compared to BMI alone (specificity  $\approx 0.39$ ; false positive = 0.89 ;  $AUC \approx 0.2$  ). The equations developed from measures that are easily applicable to clinical practice showed a high predictive ability and a greater ability to identify people with obesity than BMI. The equations developed in this study and may be useful for estimating %BF in the absence of the reference method and can be used as a practical tool for monitoring %BF and diagnosing obesity.

**Keywords:** Body fat. Anthropometry. Predictive models

## SUMÁRIO

<b>APRESENTAÇÃO .....</b>	<b>15</b>
<b>1 INTRODUÇÃO.....</b>	<b>16</b>
<b>2 OBJETIVOS.....</b>	<b>18</b>
<b>2.1 Geral .....</b>	<b>18</b>
<b>2.2 Específicos .....</b>	<b>18</b>
<b>3 REFERENCIAL TEÓRICO .....</b>	<b>19</b>
<b>3.1 Obesidade .....</b>	<b>19</b>
<b>3.2 Métodos para avaliação da composição corporal.....</b>	<b>20</b>
3.2.1 Pletismografia por deslocamento de ar – PDA .....	20
3.2.2 Medidas e Índices antropométricos baseados no peso, altura e CC.....	21
<b>3.3 Equações para predizer percentual de gordura a partir de medidas antropométricas simples, idade e sexo.....</b>	<b>24</b>
<b>3.4 Modelos para predizer percentual de gordura baseado em aprendizagem de máquinas .....</b>	<b>30</b>
<b>4 JUSTIFICATIVA .....</b>	<b>35</b>
<b>5 MATERIAL E MÉTODOS .....</b>	<b>36</b>
<b>5.1 Delineamento do estudo.....</b>	<b>36</b>
<b>5.2 População e amostra em estudo.....</b>	<b>36</b>
<b>5.3 Critérios de Inclusão.....</b>	<b>36</b>
<b>5.4 Critérios de Não Inclusão.....</b>	<b>36</b>
<b>5.5 Procedimentos de Obtenção dos Dados .....</b>	<b>36</b>
5.5.1 Dados socioeconômicos e demográficos/ hábitos de vida.....	36
5.5.2 Avaliação antropométrica e nutricional .....	37
5.5.3 Avaliação do percentual de gordural por Pletismografia por Deslocamento de Ar – PDA	37
<b>5.6 Análise Estatística .....</b>	<b>38</b>
5.6.1 Implementação da RNA e análise dos modelos desenvolvidos.....	38
5.6.2 Implementação e análise dos MRL.....	40
<b>5.7 Considerações éticas .....</b>	<b>42</b>
<b>6 RESULTADOS .....</b>	<b>43</b>
<b>6.1 Artigo 1 .....</b>	<b>43</b>

<b>6.2 Artigo 2</b> .....	<b>65</b>
<b>6 CONSIDERAÇÕES FINAIS</b> .....	<b>92</b>
<b>REFERÊNCIAS</b> .....	<b>94</b>
<b>ANEXO A – NORMAS DA REVISTA : ARTIGO 1</b> .....	<b>106</b>
<b>APÊNDICE 1 – Artigo publicado com a Orientadora</b> .....	<b>136</b>

