



OPEN ACCESS

EDITED BY

Roshan Joy Martis,
Manipal Institute of Technology Bengaluru,
India

REVIEWED BY

Gilbert Yong San Lim,
SingHealth, Singapore
Khashayar Misaghian,
Sage-Sentinel Smart Solutions, Japan
Ramkrishna Mondal,
All India Institute of Medical Sciences (Patna),
India

*CORRESPONDENCE

Dionísio D. A. Carvalho
✉ dionisio.carvalho@navi.ifrn.edu.br

RECEIVED 02 December 2024

ACCEPTED 07 May 2025

PUBLISHED 22 May 2025

CITATION

Carvalho DDA, Costa da Silva MMR,
Albuquerque GA, Cruz AS, Fernandes F,
Barbalho IMP, Santos JPQ, Morais AHF,
Coutinho KD, Campos ALPS, Gil P, Teixeira C,
Henriques J, Machado GM and Valentim RAM
(2025) A labeled dataset for osteoporosis
screening based on electromagnetic
attenuation.
Front. Digit. Health 7:1538477.
doi: 10.3389/fdgth.2025.1538477

COPYRIGHT

© 2025 Carvalho, Costa da Silva, Albuquerque,
Cruz, Fernandes, Barbalho, Santos, Morais,
Coutinho, Campos, Gil, Teixeira, Henriques,
Machado and Valentim. This is an open-access
article distributed under the terms of the
Creative Commons Attribution License (CC
BY). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A labeled dataset for osteoporosis screening based on electromagnetic attenuation

Dionísio D. A. Carvalho^{1*}, Mikael M. R. Costa da Silva¹,
Gabriela A. Albuquerque¹, Agnaldo S. Cruz¹, Felipe Fernandes^{1,2},
Ingridy M. P. Barbalho^{1,2}, João Paulo Q. Santos^{1,3},
Antonio H. F. Morais^{1,3}, Karilany D. Coutinho¹,
Antonio L. P. S. Campos⁴, Paulo Gil^{5,6}, César Teixeira⁶,
Jorge Henriques⁶, Guilherme M. Machado⁷ and
Ricardo A. M. Valentim¹

¹Laboratory for Technological Innovation in Health, Federal University of Rio Grande do Norte, Natal, RN, Brazil, ²The Health Sciences Research Unit, Nursing, Nursing School of Coimbra, Coimbra, Portugal, ³Advanced Nucleus for Technological Innovation (NAVI), Federal Institute of Rio Grande do Norte (IFRN), Natal, RN, Brazil, ⁴Electrical and Computer Engineering Graduate Program, Federal University of Rio Grande do Norte, Natal, RN, Brazil, ⁵Centre of Technology and Systems (CTS-UNINOVA), NOVA School of Science and Technology, Campus de Caparica, Caparica, Portugal, ⁶University of Coimbra, Centre for Informatics and Systems of the University of Coimbra, Coimbra, Portugal, ⁷Research Department, ECE-Engineering School, Paris, France

KEYWORDS

public health, machine learning, artificial intelligence, screening, electromagnetic waves

1 Introduction

Osteoporosis is an asymptomatic disease characterized by bone microarchitecture deterioration due to multiple risk factors, such as low bone mineral density (BMD), nutritional deficiency, a sedentary lifestyle, alcohol use, smoking, genetic factors, and the use of medications such as glucocorticoids, for example (1–4). The worldwide prevalence of osteoporosis is estimated to be 23.1% in women and 11.7% in men (5). The economic burden related to osteoporosis fractures is significant, costing approximately \$17.9 billion a year in the US alone (6).

Studies show that the early and systematic identification of people with clinical indicators of osteoporosis, combined with primary care, appropriate interventions, and the use of therapeutic drugs has reduced the risk of fractures (7, 8). Clark and colleagues (9) demonstrated in a randomized clinical trial that primary healthcare screening tools increased osteoporosis drug prescriptions by 124% and reduced the incidence of fractures in the observed group. Their results support the premise that improving preventive screening methods is essential for identifying individuals at high risk of osteoporotic fractures. Therefore, the development of low-cost digital health technologies or solutions for rapid diagnosis in primary health care is essential.

Early screening for osteoporosis is a challenging problem because the most appropriate diagnostic method, based on dual-energy X-ray absorptiometry (DXA) to measure the number of grams of mineral per square centimeter (g/cm²) (from the lumbar spine, total femur, femoral neck or middle third of the radius) (10, 11), is highly expensive. Furthermore, as this method uses ionizing radiation technology and requires large, specialized equipment, these devices are generally concentrated in major healthcare centers. Given Brazil's social inequality and geographical challenges, access to such services is often bureaucratic, slow, and, in some cases, unfeasible.

To overcome these challenges Brazilian researchers have developed a low-cost, portable device for osteoporosis screening called OSSEUS (12, 13). The device combines techniques for measuring the attenuation of electromagnetic waves passing through bone tissue, extracting patient features(risk factors), and recognizing patterns to help classify osteoporosis and provide decision support for healthcare professionals (14). A primary health care screening study using OSSEUS with a group of 505 people referred for a DXA scan found that 78.2% could start preventive care. In addition, the study also showed that 110 people (21.8%) were healthy (concerning this pathology) and did not need to be referred to specialized health centers to get a DXA screening (15).

Recognizing the importance of osteoporosis screening in primary health care and early care, such as treatment to reduce the incidence of fractures—especially in people identified by OSSEUS as having risk factors for the disease—, this study aims to provide a database of 669 people available to support future research, especially in the field of artificial intelligence and machine learning, and contribute to the development of digital health solutions in response to osteoporosis. The dataset is available at: <https://doi.org/10.5281/zenodo.14259374>.

2 Materials and methods

2.1 Study design and participants

This study consists of a data report conducted and guided by a descriptive analysis of osteoporosis and innovative technologies enabling early disease detection. The dataset comprises 669 people (575 females, 94 males), 156 from the control group, aged between 20–85 years (median age: 55; mean age: 54.5 ± 13.2), and 513 with low BMD, aged between 18–101 years (median age: 66; mean age: 65.3 ± 11.2). Table 1 shows the demographic, anthropometric, risk factor, OSSEUS, and DXA features of the population in the dataset. All participants in the dataset were volunteers who met the eligibility criteria. Participation was limited to individuals of both sexes, aged 18 or older, with a medical indication (prescription/request) for DXA. In addition to these criteria, participants were required to have intact middle finger phalanges to meet specific prerequisites for OSSEUS.

2.1.1 Ethical approval

The experimental protocol was approved by the Research Ethics Committee (CEP) of the Federal University of Rio Grande do Norte, Natal, Brazil, through a letter, under CAAE No. 39675020.0.0000.5292/2020, as well as by the Federal Institute of Rio Grande do Norte, Natal-RN, Brazil, an official letter; under CAAE No. 75015123.9.0000.0225/2023, and following the Helsinki Accords (as amended in 2004).

2.2 Procedures

2.2.1 Data collection

The data from the 669 people was collected at the Onofre Lopes University Hospital (HUOL) at the Federal University of Rio Grande

TABLE 1 General characteristics of dataset.

Features	Male (<i>n</i> = 94)		Female (<i>n</i> = 575)	
	<i>n</i>	(14,05%)	<i>n</i>	(85,95%)
Electronic health record				
Age (mean ± SD, min-max)	61 ± 15, 18–85		63 ± 12, 19–101	
Height (mean ± SD, cm)	163,1 ± 12, 9		151,8 ± 10,2	
Weight (mean ± SD, kg)	74,1 ± 15,1		67,0 ± 16,6	
Ethnicity (black)	12	12,7%	18	3,1%
Ethnicity (brown)	32	34,0%	170	29,5%
Ethnicity (white)	50	53,1%	387	67,3%
Alcohol	29	30,8%	84	14,6%
Smoking	40	42,5%	211	36,6%
Activity	42	44,6%	173	30,0%
Milk	65	69,1%	455	79,1%
Calcium	20	21,2%	217	37,7%
Vitamin D	50	53,1%	278	48,3%
Fall	22	23,4%	110	19,1%
Parents osteoporosis	23	24,4%	195	33,9%
Parents curved	8	8,5%	38	6,6%
Corticosteroids	23	24,4%	132	22,9%
Arthritis	1	1,0%	59	10,2%
Diseases	41	43,6%	260	45,2%
Menopause	0	0%	519	90,2%
Target (normal)	40	42,5%	116	20,1%
Target (LBMD)	35	37,2%	274	47,6%
Target (osteoporosis)	19	20,2%	185	32,1%
OSSEUS				
Medial length (mean ± SD, cm)	28,7 ± 2,2		27,1 ± 2,2	
Medial height (mean ± SD, cm)	14,7 ± 1,5		12,5 ± 1,1	
Medial width (mean ± SD, cm)	16,2 ± 1,5		14,1 ± 1,2	
Calibration (mean ± SD, mV)	1667,1 ± 51,5		1685,8 ± 42,6	
Attenuation (mean ± SD, mV)	1255,6 ± 109,2		1304,2 ± 110,8	
DXA				
Deviation (mean ± SD)	−1,4 ± 1,5		−2,0 ± 1,1	

do Norte (UFRN), between July 2021 and November 2023. The sampling procedure took an average of 20 min per person and consisted of three stages: (i) completion of anamnesis in the Electronic Health Record (EHR) for collection of clinical data and identification of risk factors for metabolic bone diseases; (ii) performing anthropometry with a pachymeter on the medial phalanx of the middle finger of the non-dominant hand, followed by measurement of radiofrequency signals using OSSEUS; (iii) acquisition of standard deviations calculated by DXA (GE Lunar DPX Pro) at the sites the scan was performed (10).

Figure 1 illustrates the architecture designed for integrating the three distinct data sources into a centralized database. Each person in those data sources was identified with a unique hash code, ensuring data integrity when building the consolidated database. The database includes 29 features, which have been duly anonymized. A subset of features from a subgroup of individuals in the database was used as predictor variables in a preliminary machine learning study to predict referrals for bone densitometry tests (15).

2.2.2 Data preprocessing

The data were inspected for instances with missing values or values outside the normal range (outliers). Instances with missing

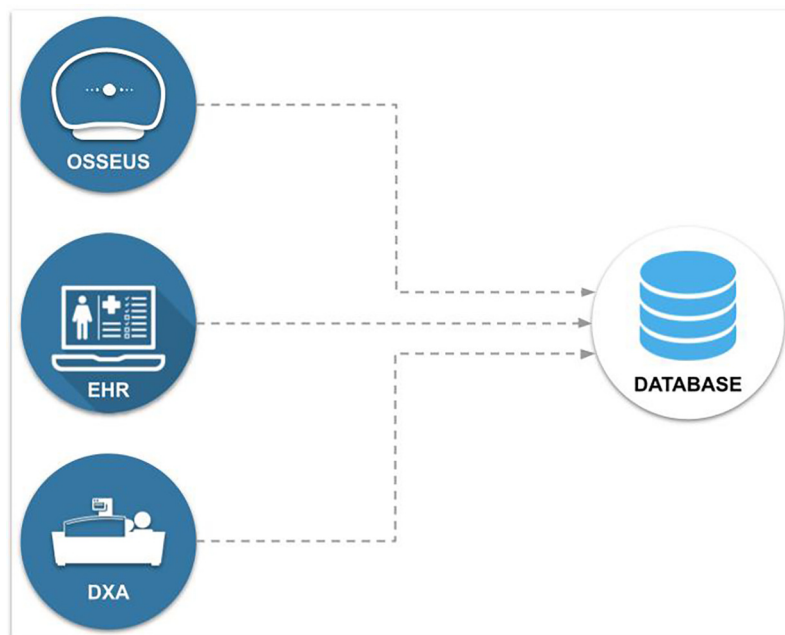


FIGURE 1
Architecture for integrating data into a database.

values for attributes related to age and the DXA scans were removed. Next, the data generated by the OSSEUS device was inspected and the instances with attenuation values exceeding those obtained during calibration were removed—since results falling outside the calibrated range may indicate measurement errors or malfunctioning of the device, compromising the validity and quality of the readings.

A new attribute, named 'worst deviation', was defined to store the lowest value among the four possible sites DXA can be performed. This attribute, along with age, gender, and menopause attributes, is considered in calculations for a diagnostic definition of metabolic bone diseases, as recommended by the Brazilian Ministry of Health in its Clinical Protocol and Therapeutic Guidelines for Osteoporosis (PCDT) (10). In summary, based on the calculation of clinical variables, menopausal women and individuals aged 50 or older are diagnosed using the T-score deviation. The remaining registries use the Z-score for diagnosis (16). Deviation values below -2.0 in the Z-score or below -1.0 in the T-score are classified as "low BMD" or "osteoporosis" (17).

The dataset also enables feature engineering, i.e., the definition, creation, or aggregation of predictive attributes. For instance, it is possible to calculate the Body Mass Index (BMI) from Electronic Health Record (EHR) data, considering body weight (kg), height (m), and the following equation:

$$\text{BMI} = \frac{\text{weight}}{\text{height}^2}.$$

In addition, it is possible to conduct investigative research using OSSEUS data, considering the real or percentage

differences between the calibration and attenuation attributes of the biomedical device and the influence of obstacles in this context. The obstacle can be determined by estimating the area or volume of the medial phalanx of the middle finger, using the following equations:

$$\text{area} = 2\pi r(r + h); \text{ volume} = \pi r^2 h,$$

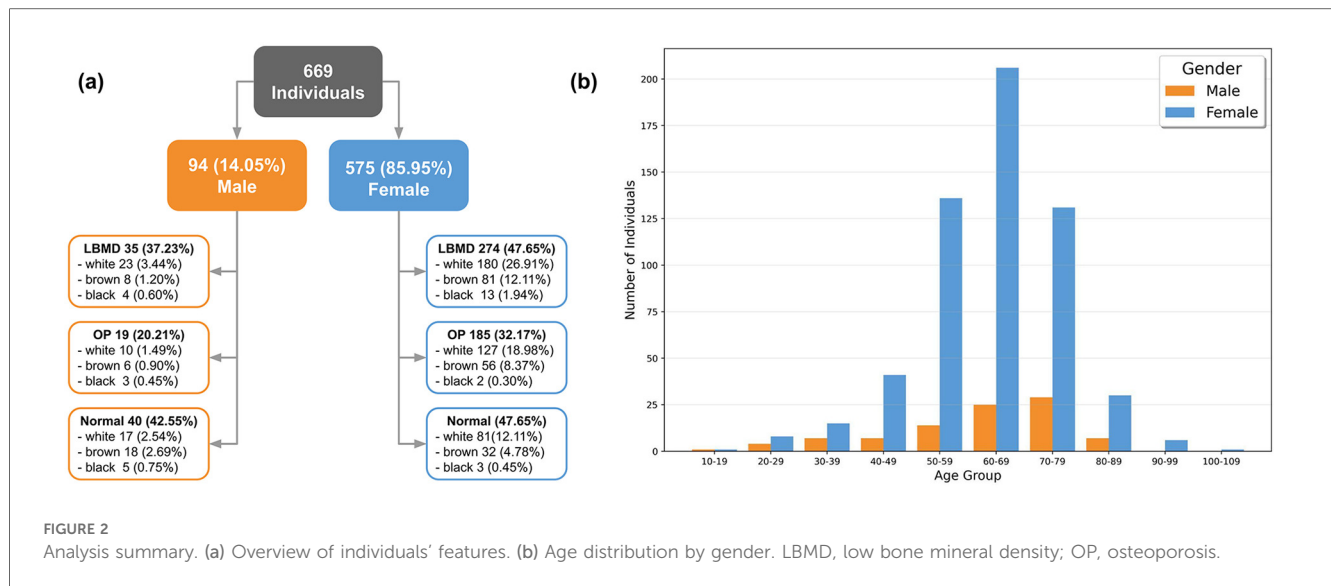
where:

- $d = \frac{\text{medial_height} + \text{medial_width}}{2}$, determines the diameter of the medial phalanx of the middle finger;
- $r = \frac{d}{2}$, defines the medial phalanx radius of the middle finger;
- $h = \text{medial_length}$, represents the length of the medial phalanx of the middle finger.

3 Descriptive analysis

The dataset presents relevant characteristics of a particular region of Brazil, which can significantly contribute to training artificial intelligence algorithms, especially machine learning. The DXA screening reports certified by specialized health professionals indicated that 156 (23.32%) individuals had normal BMD, 309 (46.19%) had low BMD, and 204 (30.49%) had osteoporosis. In the sample, 609 (91.10%) were eligible for analysis based on the T-score and 60 (8.90%) on the Z-score, as recommended by the Clinical Protocol and Therapeutic Guidelines for Osteoporosis (PCDT) (10).

Moreover, based on gender features, 575 (85.95%) participants were females, and 94 (14.05%) were males. Figure 2a breaks down the most relevant features of both groups, considering the disease-



related outcome. **Figure 2b** displays the age range of the studied group and shows a higher concentration of people in the 50–79 age bracket, totaling 541 (80.87%) individuals.

One of the groups recommended for DXA screening includes women over 65 and men aged 70 or older (10). Such group comprises 312 (46.6%) individuals with an average age of 72.8 years (ranging from 65 to 101 years), including 33 (10.5%) with normal BMD, 149 (47.7%) with low BMD, and 130 (41.6%) with osteoporosis.

Data stratification revealed that 252 people (37.67%) reported smoking, 215 (32.14%) were physically active, and 218 (32.59%) had osteoporotic parents. In addition to these characteristics, the analysis revealed a group of 302 (45.14%) individuals with underlying comorbidities and a second group of 132 (19.73%) who had previously experienced falls. Considering only the female group, which accounts for 575 participants, the data show that 519 (90.26%) reported being menopausal. In the male group, there were no reports of testosterone use. Additional characteristics can also be noted in the dataset, such as milk intake (77.7% reported consuming it) and supplementation with calcium (35.58%) and vitamin D (49.18%).

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: The dataset analyzed for this study can be found in the Zenodo: <https://doi.org/10.5281/zenodo.14259374>.

Ethics statement

The studies involving humans were approved by The experimental protocol was approved by the Research Ethics

Committee (CEP) of the Federal University of Rio Grande do Norte, Natal, Brazil, through a letter, under CAAE No. 39675020.0.0000.5292/2020, as well as by the Federal Institute of Rio Grande do Norte, Natal-RN, Brazil, an official letter; under CAAE No. 75015123.9.0000.0225/2023, and following the Helsinki Accords (as amended in 2004). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

DC: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft. MC: Data curation, Resources, Writing – review & editing. GA: Data curation, Resources, Writing – review & editing. AC: Conceptualization, Writing – review & editing. FF: Data curation, Formal analysis, Methodology, Resources, Validation, Visualization, Writing – original draft. IB: Visualization, Writing – review & editing. JS: Conceptualization, Validation, Writing – review & editing. AM: Writing – review & editing. KC: Writing – review & editing. AC: Writing – review & editing. PG: Investigation, Writing – review & editing. CT: Investigation, Writing – review & editing. JH: Investigation, Writing – review & editing. GM: Investigation, Writing – review & editing. RV: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the Brazilian Ministry of Health.

Acknowledgments

The authors express their gratitude to the Laboratory for Technological Innovation in Health (LAIS) at the Federal University of Rio Grande do Norte (UFRN) for their support. In addition, the authors would like to acknowledge the meticulous contributions of Natalia Araújo do N. Batista in translating this data report into English.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Albuquerque G, Cruz A, Carvalho D, Mayrink N, Pinheiro B, Campos A, et al. A method based on non-ionizing microwave radiation for ancillary diagnosis of osteoporosis: a pilot study. *Biomed Eng Online*. (2022) 21:70. doi: 10.1186/s12938-022-01038-y
- Compston JE, McClung MR, Leslie WD. Osteoporosis. *Lancet*. (2019) 393:364–76. doi: 10.1016/S0140-6736(18)32112-3
- Cruz AS, Lins HC, Medeiros RVA, Filho JMF, da Silva SG. Artificial intelligence on the identification of risk groups for osteoporosis, a general review. *Biomed Eng Online*. (2018) 17:12. doi: 10.1186/s12938-018-0436-1
- Pouresmaeli F, Kamalidehghan B, Kamarehei M, Goh Y. A comprehensive overview on osteoporosis and its risk factors. *Ther Clin Risk Manag*. (2018) 14:2029–49. doi: 10.2147/TCRM.S138000
- Salari N, Ghasemi H, Mohammadi L, Behzadi MH, Rabieenia E, Shohaimi S, et al. The global prevalence of osteoporosis in the world: a comprehensive systematic review and meta-analysis. *J Orthop Surg Res*. (2021) 16:609. doi: 10.1186/s13018-021-02772-0
- Clynes MA, Harvey NC, Curtis EM, Fuggle NR, Dennison EM, Cooper C. The epidemiology of osteoporosis. *Br Med Bull*. (2020) 133:105–17. doi: 10.1093/bmb/ldaa005
- Chotiyarnwong P, McCloskey EV, Harvey NC, Lorentzon M, Prieto-Alhambra D, Abrahamson B, et al. Is it time to consider population screening for fracture risk in postmenopausal women? a position paper from the international osteoporosis foundation epidemiology/quality of life working group. *Arch Osteoporos*. (2022) 17:87. doi: 10.1007/s11657-022-01117-6
- Curtis EM, Dennison EM, Cooper C, Harvey NC. Osteoporosis in 2022: care gaps to screening and personalised medicine. *Best Pract Res Clin Rheumatol*. (2022) 36:101754. doi: 10.1016/j.berh.2022.101754
- Clark EM, Gould V, Morrison L, Ades A, Dieppe P, Tobias JH. Randomized controlled trial of a primary care-based screening program to identify older women with prevalent osteoporotic vertebral fractures: cohort for skeletal health in bristol and avon (coshiba). *J Bone Miner Res*. (2012) 27:664–71. doi: 10.1002/jbmr.1478
- Brasil. Protocolo clínico e diretrizes terapêuticas da osteoporose (2023). Ministério da Saúde. Secretaria de Atenção Especializada à Saúde. Secretaria de Ciência Tecnologia, Inovação e Complexo da Saúde. Available at: <https://www.gov.br/saude/pt-br/assuntos/pcdt/arquivos/2023/portaria-conjunta-no-19-pcdt-osteoporose> (Accessed February 29, 2024).
- Kanis JA. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a who report. *Osteoporos Int*. (1994) 4:368–81. doi: 10.1007/BF01622200
- Cruz AS. *Osseus: método baseado em inteligência artificial e ondas eletromagnéticas para o diagnóstico auxiliar de doenças osteometabólicas* (Ph.D. thesis). Universidade Federal do Rio Grande do Norte, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica e de Computação, Natal, Rio Grande do Norte, Brazil (2018). p. 121f.
- Cruz AS, da Silva SG, de Castro BH. Bone density measurement through electromagnetic waves. In: *The 6th 2013 Biomedical Engineering International Conference*. (2013). p. 1–5. doi: 10.1109/BMEiCon.2013.6687655
- Pinheiro BM, Campos ALPS, de Carvalho DDA, Cruz AS, de Medeiros Valentim RA, Veras NVR, et al. The influence of antenna gain and beamwidth used in osseus in the screening process for osteoporosis. *Sci Rep*. (2021) 11:19148. doi: 10.1038/s41598-021-98204-4
- Albuquerque GA, Carvalho DDA, Cruz AS, Santos JPQ, Machado GM, Gendriz IS, et al. Osteoporosis screening using machine learning and electromagnetic waves. *Sci Rep*. (2023) 13:12865. doi: 10.1038/s41598-023-40104-w
- Brandão CMA, Camargos BM, Zerbini CA, Plapler PG, de Carvalho Mendonça LM, Albergaria B-H, et al. Scielo – Brasil – posições oficiais 2008 (2009). Posições Oficiais 2008 da Sociedade Brasileira de Densitometria Clínica (SBDens). Available at: <https://www.scielo.br/j/abem/a/jMQwSSZVxQmLXqYxt4zjNkv> (Accessed April 10, 2024).
- Netto OS, de Oliveira Lima Coutinho L, de Souza DC. Análise da nova classificação de laudos de densitometria óssea. *Radiol Bras*. (2007) 40(1):23–5. doi: 10.1590/S0100-39842007000100007

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.