

**USO DE TÉCNICAS DE APRENDIZADO DE
MÁQUINA PARA ESTIMAÇÃO DE GORDURA
CORPORAL EM ADULTOS BRASILEIROS**

**SÃO LUÍS, MA
FEVEREIRO – 2026**

HELOÍSA BAIMA DA SILVA SANTOS

**USO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA ESTIMAÇÃO DE
GORDURA CORPORAL EM ADULTOS BRASILEIROS**

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

Orientador: Prof.^a Dr.^a Ana Karina Teixeira da Cunha França

Coorientador: Prof.^a Dr.^a Alcione Miranda dos Santos

**SÃO LUÍS, MA
FEVEREIRO – 2026**

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

Santos, Heloísa Baima da Silva.

Uso de técnicas de Aprendizado de Máquina para
estimação de gordura corporal em adultos brasileiros /
Heloísa Baima da Silva Santos. - 2026.

112 p.

Coorientador(a) 1: Alcione Miranda dos Santos.

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

Dissertação (Mestrado) - Programa de Pós-graduação em
Saúde Coletiva/ccbs, Universidade Federal do Maranhão, São
Luís, 2026.

1. Composição Corporal. 2. Aprendizado de Máquina. 3.
Algoritmos. 4. Antropometria. I. Santos, Alcione Miranda
dos. II. França, Ana Karina Teixeira da Cunha. III.
Título.

USO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA ESTIMAÇÃO DE GORDURA CORPORAL EM ADULTOS BRASILEIROS

Heloísa Baima da Silva Santos

Dissertação aprovada em 11 de fevereiro de 2026 pela banca examinadora constituída dos seguintes membros:

Banca Examinadora:

Prof.^a Dr.^a Ana Karina Teixeira da Cunha França
Orientadora
Universidade Federal do Maranhão

Prof.^a Dr.^a Alcione Miranda dos Santos
Coorientadora
Universidade Federal do Maranhão

Prof. Dr. Carlos Magno Sousa Júnior
Examinador Externo
Universidade Estadual do Maranhão

Prof.^a Dr.^a Joelma Ximenes Prado Teixeira Nascimento
Examinadora Interna
Universidade Federal do Maranhão

Prof.^a Dr.^a Eduarda Gomes Bogéa
Examinadora Suplente Externa
Universidade Federal do Piauí

Prof.^a Dr.^a Luz Marina Gomez Gomez
Examinadora Suplente Interna
Universidade Federal do Maranhão

Dedico este trabalho aos meus pais, com amor
e alegria.

AGRADECIMENTOS

Ao Senhor Jesus, pela graça de poder viver essa jornada que um dia foi um sonho. Pela Sua doce e real presença em cada oração feita, em cada momento de estudo e em cada desafio vivido.

Aos meus pais, Rivaldo Santos e Gisa Baima, pelas orientações, sábios conselhos, amor e orações. Que eu possa honrá-los por toda a minha vida.

À minha orientadora, professora Ana Karina, pelo tempo dedicado e pelas preciosas recomendações. Sou grata pelos ensinamentos, conselhos, cuidado e zelo com que corrigia cada parágrafo deste trabalho. Tenho por ela grande admiração e carinho. Não poderia ter melhor orientadora.

À minha coorientadora Alcione Miranda, pela idealização do trabalho e orientações valiosas para o desenvolvimento e construção metodológica deste estudo.

Às minhas amigas que dividiram a jornada acadêmica nesses quase dois anos, Emanuellen Coelho, Elza Barbosa, Patrícia Tocantins, Fernanda Lambert, Alana Gomes, Rafiza Lago e Talitha Calvet.

Ao Mateus Silva, por todo apoio e carinho, tão importantes durante a reta final do trabalho.

Aos professores Alex Barradas e Bianca Rodrigues pelas relevantes contribuições na banca de qualificação do presente trabalho.

Ao professor Bruno Feres, pela disponibilidade e solicitude.

Ao professor Giovanni França, pelas contribuições.

Ao doutorando Victor Silveira, cujas sugestões e apontamentos contribuíram para o enriquecimento do trabalho.

Ao professor Alexandre Fonseca e à Rosiane Oliveira do Centro de Ciências de Pinheiro, pelo apoio e incentivo, que foram tão relevantes nessa jornada.

À Universidade Federal do Maranhão (UFMA), ao Programa de Pós-Graduação em Saúde Coletiva (PPGSC) por promover esse ambiente de crescimento e aprendizado, e à Secretaria do PPGSC pelo trabalho dedicado.

À Coordenação do Consórcio RPS de Pelotas, em especial ao professor Dr. Fernando C. Wehrmeister, pela parceria e por gentilmente ter disponibilizado os dados para a realização desta dissertação.

A todos que de alguma forma contribuíram para a realização deste trabalho.

“Os conhecimentos que se repartem são como o andaime que ajuda a construir o edifício do amor e da sabedoria, edifício que há de durar para sempre, inclusive quando os conhecimentos forem esquecidos.”

Agostinho de Hipona

LISTA DE FIGURAS

Figura 1 - Exemplos de problemas de natureza supervisionada - classificação e regressão.....	29
Figura 2 - Fluxograma das etapas para construção de um modelo de Aprendizado de Máquina (AM) na saúde.....	32
Figura 3 - Infográfico das etapas para construção dos modelos de Aprendizado de Máquina...	40
Figura 4 - Fluxograma das etapas de consolidação da base de dados do estudo.....	42
 Artigo	
Figura 1 - Infográfico das etapas para construção dos modelos de Aprendizado de Máquina...	73
Figura 2 - Fluxograma das etapas de consolidação da base de dados do estudo.....	74
Figura 3 - Análise unificada de interpretabilidade (SHAP) do modelo SVM Radial: importância global e impacto das variáveis em uma amostra aleatória de 500 observações.....	75
Figura 4 - Gráficos de Bland-Altman para avaliação da concordância entre valores reais vs. valores estimados do percentual de gordura corporal por modelos de Aprendizado de Máquina e Regressão Linear.....	76

LISTA DE QUADROS

Quadro 1 - Características e principais resultados de artigos publicados nos últimos anos (2020-2025) que desenvolveram modelos usando Aprendizado de Máquina para estimar gordura corporal.....	34
Quadro 2 - Hiperparâmetros do algoritmo SVM Radial e suas respectivas descrições.....	44

LISTA DE TABELAS

Artigo

Tabela 1 - Características demográficas e antropométricas de amostra de indivíduos do Consórcio de Coortes de Pelotas - RS, Brasil, 1982 e 1993.....	77
Tabela 2 - Classificação do estado nutricional, de acordo com IMC e percentual de gordura corporal (%GC), de amostra de indivíduos do Consórcio de Coortes de Pelotas - RS, Brasil, 1982 e 1993.....	78
Tabela 3 - Valores das métricas da validação cruzada ($k = 5$) dos modelos de Aprendizado de Máquina e Regressão Linear aplicados no estudo.....	79
Tabela 4 - Valores das métricas do conjunto de teste dos modelos de Aprendizado de Máquina e Regressão Linear aplicados no estudo.....	80
Tabela 5 - Coeficientes de Correlação Intraclasse (CCI) entre valores reais vs. valores estimados do percentual de gordura corporal por modelos de Aprendizado de Máquina e Regressão Linear.....	81
Tabela 6 - Validação clínica do modelo SVM Radial - Reduzido 2 para estimação do percentual de gordura corporal (%GC) com base em classificação do IMC.....	82

LISTA DE ABREVIATURAS E SIGLAS

APS	- Atenção Primária em Saúde
AM	- Aprendizado de Máquina
AMG	Massa livre de gordura
BIA	- Bioimpedância
CC	- Circunferência da cintura
CCI	- Coeficiente de Correlação Intraclassas
CQ	- Circunferência do quadril
DP	- Desvio Padrão
DXA	- Absortometria de Raio-X de Dupla Energia
%GC	- Percentual de gordura corporal
IC	- Intervalo de Confiança
IDE	- Ambiente de desenvolvimento integrado (<i>Integrated development environment</i>)
IDHM	- Índice de Desenvolvimento Humano Municipal
IMC	- Índice de Massa Corporal
MAC	- Teste de Robustez (<i>Robustness Test</i>)
MAE	- Erro Absoluto Médio (<i>Mean Absolute Error</i>)
MAPE	- Erro Percentual Absoluto Médio (<i>Mean Absolute Percentage Error</i>)
NRMSE	- Raiz do Erro Quadrático Médio Normalizado (<i>Normalized Root Mean Squared Error</i>)
OMS	- Organização Mundial da Saúde
PC	- Pregas Cutâneas
PDA	- Pletismografia por Deslocamento de Ar
R ²	- Coeficiente de Determinação
RCE	- Relação Cintura-Estatura
RCQ	- Relação Cintura-Quadril
RMSE	- Raiz do Erro Quadrático Médio (<i>Root Mean Square Error</i>)
RL	- Regressão Linear
RLS	- Regressão Linear Simples
RLM	- Regressão Linear Múltipla
RS	- Rio Grande do Sul
SEE	- Erro Padrão de Estimativa (<i>Standard Error of Estimate</i>)

SHAP	- Shapley Additive ExPlanations
SVM	- Máquina de Vetor de Suporte (<i>Support Vector Machine</i>)
TAV	- Tecido adiposo visceral
TASA	Tecido adiposo subcutâneo abdominal
TCLE	- Termo de Consentimento Livre e Esclarecido
Vigitel	- Vigilância de Fatores de Risco e Proteção para Doenças Crônicas por Inquérito Telefônico

SANTOS, Heloísa Baima da Silva, **Uso de técnicas de Aprendizado de Máquina para estimação de gordura corporal em adultos brasileiros**, 2026, Dissertação (Mestrado em Saúde Coletiva) - Programa de Pós-graduação em Saúde Coletiva, Universidade Federal do Maranhão, São Luís, 112p.

RESUMO

Considerando a frequência crescente de obesidade no Brasil e no mundo, as limitações do Índice de Massa Corporal (IMC) em diferenciar massa magra de massa gorda, bem como a ausência de um método para avaliar o percentual de gordura corporal (%GC) de forma precisa e acessível, este estudo propôs-se a desenvolver modelos capazes de estimar o %GC de adultos, a partir de dados demográficos e medidas antropométricas simples, através de técnicas de Aprendizado de Máquina (AM). Todas as análises estatísticas e a construção dos modelos foram realizadas na linguagem de programação R[®]. A população do estudo foi constituída por 7.085 adultos com 22 e 30 anos, pertencentes às coortes de Pelotas-RS de 1993 e de 1982, respectivamente. Os valores do IMC dos participantes do estudo foram calculados e classificados conforme a Organização Mundial da Saúde. Os indivíduos foram categorizados, ainda, com base nos valores de %GC, em obesos ($\geq 25\%$ para homens e $\geq 32\%$ para mulheres) e não obesos ($\leq 25\%$ para homens e $\leq 32\%$ para mulheres). Após classificação do IMC e %GC, os indivíduos com $\text{IMC} < 25 \text{ kg/m}^2$ também foram agrupados em peso normal ($\text{IMC} < 25 \text{ kg/m}^2$ e ausência de obesidade pelo %GC) e obeso de peso normal ($\text{IMC} < 25 \text{ kg/m}^2$ e presença de obesidade pelo %GC). Foram consideradas variáveis de entrada: dados demográficos (sexo e idade) e medidas antropométricas (peso, altura e circunferências da cintura (CC), quadril (CQ), punho direito e panturrilha direita). Enquanto o desfecho foi considerado o %GC mensurado pelo método da Absortometria de Raio-X de Dupla Energia (DXA). Para construção dos modelos, foi aplicado o algoritmo Máquina de Vetor de Suporte (*Support Vector Machine - SVM*) Radial, adotando-se a Regressão Linear (RL) como referência (*baseline*). A partição dos dados seguiu o método *Hold-Out*, em que 80% compuseram o conjunto de treino e 20% o conjunto de teste. Exclusivamente no grupo de treino, aplicou-se a técnica de validação cruzada com k partições (*k-fold Cross Validation*), adotando-se $k = 5$. Para avaliação do desempenho, foram calculadas as métricas: Erro Médio Absoluto (MAE), Raiz do Erro Quadrático Médio (RMSE), Raiz do Erro Quadrático Médio Normalizado (NRMSE), Erro Percentual Absoluto Médio (MAPE) e Coeficiente de Determinação (R^2). Posteriormente, no modelo com melhor desempenho, aplicou-se a função SHAP para seleção das variáveis de maior relevância para

construção de duas versões reduzidas: Reduzido 1 (somente com as cinco variáveis mais relevantes com base no cálculo de SHAP - sexo, CC, CQ, altura e peso, em ordem de importância) e Reduzido 2 (incluindo a variável idade). Aplicou-se o gráfico de Bland-Altman e o Coeficiente de Correlação Intraclass (CCI) para avaliação da concordância entre o %GC estimado pelos modelos e o %GC mensurado pela DXA. A frequência de indivíduos com sobrepeso foi de 26,5% e de obesidade 14,1%, considerando o IMC. Já pelo %GC, 52,1% dos indivíduos foram classificados com obesidade. Destaca-se que, indivíduos obesos de peso normal corresponderam a 17,3% da amostra geral. Todos os modelos de AM tiveram melhores resultados quando comparados à RL, sendo o SVM Radial - Reduzido 2 o que obteve melhor desempenho e concordância ($MAE = 2,98$; $RMSE = 3,72$; $NRMSE = 7,07$; $R^2 = 0,91$, $MAPE = 13,24\%$ e $CCI = 0,95$ [IC 95%: 0,947 – 0,957]). Conclui-se que os modelos de AM desenvolvidos, utilizando variáveis demográficas e antropométricas simples, apresentaram alto desempenho e excelente concordância para estimação do %GC de adultos, sendo potencialmente aplicáveis em pesquisas epidemiológicas e em ambientes clínicos, inclusive naqueles com escassez de recursos, especialmente na Atenção Primária à Saúde. Ressalta-se que o modelo SVM Radial – Reduzido 2 mostrou-se preciso e parcimonioso, necessitando de um menor número de variáveis para sua aplicação.

Palavras-chave: composição corporal; aprendizado de máquina; algoritmos; antropometria.

SANTOS, Heloísa Baima da Silva, **Use of Machine Learning techniques for body fat estimation in Brazilian adults**, 2026, Dissertação (Mestrado em Saúde Coletiva) - Programa de Pós-graduação em Saúde Coletiva, Universidade Federal do Maranhão, São Luís, 112p.

ABSTRACT

Considering the rising prevalence of obesity in Brazil and worldwide, the limitations of Body Mass Index (BMI) in differentiating lean mass from fat mass, as well as the absence of a precise and accessible method to assess body fat percentage (%BF), this study aimed to develop models capable of estimating %BF in adults, using demographic data and simple anthropometric measures, through Machine Learning (ML) techniques. All statistical analyses and model construction were performed using the R[®] programming language. The study population consisted of 7,085 adults aged 22 and 30 years, belonging to the 1993 and 1982 Pelotas-RS birth cohorts, respectively. Participants' BMI values were calculated and classified according to the World Health Organization. Individuals were further categorized based on %BF values into obese ($\geq 25\%$ for men and $\geq 32\%$ for women) and non-obese ($\leq 25\%$ for men and $\leq 32\%$ for women). After BMI and %BF classification, individuals with BMI $< 25 \text{ kg/m}^2$ were also grouped into normal weight (BMI $< 25 \text{ kg/m}^2$ and absence of obesity by %BF) and normal-weight obese (BMI $< 25 \text{ kg/m}^2$ and presence of obesity by %BF). Input variables considered were demographic data (sex and age) and anthropometric measures (weight, height, and waist (WC), hip (HC), right wrist, and right calf circumferences). The outcome was defined as the %BF measured by Dual-Energy X-ray Absorptiometry (DXA). For model construction, the Radial Support Vector Machine (SVM) algorithm was applied, adopting Linear Regression (LR) as a baseline. Data partitioning followed the Hold-Out method, with 80% comprising the training set and 20% the test set. Exclusively in the training group, the k-fold Cross-Validation technique was applied, adopting $k = 5$. To evaluate performance, the following metrics were calculated: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Normalized Root Mean Square Error (NRMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). Subsequently, in the best-performing model, the SHAP function was applied to select the most relevant variables for constructing two reduced versions: Reduced 1 (containing only the five most relevant variables based on SHAP calculation – sex, WC, HC, height, and weight, in order of importance) and Reduced 2 (including the variable age). The Bland-Altman plot and Intraclass Correlation Coefficient (ICC) were applied to evaluate the agreement between the %BF estimated by the models and the %BF measured by DXA. The frequency of overweight was 26.5% and obesity 14.1%, considering BMI. However, by %BF,

52.1% of individuals were classified as obese. Notably, normal-weight obese individuals corresponded to 17.3% of the general sample. All ML models yielded better results compared to LR, with the Radial SVM - Reduced 2 achieving the best performance and agreement (MAE = 2.98; RMSE = 3.72; NRMSE = 7.07; $R^2 = 0.91$, MAPE = 13.24% and ICC = 0.95 [95% CI: 0.947 – 0.957]). It is concluded that the developed ML models, using simple demographic and anthropometric variables, presented high performance and excellent agreement for estimating %BF in adults, being potentially applicable in epidemiological research and clinical settings, including those with scarce resources, especially in Primary Health Care. It is noteworthy that the Radial SVM – Reduced 2 model proved to be accurate and parsimonious, requiring a smaller number of variables for its application.

Keywords: body composition; machine learning; algorithms; anthropometry.

SUMÁRIO

1.	INTRODUÇÃO.....	17
2.	OBJETIVOS	19
2.1	Geral.....	19
2.2	Específicos	19
3.	REFERENCIAL TEÓRICO	20
3.1	Obesidade	20
3.2	Métodos para Avaliação Nutricional	21
3.3	Modelos para Estimacão de Percentual de Gordura Corporal (%GC)	27
3.4	Aprendizado de Máquina (AM) Aplicado à Saúde	28
3.5	Uso de Aprendizado de Máquina para Estimar o %GC	32
4.	MATERIAL E MÉTODOS	38
4.1	Delineamento do Estudo	38
4.2	Local de Estudo.....	38
4.3	População de Estudo	38
4.4	Crterios de Inclusão	38
4.5	Crterios de Não Inclusão.....	38
4.6	Coleta de Dados	39
4.6.1	Dados demográficos	39
4.6.2	Avaliacão antropométrica e da composicão corporal	39
4.7	Processamento e Análise dos Dados.....	40
4.7.1	Preparacão dos dados e implementacão de algoritmos de AM	41
4.7.2	Avaliacão dos modelos desenvolvidos	44
4.8	Consideracões Éticas	46
5.	RESULTADOS	47
5.1	Artigo	47
6.	CONSIDERACões FINAIS	83
	REFERÊNCIAS	85
	ANEXO A – NORMAS DA REVISTA	96