Clinical Radiology 83 (2025) 106831

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Automated vertebral compression fracture detection and quantification on opportunistic CT scans: a performance evaluation

D. Guenoun [a,b], M.S. Quemeneur [a], A. Ayobi [c,*], C. Castineira [c], S. Quenet [c], J. Kiewsky [c], M. Mahfoud [c], C. Avare [c], Y. Chaibi [c], P. Champsaur [a]

[a] Department of Radiology, Institute for Locomotion, Sainte-Marguerite Hospital, APHM, 13009 Marseille, France
[b] Institute of Movement Sciences (ISM), CNRS, Aix Marseille University, 13005 Marseille, France
[c] Avicenna.AI, 375 Avenue Du Mistral, 13600 La Ciotat, France

ARTICLE INFORMATION

Article history:
Received 4 July 2024
Received in revised form
14 January 2025
Accepted 21 January 2025

AIM: Since the majority of vertebral compression fractures (VCFs) are asymptomatic, they often go undetected on opportunistic CT scans. To reduce rates of undiagnosed osteoporosis, we developed a deep learning (DL)-based algorithm using 2D/3D U-Nets convolutional neural networks to opportunistically screen for VCF on CT scans. This study aimed to evaluate the performance of the algorithm using external real-world data.

MATERIALS AND METHODS: CT scans acquired for various indications other than a suspicion of VCF from January 2019 to August 2020 were retrospectively and consecutively collected. The algorithm was designed to label each vertebra, detect VCF, measure vertebral height loss (VHL) and calculate mean Hounsfield Units (mean HU) for vertebral bone attenuation. For the ground truth, two board-certified radiologists defined if VCF was present and performed the measurements. The algorithm analyzed the scans and the results were compared to the experts' assessments.

RESULTS: A total of 100 patients (mean age: 76.6 years ± 10.1[SD], 72% women) were evaluated. The overall labeling agreement was 94.9% (95%CI: 93.7%–95.9%). Regarding VHL, the 95% limits of agreement (LoA) between the algorithm and the radiologists was [-9.3, 8.6]; 94.1% of the differences lay within the radiologists' LoA and the intraclass correlation coefficient was 0.854 (95%CI: 0.822–0.881). For the mean HU, Pearson's correlation was 0.89 (95%CI: 0.84 –0.92; p-value <0.0001). Finally, the algorithm's VCF screening sensitivity and specificity were 92.3% (95%CI: 81.5%–97.9%) and 91.7% (95%CI: 80.0%–97.7%), respectively.

CONCLUSIONS: This automated tool for screening and quantification of opportunistic VCF demonstrated high reliability and performance that may facilitate radiologists' task and improve opportunistic osteoporosis assessments.

* Guarantor and correspondent: A. Ayobi, Avicenna.AI, 375 Avenue du Mistral, 13600, La Ciotat, France.
E-mail address: angela.ayobi@avicenna.ai (A. Ayobi).

https://doi.org/10.1016/j.crad.2025.106831

Introduction

Vertebral Compression Fractures (VCF) of the thoracolumbar spine are commonly due to osteoporosis, but can also result from trauma, infection or neoplasm.[1] Genant's classification defines VCF as mild (vertebral height loss (VHL) between 20-25%, grade 1), moderate (VHL between 25-40%, grade 2), and severe (VHL $>$ 40%, grade 3).[2] Vertebrae with VHL $\leq$ 20% are normal (grade 0).

VCF prevalence among Europeans over 50 years old ranges between 18-26%.[3] In the United States, approximately 1–1.5 million VCFs occur annually.[4] VCFs are the most common osteoporotic fractures after hip fractures.[5] Women are considered to be at a higher risk with 25% of postmenopausal women experiencing VCF during their lifetime.[6]

Despite their frequency, only 15–30% of VCFs are symptomatic and only 30% receive medical attention. Early signs are often undetected, leading to underdiagnosis and undertreatment.[3,7,8] Less than 10% of fractures visible on CT images are mentioned in the radiology report[9,10] resulting in fewer patients receiving osteoporosis management plans to prevent secondary fractures.[11] Undetected VCFs significantly decrease quality of life, double age-adjusted mortality risk, and triple the risk of subsequent fractures compared to the general population.[3]

Osteoporosis can be detected on CT scans acquired for other medical indications by assessing vertebral fracture and bone attenuation.[3,6,7] This opportunistic analysis provides a fast and reliable alternative to additional dual x-ray absorptiometry (DXA) exams.[12] To reduce underdiagnosis, the International Osteoporosis Foundation and the European Society of Musculoskeletal Radiology encourage radiologists to report osteoporotic fractures on opportunistic CT scans.[13–16]

Automated deep learning (DL)-based tools could assist physicians in identifying patients with VCF on CT scans, aiding in early detection, correct diagnosis, and reporting of osteoporotic fractures.[17] Algorithms for screening VCF on CT scans have shown sensitivity and specificity ranging from 65% to 99%.[17–22] Previous studies have analyzed algorithms that provide selected components of the VCF diagnostic pipeline, such as vertebral labeling or quantitative VCF measures including VHL or mean Hounsfield Units (mean HU) vertebral bone attenuation.[23–30] However, no study has simultaneously evaluated an algorithm, which provides all four components of the VCF diagnostic pipeline to assist physicians in osteoporosis assessment. Our study aimed to evaluate a DL-based prototype algorithm designed to label each vertebra, calculate VHL, detect positive VCF cases and provide the mean HU. The software was previously trained and validated internally, thus, this study aimed to validate the software on real-world opportunistic CT scans acquired as part of clinical routine from an external source. The objective was to demonstrate that automated software can evaluate four components of the VCF diagnostic pipeline and may thereby improve the early detection of osteoporosis.

Materials and Methods

Study design

This study was conducted with approval from our local institutional medical ethics committee and received no funding. It was a single-center, retrospective, observational and cross-sectional diagnostic study with data collected as part of clinical routine. It aims to compare the algorithm results against the ground truth based on four objectives:

1. Labeling of each visible vertebra from T1-L5,
2. Quantifying of the percentage of VHL,
3. Detecting of cases with grade 2 or 3 VCF,
4. Measuring of the mean HU vertebral bone attenuation within the first grade 0 or 1 vertebrae among L1-L4, starting with L1.

Vertebral labeling consisted of identifying and naming (e.g. "T3") each thoracic and lumbar visible vertebra in every case. The percentage of VHL, a measurement used to evaluate the degree of a VCF,[2] was computed for each visible vertebra from T1-L5. The VHL were obtained based on the reduction in the anterior, middle, and posterior vertebral heights ($\boldsymbol{h}_a$, $\boldsymbol{h}_m$ and $\boldsymbol{h}_p$- see Fig 1), following Genant's quantitative methodology.[2]

In order to analyze the vertebral compression detection, a per-case-level analysis was conducted by considering one case as positive if at least one vertebra presented a grade 2 or 3 VCF (moderate or severe deformity).[2] The mean HU was calculated by placing a region of interest on the trabecular bone at the first grade 0 or 1 vertebra in the range of L1-L4, starting with L1. Indeed, if L1 has VHL $>$ 25%, the measurements are performed in L2, and so on.

Data collection

Chest-abdominal-pelvis (CAP) CT scans, acquired as part of clinical routine from January 2019 to August 2020 for medical indications other than suspicion of VCF (cancer,

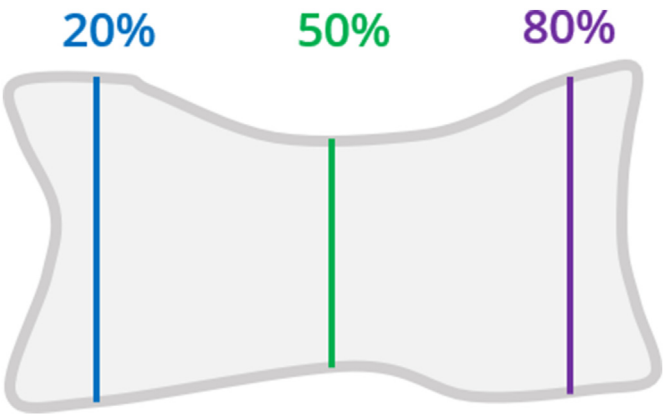

Figure 1 Example of the three vertebral height measurements placed at the anterior (blue), middle (green) and posterior (purple) aspects of the vertebral body.

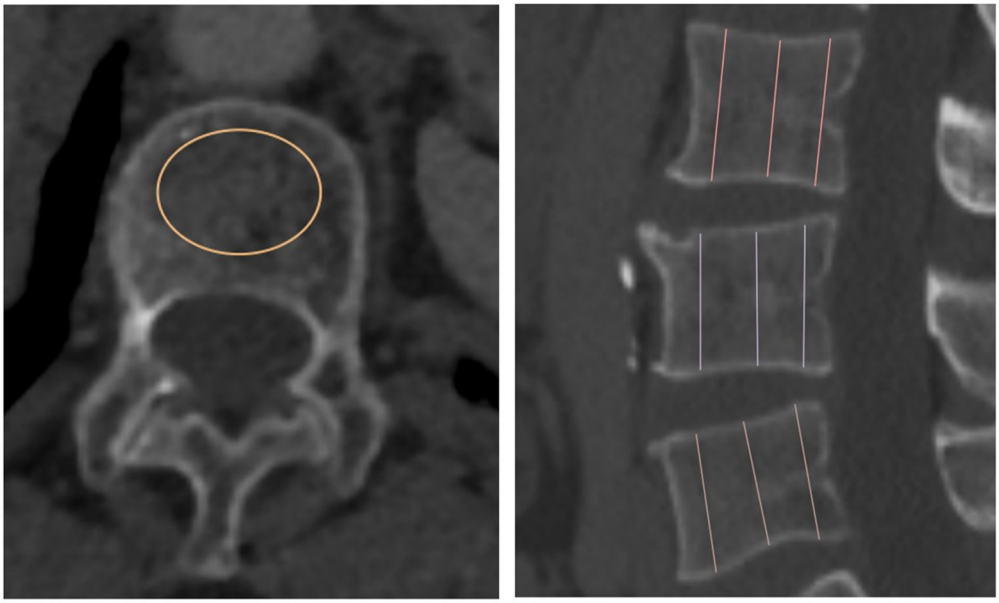

Figure 2 (a) Example of placement of region of interest (ROI) for mean HU assessment in the first lumbar vertebra (L1). Axial image of L1 in bone window (C: 300; W: 1600). A ROI is placed in the upper part of L1, between the endplate and the entrance of vessels at the midportion. The ROI was as large as possible and positioned in a homogenous area without inclusion of the cortex. (b) Example of the 3 heights placed between the anterior and posterior vertebrae cortical part at 20%, 50% and 80% of the vertebral body. Midsagittal vertebrae image in bone window (C: 300; W: 1600).

HIV, neoplasia, infection, etc.) were retrospectively and consecutively collected from our hospital. CT scans both with and without contrast were accepted. All patients were scanned on a 64-row scanner (LightSpeed VCT, GE Medical Systems) in helical mode with a 0.625 mm slice collimation, tube voltage of 120 kVp, and image matrix of 512 $\times$ 512. A soft kernel was used and images were reconstructed at a 1.25 mm slice thickness. For contrast CTs, iohexol (Omnipaque, GE Healthcare AS) with an amount of 90 mL was used as a contrast agent.

All patients over 50 years old were eligible. Non-inclusion criteria were material in the vertebrae or significant artifacts. All eligible CT scans were transferred from the hospital's Picture Archiving and Communication System (PACS) and the first 100 cases that met the described criteria were consecutively included in the study.

Ground truth

In order to establish the ground truth (GT), two board-certified radiologists with 19 and 7 years of experience conducted the scan analysis. They labeled by consensus all visible vertebrae from T1 to L5. For the VHL analysis, they assessed the scans employing Genant's methodology.[2] Each measurement consisted of three segments delineating the anterior, posterior, and middle vertebral heights. The segments were placed on each visible vertebra in the sagittal plane at a slice passing through the middle of the vertebral body. This slice was chosen using the MPR view to minimize bias by placing the coronal plane in the middle of the vertebral body. The segments connected the middle of the lower and upper cortical bone (Fig 2b). Deformations due to the presence of Schmorl's nodes were ignored in the VHL measurements by considering the cortical lines of the vertebral end-plates to be continuous. Both radiologists analyzed the same cases separately, blinded to the analysis of the other. The same two radiologists then jointly reviewed the same cases and defined the final VHL ground truth by consensus. Vertebrae containing cement or hardware were excluded from measurement.

Regarding VCF screening, the analysis was performed at a per-case level. The truthers defined, by consensus, a case as positive (presence of at least one vertebra with a compression fracture) after analyzing the visual appearance of the vertebra, taking into account the confounding factors (artifacts, Schmorl's nodes, Scheuermann disease, etc) and the measurements, as described by Genant's method.[2]

For the mean HU, the experts placed an elliptical region of interest (ROI) in the anterior part of the vertebral body (without including the cortical bone, nodes, veins or arthrosis signs) at an axial slice selected in the upper third of the vertebra, as they usually do in clinical practice (Fig 2a).[31] Indeed, according to the literature, the assessment of vertebral bone HU is most frequently performed on L1 for thoraco-abdominal or abdominal scans[32]; but as mean HU measurements on vertebrae with fractures are deemed unreliable, the ROI was placed on the first unaffected vertebrae (VHL $<$25% and without cement or hardware) among the lumbar segment (L1-L4).[33] One mean HU measurement per case was performed by the GT. This study did not incorporate any significant adjustments in HU measurements, based on collective research suggesting that

differences in trabecular BMD measurements between contrast-enhanced and unenhanced CT scans are generally minor and unlikely to affect osteoporosis prediction significantly.[34,35]

Deep learning-based algorithm

The DL-based prototype algorithm was developed by Avicenna.AI (CINA-VCF Quantix v0.60, La Ciotat, France). It was trained on 12,402 vertebrae from 1353 CT cases acquired in US and French centers during 2021 and 2022. Datasets reflected the variety that exists in clinical settings, with an adequate distribution in terms of CT manufacturers (more than 64 scanner models from Siemens, Philips, GE and Canon), patient age, patient gender, contrast presence, field of view and slice thickness. Thoracic and lumbar vertebrae were homogeneously represented and there was a great diversity in the VHLs. The algorithm was validated on a separate pilot dataset of 1,994 vertebrae from 152 cases and achieved a sensitivity of 92% (95% CI: 82%–97%), a specificity of 99% (95% CI: 93%–100%) and an accuracy of 96% (95% CI: 92%–99%) for VCF detection, an accuracy of 98% (95% CI: 94.3%–99.6%) for vertebral labeling, a LoA of ±10% for VHL and strong correlations (range of 0.6–0.8) for mean HU measurements.

The application was designed using a cascade of task-focused networks. They all feature architectures based on 2D and 3D U-Nets derived from Convolutional Neural Networks (CNNs), which perform very well in segmentation and landmark regression tasks.[36] The algorithm locates the spine in order to reduce data size and standardize the field of view. Each lumbar and thoracic vertebral body center is then located and labeled. A post-processing step allows vertebrae with cement or surgical material to be filtered out.The subsequent phase identifies six measurement landmarks per vertebra, which correspond to the endpoints of the anterior, middle, and posterior measurement segments of the vertebral body (Fig 1). These landmarks enable the calculation of the VHL for the corresponding vertebra. The last step uses the segmentation of each vertebra of interest to place an elliptical ROI at the axial section and calculate the mean HU. Four results are automatically generated:

- The label of all visible thoracic and lumbar vertebrae.
- The VHL of each vertebra displayed in yellow, orange and red for Genant's grade 1, 2 and 3 VCF, respectively.[2]
- A passive notification for cases with at least one vertebra with moderate or severe VCF.
- The mean HU within all healthy (grade 0 or 1 without cementoplasty or hardware) vertebrae L1-L4.

The DL-based model is property of Avicenna.AI and can be obtained upon reasonable request and with the approval of the Regulatory Affairs Department of Avicenna.AI.

Statistical analysis

The labeling evaluation was conducted at the per-vertebra level. The overall agreement percentage and 95% the Clopper Pearson confidence interval (95% CI) between the algorithm and the GT were computed. Vertebral height loss (VHL) was compared using a Bland–Altman Plot by calculating Limits of Agreement (LoA) at a per-vertebra level between the algorithm and the GT, and between each individual truther. The percentage of differences between the software and the GT that lie within the truthers' LoA was also evaluated. Moreover, the intraclass correlation coefficient was computed to assess the agreement between the algorithm and the GT. For the VCF screening, the software's sensitivity, specificity and accuracy on a per-case level were computed, together with their 95% CI.

The Pearson correlation coefficient was computed to evaluate the linear relationship between the algorithm's mean HU measurements and the GT. For each case, only the mean HU measured on the same vertebra by the GT and the software were taken into account. In fact, the algorithm provides mean HU measurements on all healthy vertebrae (grade 0 or 1 without cement or hardware) among L1-L4, but the GT measured only the first healthy vertebra among this lumbar segment. All the statistical analyses were performed using MedCalc Statistical Software (v20.015, MedCalc Software Ltd).

Prior to the analyses, an initial sample size calculation was carried out for each metric based on conventional clinicians' performance,[32,37] and assuming a significance of 5% (α-level) and a power of 80% (β-level of 20%). Regarding vertebral labeling, expecting a 95% accuracy with 95% confidence interval not wider than ±5%, requires at least 239 vertebrae.[38] For VHL measurements, an estimated expected mean of differences of -0.35, standard deviation of differences of 4.5 and a maximum allowed difference between methods of 10.5, yields a minimum of 274 required vertebrae.[39] For VCF screening, at least 73 cases are required for an expected sensitivity and specificity of 90%, a prevalence of 52% and a 95% confidence interval not wider than ±10%.[40] Finally, for the mean HU, at least 9 cases are required for a correlation coefficient of 0.8.[41]

Results

A total of 1,012 CAP CTs acquired between January 2019 and August 2020 were collected. 835 were rejected due to age < 50 years, significant artifacts and/or presence of material. From the remaining 177 cases, the first 100 were consecutively selected. Among them, 11% were non-contrast CT exams, mean age was 76.6 years (SD: 10.1) and 72% were women. After the GT assessment, the rate of positive exams (cases with at least one grade 2 or 3 VCF) was 52%. Additional baseline characteristics are presented in Table 1. From this cohort, 20 cases (340 vertebrae) were randomly selected for VHL measurements. A flowchart of the cohort selection is shown in Fig 3. There was no missing data.

Labeling of vertebrae

A final cohort of 1,700 vertebrae (100 CTs x 17 vertebrae each) was analyzed. 17/1700 (1%) of vertebrae presented

Table 1
Patient and CT scans characteristics.

Characteristic	Cases without VCF (n = 48, 816 vertebrae)	Cases with VCF (n = 52, 884 vertebrae)
Female, n (%)	35 (72.9%)	37 (71.2%)
Age (years), mean ± SD	78.13 ± 10.2	78.5 ± 9.5
Contrast CT, n (%)	42 (87.5%)	47 (90.4%)
Mean HU ± SD per vertebra[a]	133.6 ± 48.1	109.2 ± 33.9
Number of vertebrae with VCF, n (%)	0/816 (0%)	178/884 (20.1%)

[a] Mean HU within all grade 0 or 1 vertebrae among L1-L4.

cement or hardware and were thus excluded from analysis. The overall agreement between the software and the GT was 94.9% (95% CI: 93.7%–95.9%). Discrepancies were mainly caused by L5 lumbar sacralization or strong spine local curvature. An example of labeling is presented in Fig 5.

Vertebral height loss measurements

All the VHL measurements defined by the GT were compared to those of the artificial intelligence (AI) tool. A total of 340 vertebrae from 20 patients were measured by the GT but 4/340 (1.2%) were excluded from analysis due to presence of cement. The 95% LoA between truther 1 and truther 2 were [-8.1, 11.3]. Between the AI-based device and the GT the 95% LoA were [-9.3, 8.6] and 94.1% (317/337) of the differences between the AI and the GT lay within the ground truthers LoA's. Also, the mean difference between the software and the GT was -0.4, indicating a negligible systematic bias. ICC between truther 1 and truther 2 (inter-rater reliability) was 0.864 (95% CI: 0.816–0.897) and between the algorithm and the GT was 0.854 (95% CI: 0.822–0.881). Figure 4 presents the Bland–Altman Plot for the VHL measurements and Fig 5a shows an example of VHL measurements provided by the algorithm.

VCF detection

The whole dataset (100 cases, 52% positive) was included in the per-case analysis of the device's ability to detect a grade 2 or grade 3 VCF. The results for sensitivity, specificity and accuracy are presented in Table 2. Regarding the discrepancies, 3/4 of false negatives (FNs) and 3/4 of false positives (FPs) were cases close to the grade 2/grade 3 boundary, with VHL around 25% according to the software (i.e. 24.2%, 23.8%, 22.5% for FNs; and 25.5%, 27% and 27.1% for

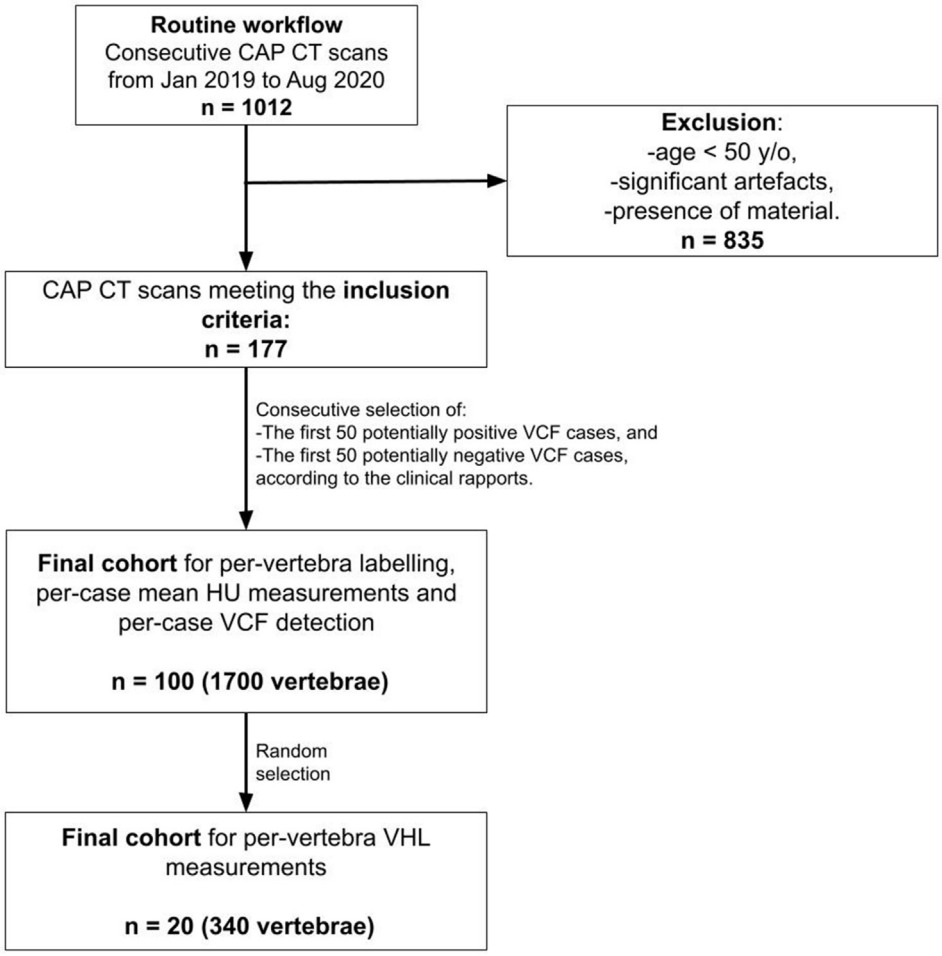

Figure 3 Flowchart of the final study cohort.

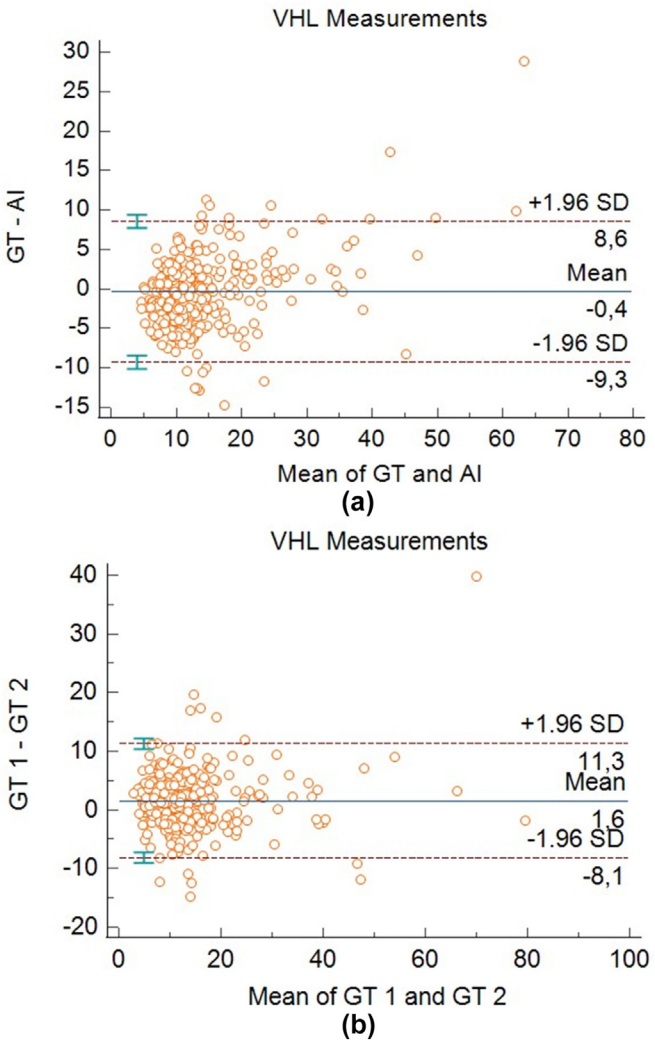

Figure 4 (a) Bland–Altman plot between AI and GT with 95% limits of agreement. (b) Bland–Altman VHL plot between both radiologists with 95% limits of agreement.

FPs) and hence, very close to the positive threshold. One example of these FNs is shown in Fig 5b. The remaining FN was due to a compression not visible in the plane passing through the midsagittal plane of the vertebral body and the remaining FP was caused by the natural deformation of L5.

Mean HU vertebral bone attenuation

The linear regression analysis of the mean HU calculated by the software and the GT was computed on a per-case level. Among the 100 cases, 7/100 (7%) cases were excluded because the mean HU was not measured on the same vertebra (e.g., the algorithm detected L1 as compressed and therefore measured the mean HU form L2 to L4, while according to the GT, L1 was not compressed so the operator measured the mean HU at L1). The Pearson correlation coefficient was 0.89 (95% CI: 0.84–0.92; p-value <0.0001), indicating a strong correlation between the algorithm and the GT. Figure 6 presents the scatter diagram of the mean HU measurements and Fig 5a presents an example of the mean HU provided by the DL tool.

Discussion

Vertebral labeling, VHL assessment, vertebral fracture detection and mean HU measurements are essential in spine processing workflow and significantly associated with osteoporotic vertebral fractures.[3] Using routine CTs for opportunistic osteoporosis screening reduces costs and radiation compared to DXA exams.[12] The Royal Osteoporosis Society prioritizes opportunistic detection of osteoporosis and VCF from CT.[16] This is the first proof of concept study evaluating the performances of four components of an automated AI software expected to improve the early detection of osteoporosis on CT scans.

The software demonstrated proficiency in vertebral labeling by correctly naming 94.9% of vertebrae and optimizing a time-consuming clinical task. Our results align with the state of the art which presents accuracies ranging from 84.62% to 97.5% for similar algorithms.[23–28]

Regarding the VHL, the software's measurements closely matched human variabilities, with 94.1% of differences between the AI device and the GT falling within the range of the truthers' disagreements. Clinicians' 95% LoA for conventional assessments range from $\pm$5.97 to $\pm$11.75 for intra-rater agreement and $\pm$7.25 to $\pm$12.53 for inter-rater agreement.[37] FDA-cleared devices reported 87.5%–96.06% of differences lie within the truthers' 95% LoA.[27,29] Moreover, the device reported an ICC of 0.85, surpassing the range of 0.53–0.82 typically reported for conventional clinicians' assessments.[37]

VHL measurement differences should be considered relative to the VHL percentage. A 15% difference between the GT and software for a vertebra with 10% VHL is more critical than for one with 40% VHL. In the first situation the case will fall into the wrong category (positive rather than negative) and the software will alert the radiologist for a vertebra that is not compressed, which is considered a critical situation. In the second case, even with a 25% error in measurement, the degree of vertebral compression remains significant, justifying the alert. In our study, we obtained lower discrepancies in vertebrae with VHL $<$ 15% and most of the greatest differences were observed in vertebrae with VHL $>$ 30% (highly compressed). Indeed, as the vertebra becomes smaller, even a minor error can have a notable impact on the predicted VHL. For example, for one case, the GT detected a VHL of 78% and the AI a VHL of 49%, yielding a difference of 29%. Even if the measurement was not accurate, the software was able to flag a case requiring patient care.

The software demonstrated robust performance at a per-case level in opportunistic VCF detection, classifying positive and negative cases with a sensitivity of 92.3% and a specificity of 91.7%. Misclassifications primarily occurred in borderline cases that were even difficult to assess by the radiologists. These results are consistent with the literature; other AI algorithms obtained sensitivities ranging from

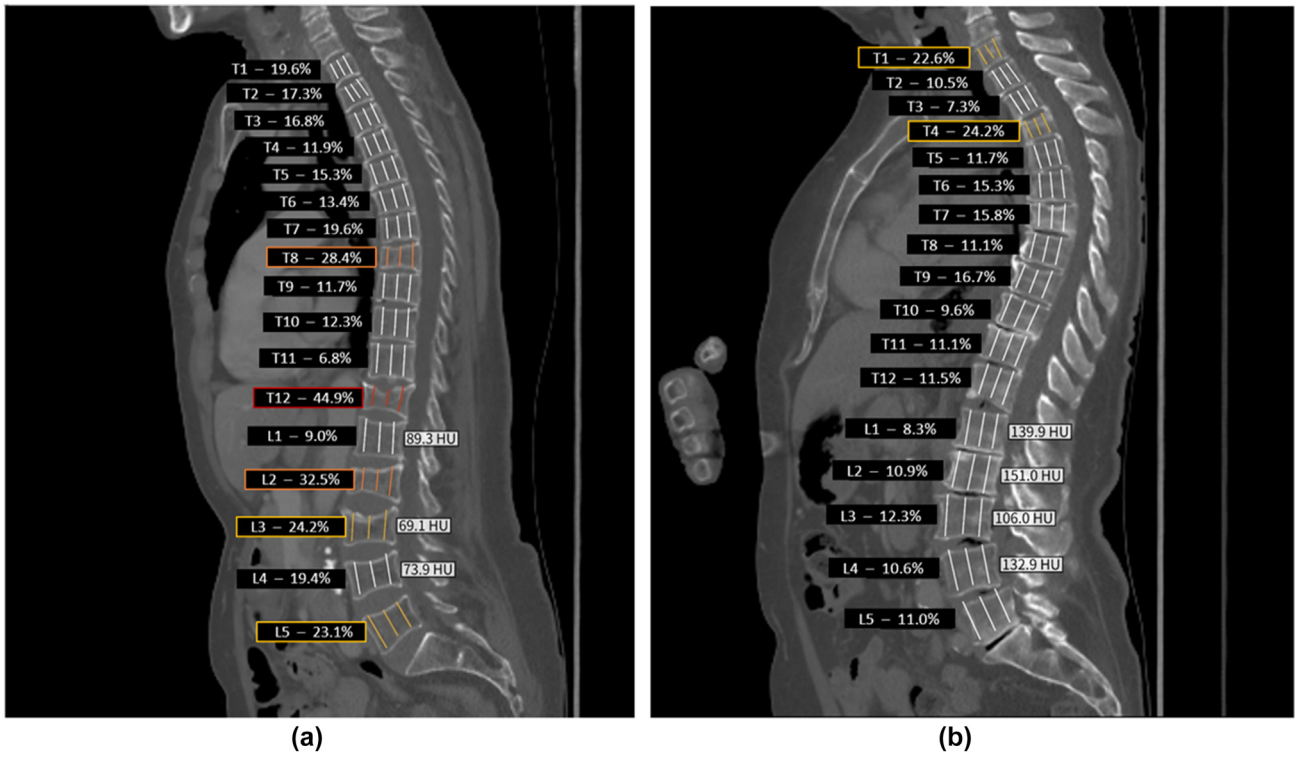

Figure 5 Example of the software results. (a) For this case, T8, L2 and L4 were considered as grade 2 compressions (VHL of 26.9%, 35.1% and 27.7%, respectively), and T12 as grade 3 compressions (VHL of 49.1%), according to the ground truth. The other vertebrae were considered as grade 0 or 1 compressions. The algorithm identified T8 and L2 as grade 2 (orange color) and T12 as grade 3 (red color). Regarding the mean HU, the ground truth measured 85 HU for L1. All the vertebrae were correctly labeled by the AI. (b) Example of a false negative case. The truthers detected a VHL > 25% at T4 while the algorithm detected a VHL of 24.2%, very close to the threshold limit.

Table 2
AI per-case VCF detection performance (TP = true positives, FN = false negatives, FP = false positives, TN = true negatives).

Statistic	Per-case analysis (n = 100)
Sensitivity (95% CI)	92.3% (81.5%–97.9%)
	TP = 48; FN = 4
Specificity (95% CI)	91.7% (80.0%–97.7%)
	TN = 44; FP = 4
Accuracy (95% CI)	92.0% (84.8%–96.5%)

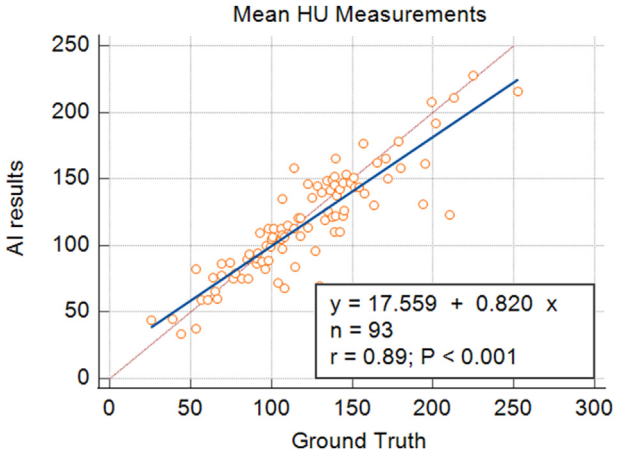

Figure 6 Mean HU linear regression relationship between the ground truth and the AI.

65%-98.7% and specificities ranging from 65%-95.8%, with misdiagnosis occurring when vertebrae were slightly deformed.[17–22] Similarly, radiologists in clinical routine presented accuracies ranging between 82%-94%, and most misreported cases were due to a lack of information regarding patient's history of osteoporosis or relative deformations[18,37] This DL tool provided reliable results comparable to previous findings for VCF screening. It accurately alerts physicians about opportunistic VCF, enabling early diagnosis and patient management.

A Pearson correlation coefficient of 0.89 between the AI mean HU measurements and the GT indicated a high degree of linear relationship. A similar software obtained a Pearson correlation coefficient of 0.8,[30] which is in line with our results. In clinical routine, mean HU measurements are predictive of future fragility fractures and correlate with DXA-based bone mineral density (BMD) analyzes and osteoporosis prediction.[7,12,42] Radiologists' assessments for mean HU are considered reliable with intraclass correlation coefficients ranging from 0.70 to 0.94.[32,43,44] In opportunistic cases, these measurements are typically not conducted. Therefore, a rapid, precise, and automated mean HU measurement could streamline radiologists' tasks and enhance BMD analyses.

This study has limitations. First, the retrospective selection of patients over 50 with CAP CT may cause selection

bias. Second, the small cohort limits rare confounding factors and reduces statistical power. Third, being a single-center, single-scanner study may affect generalizability. The algorithm may exhibit a different performance when applied to images from other scanners due to variations in image quality, noise characteristics, and scanning protocols. However, since it was trained in more than 64 scanner models from 4 different scanner manufacturers, the variations are likely minimal. Further research with multiple scanners and institutions is needed to assure a valuable transferability. A prospective study would also be valuable to assess real-time data and the impact of opportunistic VCF detection on patient management and outcomes. Additionally, therapeutic management should involve general practitioners and rheumatologists.

In conclusion, these results highlight the AI-based tool's potential to expedite accurate assessments, particularly in opportunistic cases where such measurements are not commonly performed. These functionalities provide a reliable and accurate system that may assist clinicians' decision-support systems, reduce misdiagnosed rates and improve patient care.

Author contribution

1. Guarantor of integrity of the entire study: A. Ayobi
2. Study concepts and design: A. Ayobi, D. Guenoun, C. Castineira, P. Champsaur
3. Literature research; Manuscript preparation: A. Ayobi, C. Castineira
4. Clinical studies: D. Guenoun, M.S. Quemeneur, P. Champsaur
5. Experimental studies/data analysis: S. Quenet, J. Kiewsky, M. Mahfoud, C. Avare
6. Statistical analysis: A. Ayobi, Y. Chaibi, C. Avare
7. Manuscript editing: D. Guenoun, A. Ayobi, M.S. Quemeneur, P. Champsaur, S. Quenet, J. Kiewsky, M. Mahfoud, C. Avare, C. Castineira, Y. Chaibi

Conflict of interest

The authors declare the following financial interests/personal relationships that may be considered as potential competing interests: Angela Ayobi, Charlotte Castineira, Sarah Quenet, Julie Kiewsky, Mohammed Mahfoud, Christophe Avare and Yasmina Chaibi report financial support was provided by Avicenna. AI through employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Acknowledgments

The authors thank the clinicians from Institute for Locomotion at Sainte-Marguerite Hospital, who performed the preliminary data collection process. They also thank the patients who provided anonymously the CT data for this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crad.2025.106831.

References

1. Alexandru D, So W. Evaluation and management of vertebral compression fractures. *Perm J* 2012;**16**:46–51. https://doi.org/10.7812/TPP/12-037.
2. Genant HK, Wu CY, van Kuijk C, *et al*. Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res* 1993;**8**:1137–48. https://doi.org/10.1002/jbmr.5650080915.
3. Löffler MT, Jacob A, Scharr A, *et al*. Automatic opportunistic osteoporosis screening in routine CT: improved prediction of patients with prevalent vertebral fractures compared to DXA. *Eur Radiol* 2021;**31**:6069–77. https://doi.org/10.1007/s00330-020-07655-2.
4. Donnally C, DiPompeo C, Varacallo M. *Vertebral compression fractures. StatPearls internet*. Treasure Island (FL): StatPearls Publishing; 2023.
5. Ensrud KE. Epidemiology of fracture risk with advancing age. *J Gerontol A Biol Sci Med Sci* 2013;**68**:1236–42. https://doi.org/10.1093/gerona/glt092.
6. Alsoof D, Anderson G, McDonald CL, *et al*. Diagnosis and management of vertebral compression fracture. *Am J Med* 2022;**135**:815–21. https://doi.org/10.1016/j.amjmed.2022.02.035.
7. Pickhardt PJ, Pooler BD, Lauder T, *et al*. Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications. *Ann Intern Med* 2013;**158**:588. https://doi.org/10.7326/0003-4819-158-8-201304160-00003.
8. State of the Nation report: Vertebral fracture identification in 2021 2021. https://strwebprdmedia.blob.core.windows.net/media/fcodhs1m/state-of-the-nation-report-vertebral-fracture-identification-in-2021.pdf (accessed January 15, 2023).
9. Carberry GA, Pooler BD, Binkley N, *et al*. Unreported vertebral body compression fractures at abdominal multidetector CT. *Radiology* 2013;**268**:120–6. https://doi.org/10.1148/radiol.13121632.
10. Bartalena T. Incidental vertebral compression fractures in imaging studies: lessons not learned by radiologists. *World J Radiol* 2010;**2**:399. https://doi.org/10.4329/wjr.v2.i10.399.
11. Miller PD. Underdiagnoses and undertreatment of osteoporosis: the battle to Be won. *J Clin Endocrinol Metab* 2016;**101**:852–9. https://doi.org/10.1210/jc.2015-3156.
12. Lenchik L, Weaver AA, Ward RJ, *et al*. Opportunistic screening for osteoporosis using computed tomography: state of the art and argument for paradigm shift. *Curr Rheumatol Rep* 2018;**20**:74. https://doi.org/10.1007/s11926-018-0784-7.
13. Gruber M, Dinges J, Müller D, *et al*. Impact of specific training in detecting osteoporotic vertebral fractures on routine chest radiographs. *RöFo - Fortschritte Auf Dem Geb Röntgenstrahlen Bildgeb Verfahr* 2013;**185**:1074–80. https://doi.org/10.1055/s-0033-1335230.
14. Gold LS, Cody RF, Tan WK, *et al*. Osteoporosis identification among previously undiagnosed individuals with vertebral fractures. *Osteoporos Int* 2022;**33**:1925–35. https://doi.org/10.1007/s00198-022-06450-7.
15. Baum T, Bauer JS, Klinder T, *et al*. Automatic detection of osteoporotic vertebral fractures in routine thoracic and abdominal MDCT. *Eur Radiol* 2014;**24**:872–80. https://doi.org/10.1007/s00330-013-3089-2.
16. Aggarwal V, Maslen C, Abel RL, *et al*. Opportunistic diagnosis of osteoporosis, fragile bone strength and vertebral fractures from routine CT scans; a review of approved technology systems and pathways to implementation. *Ther Adv Musculoskelet Dis* 2021;**13**:1759720X2110240, https://doi.org/10.1177/1759720X211024029.
17. Roux C, Rozes A, Reizine D, *et al*. Fully automated opportunistic screening of vertebral fractures and osteoporosis on more than 150 000 routine computed tomography scans. *Rheumatology* 2022;**61**:3269–78. https://doi.org/10.1093/rheumatology/keab878.

18. Tomita N, Cheung YY, Hassanpour S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. *Comput Biol Med* 2018;**98**:8–15. https://doi.org/10.1016/j.compbiomed.2018.05.011.
19. Burns JE, Yao J, Summers RM. Vertebral body compression fractures and bone density: automated detection and classification on CT images. *Radiology* 2017;**284**:788–97. https://doi.org/10.1148/radiol.2017162100.
20. Kolanu N, Silverstone EJ, Ho BH, *et al*. Clinical utility of computer-aided diagnosis of vertebral fractures from computed tomography images. *J Bone Miner Res* 2020;**35**:2307–12. https://doi.org/10.1002/jbmr.4146.
21. K192901. 510(k) premarket notification. https://www.accessdata.fda.gov/cdrh_docs/pdf19/K192901.pdf. [Accessed 15 December 2023]; 2020.
22. Page JH, Moser FG, Maya MM, *et al*. Opportunistic CT screening—machine learning algorithm identifies majority of vertebral compression fractures: a cohort study. *JBMR Plus* 2023;**7**:e10778, https://doi.org/10.1002/jbm4.10778.
23. Sekuboyina A, Husseini ME, Bayat A, *et al*. VerSe: a vertebrae labelling and segmentation benchmark for multi-detector CT images. *Med Image Anal* 2021;**73**:102166, https://doi.org/10.1016/j.media.2021.102166.
24. Netherton TJ, Rhee DJ, Cardenas CE, *et al*. Evaluation of a multiview architecture for automatic vertebral labeling of palliative radiotherapy simulation CT images. *Med Phys* 2020;**47**:5592–608. https://doi.org/10.1002/mp.14415.
25. Payer C, Štern D, Bischof H, *et al*. Coarse to fine vertebrae localization and segmentation with SpatialConfiguration-Net and U-Net. In: *Proc. 15th Int. Jt. Conf. Comput. Vis. Imaging Comput. Graph. Theory Appl. Valletta*. Malta: SCITEPRESS - Science and Technology Publications; 2020. p. 124–33. https://doi.org/10.5220/0008975201240133.
26. Chen D, Bai Y, Zhao W, *et al. Deep reasoning networks for unsupervised pattern de-mixing with constraint reasoning*. 2020.
27. K213944. *510(k) premarket notification*. 2022.
28. Pan Y, Shi D, Wang H, *et al*. Automatic opportunistic osteoporosis screening using low-dose chest computed tomography scans obtained for lung cancer screening. *Eur Radiol* 2020;**30**:4107–16. https://doi.org/10.1007/s00330-020-06679-y.
29. K193267. *510(k) Premarket Notification.pdf*. 2020.
30. Pickhardt PJ, Lee SJ, Liu J, *et al*. Population-based opportunistic osteoporosis screening: validation of a fully automated CT tool for assessing longitudinal BMD changes. *Br J Radiol* 2019;**92**:20180726, https://doi.org/10.1259/bjr.20180726.
31. Guenoun D, Champsaur P. Scanner opportuniste pour le diagnostic d'ostéo- porose et la détection de fracture n.d.
32. Pompe E, De Jong PA, De Jong WU, *et al*. Inter-observer and inter-examination variability of manual vertebral bone attenuation measurements on computed tomography. *Eur Radiol* 2016;**26**:3046–53. https://doi.org/10.1007/s00330-015-4145-x.
33. Boutin RD, Lenchik L. Value-added opportunistic CT: insights into osteoporosis and sarcopenia. *Am J Roentgenol* 2020;**215**:582–94. https://doi.org/10.2214/AJR.20.22874.
34. Jang S, Graffy PM, Ziemlewicz TJ, *et al*. Opportunistic osteoporosis screening at routine abdominal and thoracic CT: normative L1 trabecular attenuation values in more than 20 000 adults. *Radiology* 2019;**291**:360–7. https://doi.org/10.1148/radiol.2019181648.
35. Pickhardt PJ, Lauder T, Pooler BD, *et al*. Effect of IV contrast on lumbar trabecular attenuation at routine abdominal CT: correlation with DXA and implications for opportunistic osteoporosis screening. *Osteoporos Int* 2016;**27**:147–52. https://doi.org/10.1007/s00198-015-3224-9.
36. Ronneberger O, Fischer P, Brox T. *U-net: convolutional networks for biomedical image segmentation*. 2015.
37. Buckens CF, De Jong PA, Mol C, *et al*. Intra and interobserver reliability and agreement of semiquantitative vertebral fracture assessment on chest computed tomography. *PLoS ONE* 2013;**8**:e71204, https://doi.org/10.1371/journal.pone.0071204.
38. Machin D, Campbell MJ, Tan SB, Tan SH. Sample Size Tables for Clinical Studies, 3rd Edition | Wiley. WileyCom n.d. https://www.wiley.com/en-dk/Sample+Size+Tables+for+Clinical+Studies%2C+3rd+Edition-p-9781444300710 (accessed October 22, 2024).
39. Lu M-J, Zhong W-H, Liu Y-X, *et al*. Sample size for assessing agreement between two methods of measurement by Bland–Altman method. *Int J Biostat* 2016;**12**, https://doi.org/10.1515/ijb-2015-0039.
40. Buderer NMF, Statistical Methodology I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emerg Med* 1996;**3**:895–900. https://doi.org/10.1111/j.1553-2712.1996.tb03538.x.
41. Martin Bland Harms M. An introduction to medical statistics (3rd ed.) *Physiotherapy* 2010;**96**:82. https://doi.org/10.1016/j.physio.2008.05.001.
42. Lee SJ, Graffy PM, Zea RD, *et al*. Future osteoporotic fracture risk related to lumbar vertebral trabecular attenuation measured at routine body CT: vertebral trabecular attenuation values and osteoporotic fracture risk. *J Bone Miner Res* 2018;**33**:860–7. https://doi.org/10.1002/jbmr.3383.
43. Hendrickson NR, Pickhardt PJ, Rosas HG, Anderson PA. Bone mineral density T-scores derived from CT attenuation numbers (hounsfield units): clinical utility and correlation with dual-energy x-ray absorptiometry n.d.
44. Lee SY, Kwon S-S, Kim HS, *et al*. Reliability and validity of lower extremity computed tomography as a screening tool for osteoporosis. *Osteoporos Int* 2015;**26**:1387–94. https://doi.org/10.1007/s00198-014-3013-x.