COMPUTED TOMOGRAPHY

Deep learning-driven incidental detection of vertebral fractures in cancer patients: advancing diagnostic precision and clinical management

El Mehdi Mniai[1,7] · Vladimir Laletin[2] · Lambros Tselikas[3,4] · Tarek Assi[6] · Baptiste Bonnet[3,4] · Astrid Orfali Camez[1] · Amir Zemmouri[1] · Serge Muller[1] · Tania Moussa[1] · Yasmina Chaibi[2] · Julie Kiewsky[2] · Sarah Quenet[2] · Christophe Avare[2] · Nathalie Lassau[1,5] · Corinne Balleyguier[1,5] · Angela Ayobi[2] · Samy Ammari[1,5]

Abstract

Purpose Vertebral compression fractures (VCFs) are the most prevalent skeletal manifestations of osteoporosis in cancer patients. Yet, they are frequently missed or not reported in routine clinical radiology, adversely impacting patient outcomes and quality of life. This study evaluates the diagnostic performance of a deep-learning (DL)-based application and its potential to reduce the miss rate of incidental VCFs in a high-risk cancer population.

Materials and methods We retrospectively analysed thoraco-abdomino-pelvic (TAP) CT scans from 1556 patients with stage IV cancer collected consecutively over a 4-month period (September–December 2023) in a tertiary cancer center. A DL-based application flagged cases positive for VCFs, which were subsequently reviewed by two expert radiologists for validation. Additionally, grade 3 fractures identified by the application were independently assessed by two expert interventional radiologists to determine their eligibility for vertebroplasty.

Results Of the 1556 cases, 501 were flagged as positive for VCF by the application, with 436 confirmed as true positives by expert review, yielding a positive predictive value (PPV) of 87%. Common causes of false positives included sclerotic vertebral metastases, scoliosis, and vertebrae misidentification. Notably, 83.5% (364/436) of true positive VCFs were absent from radiology reports, indicating a substantial non-report rate in routine practice. Ten grade 3 fractures were overlooked or not reported by radiologists. Among them, 9 were deemed suitable for vertebroplasty by expert interventional radiologists.

Conclusion This study underscores the potential of DL-based applications to improve the detection of VCFs. The analyzed tool can assist radiologists in detecting more incidental vertebral fractures in adult cancer patients, optimising timely treatment and reducing associated morbidity and economic burden. Moreover, it might enhance patient access to interventional treatments such as vertebroplasty. These findings highlight the transformative role that DL can play in optimising clinical management and outcomes for osteoporosis-related VCFs in cancer patients.

Keywords Vertebral compression fracture · VCF · Deep-learning · Cancer · AI · Radiology

---

El Mehdi Mniai, Vladimir Laletin, Angela Ayobi and Samy Ammari have contributed equally to this work.

✉ Vladimir Laletin
vladimir.laletin@avicenna.ai

1 Department of Radiology, Gustave Roussy, 94805 Villejuif, France

2 Avicenna.AI, 375 Avenue du Mistral, 13600 La Ciotat, France

3 Département d'Anesthésie, Chirurgie et Interventionnel (DACI), Service de Radiologie Interventionnelle, Gustave Roussy, 94805 Villejuif, France

4 Centre d'Investigation Clinique BIOTHERIS, INSERM CIC 1428, 94805 Villejuif, France

5 Laboratoire d'Imagerie Biomédicale Multimodale Paris-Saclay, Inserm, CNRS, CEA, BIOMAPS, UMR 1281, Université Paris-Saclay, 94800 Villejuif, France

6 Division of International Patients Care, Gustave Roussy, 94805 Villejuif, France

7 Mohammed VI University of Sciences and Health – UM6SS, Casablanca, Morocco

Introduction

Vertebral fractures are the most common type of fracture linked to osteoporosis, accounting for up to 50% of all fractures in individuals with this condition [1–3]. These fractures predominantly involve vertebral compression fractures (VCFs) of the thoracic and lumbar vertebral bodies [4, 5]. The overall prevalence of vertebral fractures worldwide was estimated at approximately 20% [6]. Despite this high prevalence, VCFs frequently remain asymptomatic and clinically undetected, resulting in underdiagnosis and suboptimal patient management [7–9]. Undiagnosed VCFs are associated with increased mortality, a heightened risk of subsequent fractures—particularly hip neck fractures—and a greater likelihood of immobility [1, 10–12]. Collectively, these factors contribute to poor prognoses and significantly diminished quality of life [13, 14].

Cancer patients face an elevated risk of vertebral fractures. Indeed, the spine is the most frequent site of skeletal metastases, which trigger osteoclast activation, irregular bone trabeculae formation, and bone resorption [15, 16]. Cancer therapies like chemotherapy agents and glucocorticoid treatment can reduce bone mineral mass through direct bone toxicities and indirect sex hormone lowering mechanisms [17, 18]. Additionally, sex hormone level lowering is a strategy for the treatment of hormone-dependent tumors [17, 19]. Moreover, cancer predominantly affects an aging population, which is already susceptible to bone pathology and osteoporosis due to age-related declines in sex hormone levels [19]. Radiotherapy and radionuclide treatments, as well as cancer immunotherapy, favor bone mineral loss and bone fractures [17, 20–22]. Hence, testing for VCF and osteoporosis might significantly improve cancer patient management and outcomes.

Computer tomography (CT) is a widespread examination in cancer patients and was demonstrated to be of high usability for opportunistic osteoporosis and VCF fracture screening [23, 24]. However, the implementation of opportunistic pathology detection may be slowed by the excessive demand placed on radiologists. With the advancement of deep learning (DL) solutions in radiology, this task can be at least partially automated, reducing the clinical workload and enhancing radiologists' efficiency [25, 26]. Deep learning applications in radiology are undergoing significant advancements, demonstrating their potential as complementary tools to augment radiologists' performance, reduce involvement in repetitive tasks, and mitigate professional fatigue [27–29]. Multiple tools for automated osteoporosis and VCF screening have been proposed recently [30–40]. In this study, we aimed to assess the diagnostic performance of an automated deep learning tool

for VCF detection and quantification (CINA-VCF Quantix, Avicenna.AI, La Ciotat, France) in an asymptomatic oncology population undergoing CT imaging as part of their regular oncological care. Furthermore, we examined the potential impact on clinical management identifying patients with VCF that would benefit from vertebroplasty and that were not reported by radiologists during standard evaluations.

Materials and methods

Ethical considerations

This study was approved by the Institutional Review Board of Gustave Roussy Cancer Campus (no.: 2024-381). The need for written informed consent was waived. Avicenna.AI provided the VCF detection DL-based algorithm for this study. The study received no financial support.

Data collection and study design

All consecutive thoraco-abdomino-pelvic (TAP) CT scans performed from September to December 2023 at the Institute Gustave Roussy (Villejuif, France) for patients with stage IV cancer were retrospectively collected. CT scans were acquired on GE Healthcare GE Optima CT660 (GE Healthcare, Milwaukee, WI, USA) and Siemens Healthineers SOMATOM® Force Dual Source CT system (Siemens Healthineers, Erlangen, Germany). Demographic data of included patients were collected, including age, gender, cancer type, bone metastatic status, and spinal treatment status.

All collected cases were analyzed by CINA-VCF Quantix DL-based application version 0.7 (Avicenna.AI, La Ciotat, France) for VCF detection. The cases flagged positive for VCF (grade 1, grade 2, and grade 3, corresponding to—20–25%, 26–40%, and > 40% of vertebral height loss, respectively, according to Genants' classification [41]) were reviewed by two expert radiologists (one senior and one junior) in order to validate the VCF findings. The DL-based application performance was retrospectively compared to the radiological assessment, based on the available reports.

All grade 3 fractures detected by DL application were independently analyzed by two expert interventional radiologists, with 10 and 2 years of experience respectively, to assess whether they would have treated these severe vertebral fractures with vertebroplasty or not.

Finally, as previously reported, to evaluate osteoporosis risk, anterior trabecular CT attenuation values of the L1 vertebra in mean Hounsfield units (HU) were evaluated [42]. Using a region-of-interest (ROI) approach, mean HU values were automatically computed by the DL application for all

scans in which the L1 vertebra was visible in the absence of sclerotic metastases and fractures.

Deep learning algorithm and training methodology

The DL-based algorithm, developed by Avicenna.AI (CINA-VCF Quantix v0.7, La Ciotat, France), was developed and trained as previously reported [43]. In brief, the algorithm was built on 2D/3D U-Net-based CNN architectures. It identifies and standardizes the spine, detects and labels thoracic and lumbar vertebrae, and excludes vertebrae with cement or other materials. The model was trained on 12,402 vertebrae collected from 886 cases sourced from both U.S. and French centers between 2021 and 2022. The datasets encompassed a broad clinical diversity, ensuring balanced representation across scanner manufacturers, patient age groups, genders, contrast types, fields of view, and slice thicknesses. Specifically, the vertebral height loss (VHL) algorithm was trained on a representative subset of the training dataset comprising 325 cases, representing 3576 individually annotated vertebrae. Each vertebral level between T1 and L5 was represented, with L1 the lowest frequency (100 occurrences) and T11–T12 the highest (198 occurrences), reflecting the targeted fields of view coverage. The annotation process was the same for the training and pilot validation datasets, resulting for each vertebra in six coplanar points per vertebra corresponding to the endpoints of the vertebral body's anterior, middle, and posterior measurement segments. The model was then trained to reproduce the landmarks positioned on a 2D patch centered around the vertebral body in the sagittal plane defined by the measures. The algorithm quantifies intra and intervertebral height loss relative to neighboring vertebrae, calculate VHL and derives the Genant grade from it. As a final step, an ad-hoc algorithm defines an elliptical region-of-interest (ROI) placement on the mid-vertebra body level, in the axial plane, for mean HU measurement. Outputs include vertebral labels, VHL grades (Genant's 1–3 grade fractures), and mean HU values for L1–L4 or T8–T11 if the lumbar spine is not in the field of view.

To minimize the biases in the model training, the selection process specifically targeted a distribution of the grades at the vertebra level close to the general population prevalence (89.1% grade 0, 4.5% grade 1, 4.7% grade 2, and 1.7% grade 3). It included 55.7% of patients with at least a vertebra of grade > 0 and 44.3% of patients with at least one confounding factor. Due to the limited number of patients with bone metastasis, data augmentation included Hounsfield Units shifts. A stratified sampling strategy was used during the minibatch construction to compensate for the generally low prevalence of high grades.

The algorithm was evaluated on an independent pilot dataset comprising 1,994 vertebrae from 152 cases. It demonstrated a sensitivity of 92% (95% CI 82–97%), a specificity of 99% (95% CI 93–100%), and an overall accuracy of 96% (95% CI 92–99%) for detecting VCFs. Vertebral labeling reached an accuracy of 98% (95% CI 94.3–99.6%), with a 95% limit of agreement for the height loss at the vertebra level of [− 9.89%, 10.57%] and a bias of 0.34%. A strong correlation (ranging from 0.6 to 0.8) for mean Hounsfield Unit (HU) measurements was achieved.

Statistical analysis

Per-patient and per-vertebra positive predictive values (PPV) for VCF detection were calculated by comparing DL-based application results with the evaluations made by two clinical expert radiologists who evaluated only flagged as positive by DL application cases. The per-patient and per-vertebra VCF non-reported rates were evaluated by comparing the software results with clinical reports. Finally, the vertebroplasty missed rate was calculated based on the comparison of a documented clinical report and the independent evaluation of two expert interventional radiologists. All the statistical analyses were performed using MedCalc Statistical Software (v20.015, MedCalc Software Ltd., Ostend, Belgium).

Results

Study population

For this study, a total of 3055 computed tomography (CT) scans were collected over a 4-month period, spanning from September to December 2023. All cases were processed by CINA-VCF Quantix application. Of these, 1499 scans (49.1%) were excluded by the DL-based application due to non-compliance with the acquisition protocol (Fig. 1). The exclusions were attributed to patient age (as the application processes cases from patients aged 50 years and older) and incompatible slice thickness or reconstruction kernel. Among the 1556 CT scans retained in the study, 506 instances were flagged as positive for vertebral compression fracture by the application. After a detailed review, among positively flagged CT scans, five corresponded to secondary examination of the same patients with VCF during the study period, so they were excluded as duplicated data (Fig. 1). Finally, the 501 flagged as positive individual patients by CINA-VCF Quantix were retained for statistical analysis. The mean age for the 501 patients was 63.6 ± 11.5 years, and there were 231 (46.1%) of women. For the patients for whom the data on metastasis, osteoporosis, or interventional treatment of VCF were accessible, the distribution was as follows: mean bone density was 126.04 ± 46.32 HU ($n = 431$); bone metastases were present in 140 (31.46%) out of 445 patients with available data; osteoporosis was present in 217 (50.46%) out of 430 patients with accessible record. Finally,

Fig. 1 Study flowchart and software performances. From 3,055 collected CT scans, 1556 scans passed the acquisition protocol of the CINA-VCF Quantix application. 501 individual patients were flagged positive and reviewed by two expert radiologists. The number of true positive cases was established and the positive predictive value (PPV) was calculated. The data was analyzed *per scan* and *per vertebrae* level. The PPV was calculated for all fractures and for grade 1–2 and grade 3 fractures separately. Some cases included both grade 1–2 and grade 3 fractures
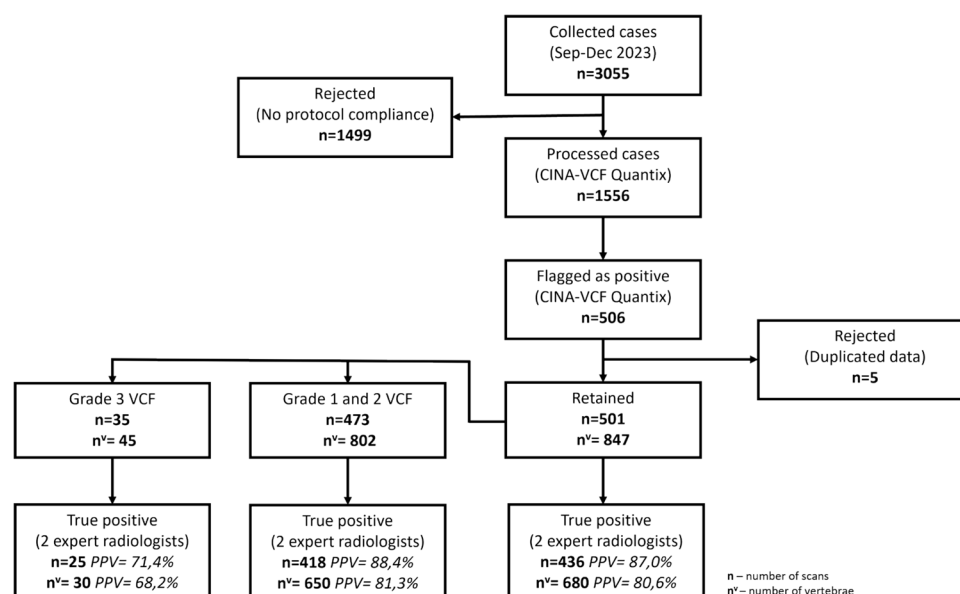
```
                        Collected cases
                        (Sep-Dec 2023)
                          n=3055

   Rejected
   (No protocol compliance)          Processed cases
   n=1499                            (CINA-VCF Quantix)
                                       n=1556

                                   Flagged as positive
                                   (CINA-VCF Quantix)      Rejected
                                       n=506              (Duplicated data)
                                                              n=5

   Grade 3 VCF      Grade 1 and 2 VCF      Retained
   n=35             n=473                   n=501
   nᵛ= 45           nᵛ= 802                 nᵛ= 847

   True positive    True positive          True positive
   (2 expert        (2 expert              (2 expert
   radiologists)    radiologists)          radiologists)
   n=25 PPV= 71,4%  n=418 PPV= 88,4%       n=436 PPV= 87,0%     n – number of scans
   nᵛ= 30 PPV=68,2% nᵛ= 650 PPV= 81,3%     nᵛ= 680 PPV= 80,6%   nᵛ – number of vertebrae
```

Table 1 Patient cohort characteristics

Characteristic	Flagged positive by CINA (VCF Quantix) Total: 501	All TP cases Total: 436	Grades 1 and 2 TP fractures Total: 418	Grade 3 TP fractures Total: 25
Mean bone density (L1) *mean HU ± SD*	126.04 ± 46.32 (*n* = 431)*	126.00 ± 46.03 (*n* = 421)	127.56 ± 45.61 (*n* = 404)	84.88 ± 40.59 (*n* = 24)
Sex *women, %*	46.10% (231/501)	48.17% (210/436)	47.61% (199/418)	52% (13/25)
Age *mean ± SD*	63.59 ± 11.49 (*n* = 501)	63.35 ± 11.56 (*n* = 436)	63.19 ± 11.61 (*n* = 418)	68.36 ± 9.57 (*n* = 25)
Bone metastases *% of patients*	31.46% (140/445)	31.87% (138/433)	31.33% (131/415)	48% (12/25)
Osteoporosis *% of patients*	50.46% (217/430)	50.48% (212/420)	49.63% (200/403)	75% (18/24)
Vertebroplasty *% of patients*	7.26% (32/441)	7.18% (31/433)	6.3% (26/414)	20% (5/25)

*In parenthesis (): number of available data

32 (7.26%) out of 441 patients with available information had previous interventional treatment of VCF. Table 1 and Fig. 1. Supplementary Fig. 1 presents particular situations in the dataset.

Diagnostic performance of the DL-based application

After a careful review of 501 flagged as positive by DL-based application cases, two expert radiologists in a consensus analysis confirmed 436 cases as true positive (TP) for VCF. Regarding primary tumor distribution for TP cases, lung, head and neck, and skin (excluding melanoma) tumors contributed to 19.2%, 18.8%, and 10% of cases, respectively. Hormone-dependent tumors were equally presented in the dataset with 8.6% of cases with prostate cancer and 7.6% of cases with breast tumors. Supplementary Fig. 2 presents additional data distribution in terms of primary tumor sites. Sixty-five cases were considered as being false positive (FP), leading to a PPV of 87.0% (95% CI 83.8–89.8%) (Fig. 1).

Representative examples of true negative (TN) and true positive (TP) notifications are presented in Fig. 2.

Regarding individual vertebrae, among the 436 true positive cases, there were 683 individual VCFs. Of these, 680 were correctly identified by the DL-based application; hence, three VCFs were missed. Additionally, the application identified 164 vertebrae as fractured; however, expert radiologists classified them as false positives, leading to a per-vertebra PPV of 80.6% (95% CI 77.7–83.2%) (Fig. 1). Examples of false negative and false positive detections are presented in Fig. 3. Regarding the VCF grade, 350 cases were considered as grade 1 (1 FN, 75 FPs, and 274 TPs), 452 cases as grade 2 (1 FN, 75 FPs, and 376 TPs), and 45 cases as grade 3 (1 FN, 14 FPs, and 30 TPs), according to the experts.

The main reasons related to false positives were: sclerotic vertebral metastases, which impede vertebral height measurement; scoliosis; and incorrect centroid detection due to the vertebrae deformation or collapsing (Fig. 3A–C). In this

Fig. 2 Examples of negative and positive cases detections. **A** Negative case. The DL-based application quantifies intra and intervertebral height loss relative to neighboring vertebrae. The threshold for positive cases is set to 20% of height loss. Variation of height loss is presented in percentage for each vertebra. Mean HU is indicated for L1–L4 vertebrae. **B** Positive case. Grade 2 vertebral fracture of T7 in an osteoporotic patient—Height loss estimated at 32.2%—Mean HU L1 trabecular attenuation: 96.7 HU (physiological reference 122–198 HU). **C** Positive case. Grade 2 vertebral fracture of L4 in a patient with lytic bone metastasis. **D** Positive case. T12 grade 3 vertebral fracture and L3, L5 grade 1 vertebral fracture identified
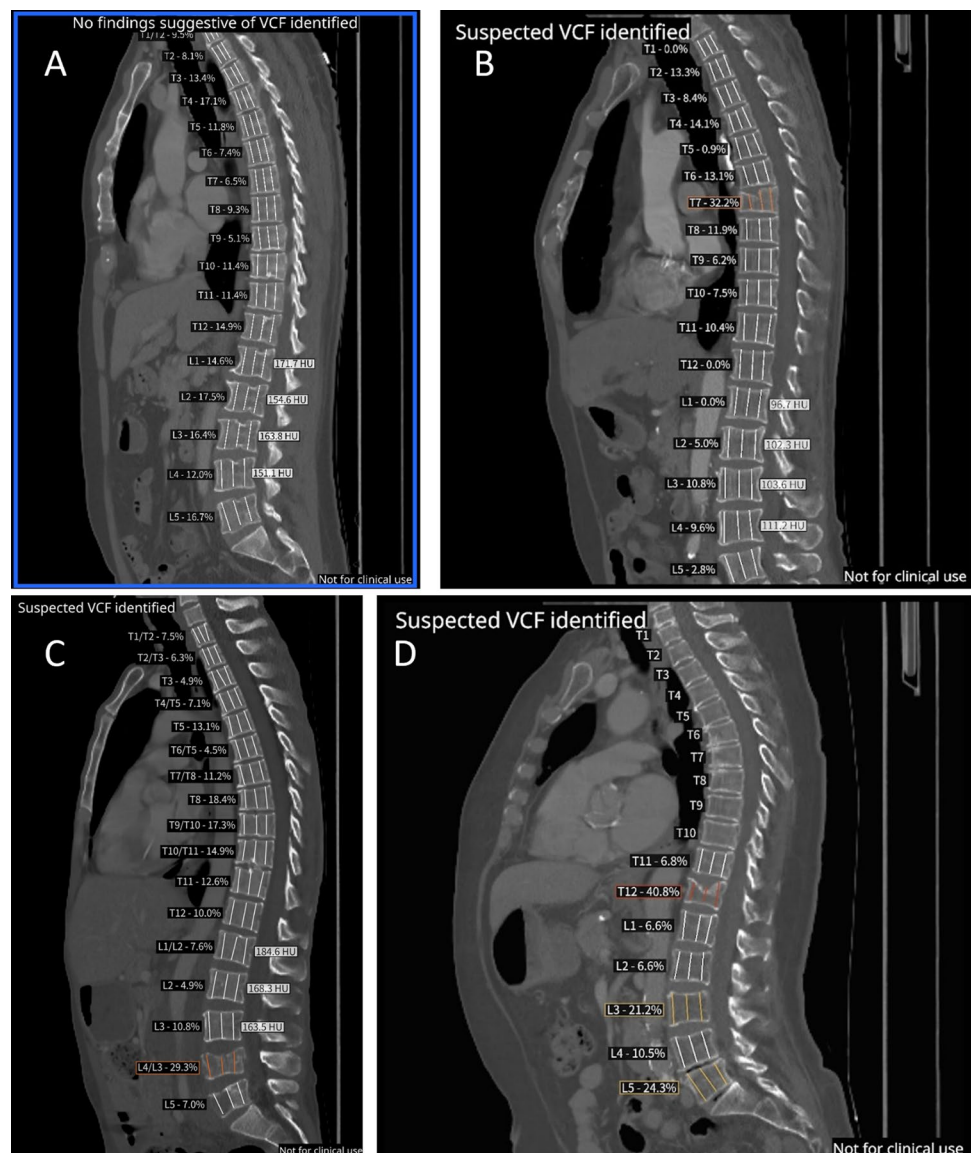
latter case, the DL-based application displays a dot in the region to alert the user that there is a high probability that the displayed height loss may not be correct (Fig. 3C).

Concerning the false negatives (Fig. 3D), two main sources of inaccuracies of the current algorithm were identified, related to vertebral body count and vertebral centroid identification. Both sources of false negatives will be addressed in the Discussion.

Detection of VCFs not reported by radiologists

Among 436 true positive cases, 364 (83.5%) were not mentioned in radiology reports. Therefore, these patients can be considered missed or unreported by radiologists in routine clinical practice (Fig. 4). Regarding per-vertebra findings, of the 680 fractures correctly detected by the DL-based application, 550 (80.9%) were not reported by radiologists (Fig. 4). Of these fractures, 540 were classified as grade 1 or grade 2, accounting for 83% of all 650 true positive grade 1 and grade 2 fractures. These corresponded to 85.7% (358 of 418) of patients. Moreover, 10 grade 3 fractures out of 30 (33.3%) were not reported by radiologists (Fig. 4). These corresponded to 36.0% (9 of 25) of patients.

Regarding the fracture localization, among 650 grade 1 and 2 fractures, 317 (48.8%) thoracic and 315 (48.5%) lumbar vertebrae were fractured. No difference in application performance, neither for thoracic nor for lumbar vertebrae, has been noticed. Among 202 vertebral fractures of patients with grade 1 and 2 fractures presenting vertebral metastases, 41 (20.3%) fractures were related to the metastatic process. Moreover, regarding grade 3 VCFs, 22 (73.3%) out of 30 fractures were localized in the thoracic and 8 (26.6%) in the
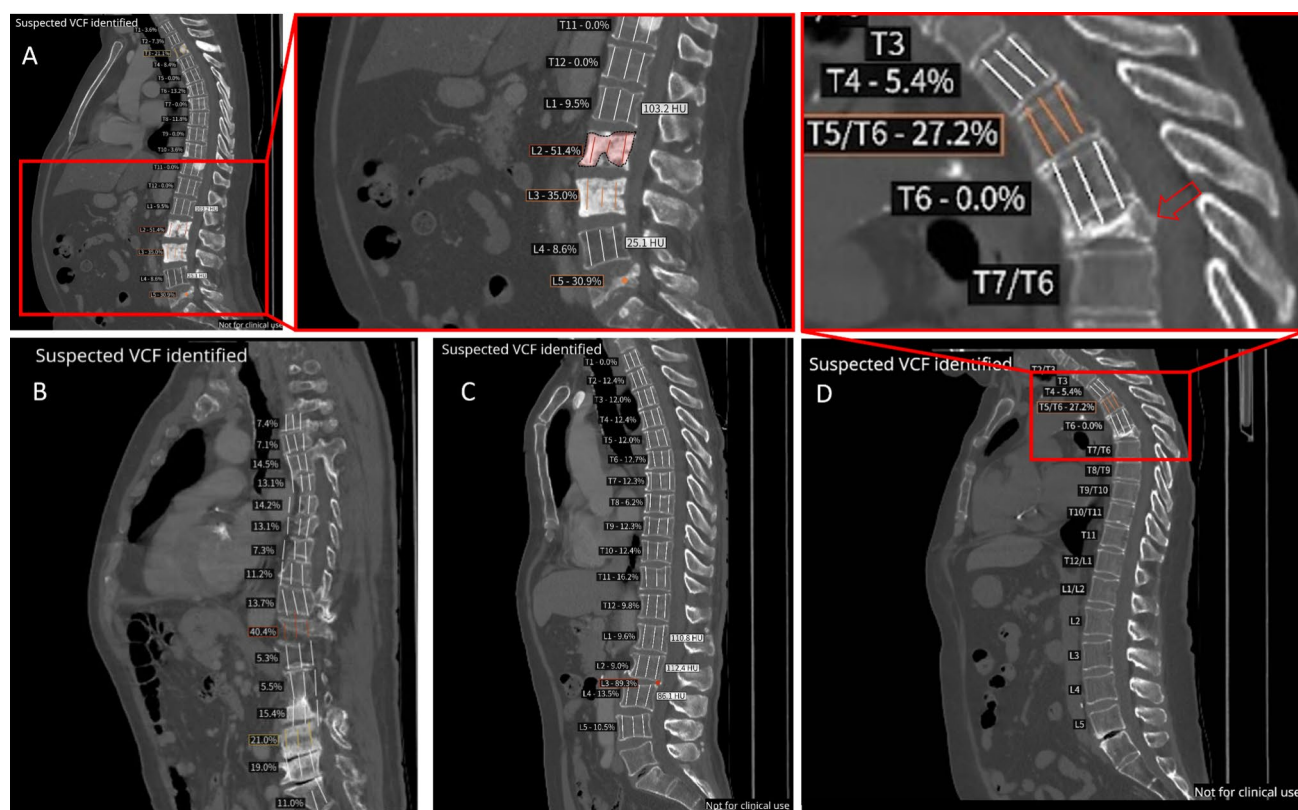
Fig. 3 Examples of false positive and false negative notifications. **A** False-positive case due to an error in vertebral height measurement because of sclerotic vertebral metastases. **B** False-positive case due to incorrect height measurement linked to scoliosis. **C** False-positive case due to incorrect vertebra centroid detection. AI displays a circle in the region to alert that the height loss may not be correct. **D** False-negative for grade 3 T7 vertebral fracture

Fig. 4 Vertebral fractures not reported by radiologists. Among 501 cases (841 vertebrae) flagged for vertebral fracture by DL-based application, 436 cases (680 vertebrae) were confirmed to be true positives by two expert radiologists. The analysis of radiology reports revealed the number of non-reported fractures by radiologists in general practice. Two expert interventional radiologists independently reviewed severe (grade 3) vertebral fractures. Among 10 non-reported cases by radiologists, nine would have been indicated to vertebroplasty
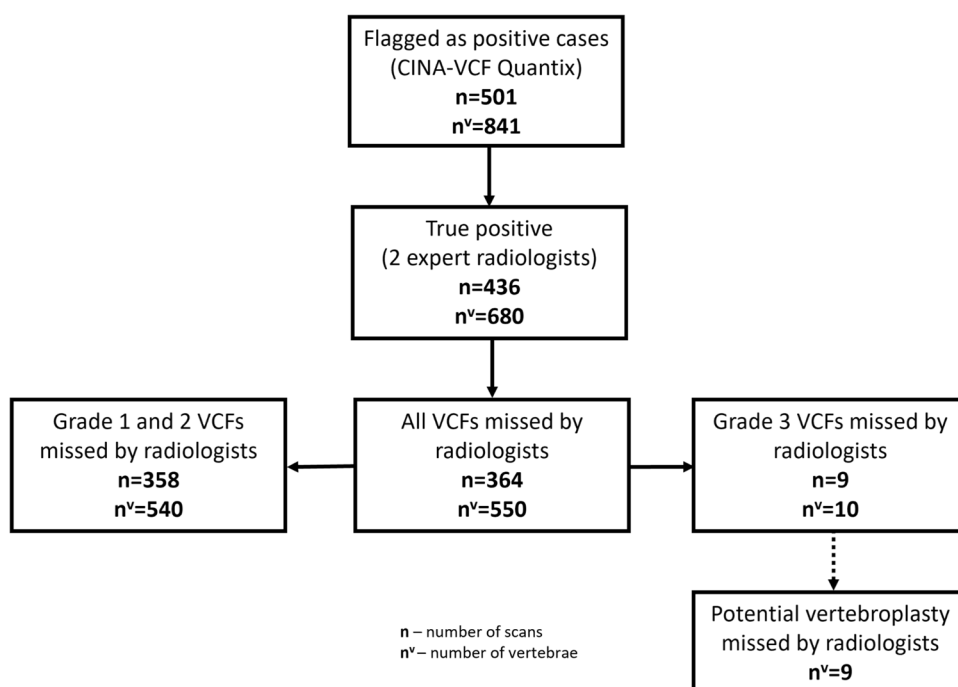
lumbar spine. Among the 13 fractures of patients with grade 3 VCFs presenting vertebral metastases, 6 (46.2%) fractures were related to the metastatic process.

Outcomes for clinical management

Finally, two interventional radiology experts, with 10 and 2 years of clinical experience, independently reviewed all grade 3 fractures to assess the potential indication for vertebroplasty in these severe cases. Among the 30 fractures reviewed, cementoplasty was recommended in 27 cases. Of the 10 fractures initially overlooked by radiologists, cementoplasty was deemed appropriate in nine cases (Fig. 4). This analysis highlights that 33% (9/27) of fractures requiring intervention were detected by the DL-based application but were not reported during routine clinical practice by radiologists.

Discussion

The purpose of this study was to evaluate the diagnostic performance of a DL-based application for incidental VCF detection on routine CT scans of asymptomatic oncological patients. The DL-based tool demonstrated a per-case PPV of 87.0% and a per-vertebra PPV of 80.6% for all VCFs. For grade 3 VCFs, the per-case and per-vertebra PPV were 71.4% and 68.2%. Regarding grade 1–2 VCFs, these values were 88.4% and 81.3%, respectively. Radiologists did not describe in the reports 83.5% of cases with VCF and 80.9% of individual vertebrae fractures detected by DL-based application. Finally, of the 10 vertebrae with grade 3 fractures that were not reported by radiologists and detected by the DL-based tool, nine required an interventional procedure, as assessed by two experienced interventional radiologists.

The automated detection of VCFs is a growing area of research. Various solutions have been proposed across different imaging modalities. Deep learning (DL) tools for standard radiographs have demonstrated high accuracy in VCF detection, exceeding 98.5% [44, 45]. Moreover, these tools have proven to be cost-effective for healthcare systems [46]. Automated VCF detection on MRI has shown performance comparable to that of spine surgeons [47]. Notably, a DL application for MRI-based VCF detection significantly improved accuracy among less-experienced radiologists [48]. DL algorithms for CT-based VCF detection are also advancing, with tools capable of vertebral labeling, segmentation, and metastasis detection [35, 39]. Additionally, CT has proven to be an optimal modality for automated osteoporosis screening, which is particularly relevant for applications detecting VCFs [34, 36, 37]. The performance of DL tools for VCF detection on CT has ranged from 47.4 to 98.7% in sensitivity and from 63.9 to 95.8% in specificity

[31, 33, 38, 40]. The application currently under evaluation, CINA-VCF Quantix, previously demonstrated a sensitivity of 92.3% and specificity of 91.7% in a previous study [43].

Regarding the current study, the application demonstrated relatively high PPV values (87.0% for overall per-case PPV). The lowest PPV was documented for grade 3 VCFs (71.4% for overall per-case PPV). It represented 10 false positive cases out of 35 alerts generated by DL-based applications. The main reasons for these false positive detections were sclerotic vertebral metastases, scoliosis, and incorrect centroid detection. All these causes lead to incorrect vertebral height measurement. Given this PPV, radiologists would need to review an additional one-eighth of the flagged cases. However, since these cases are associated with confounding factors and pathologies highly impacting patient life quality, reviewing these cases can be very advantageous. Regarding the previously published literature, PPV of 46.7% [32] and 13.9% [40] have been previously reported for other automated DL-based tools for VCF detection. Therefore, the currently evaluated application performed better and demonstrated interesting performance for its implementation in clinical practice.

Regarding false positives, the DL-based application was found to generate erroneous alerts in the presence of confounding factors such as sclerotic metastases and scoliosis. To mitigate these inaccuracies, a dedicated dataset enriched with cases exhibiting these conditions could be employed for training. However, false positives associated with scoliosis, particularly those arising from inaccurate centroid estimations, might persist despite this strategy. To address this limitation, the implementation of a specialized algorithm for the reorientation of scoliotic vertebrae along the craniocaudal axis might be proposed.

Concerning the false negatives, two main sources of inaccuracies were identified. The first source of inaccuracies takes place at the measure level. The landmark prediction algorithm precision depends both on the acquisition resolution and vertebral body presentation. Because the model somehow learned to identify patterns associated with the vertebral body contours, it can be confused by either anatomical variations or confounding factors that change the general shape or intensity of these contours leading to errors in the height loss ratio and potentially a vertebra classified as grade 0 instead of 1 or more. The second source of inaccuracies takes place in the preliminary steps of general processing where the vertebral body centers are located using a dedicated deep learning model. These locations are associated with a probability for each candidate vertebra. When this probability is below a confidence level, the candidate vertebra is ignored both for the measures and labelling. When the ignored vertebra is the only one with a high grade, the case can be flagged as negative. Nonetheless, future versions of the device will need to address the current limitations

revealed by the study. Both sources of errors will clearly benefit from more data, with a focus on adding patients with more high-grade vertebra and confounding factors which are currently either ignored during pre-processing or subject to landmark misplacements. However, given that in the more extreme cases the landmark approach may fail (there is a floor in the landmark location accuracy due to the resolution), we may complement the current approach with a classifier dedicated to flag these extreme conditions independently of the measures themselves.

No difference was observed in the DL application performance for detecting VCF in the thoracic or lumbar spine. This consistency is likely due to the training methodology, in which the model was trained at the single-vertebra level. Vertebral height is assessed individually and compared to neighboring vertebrae to compute the vertebral height loss (VHL) based on intervertebral height differences. As a result, the algorithm operates independently of the anatomical region, exhibiting no bias toward thoracic or lumbar vertebrae.

In this study on an asymptomatic oncology patient population, we reported that among true positive cases detected by DL-based application, radiologists did not report 83.5% of VCF-positive patients. This data aligns with previously published data, suggesting that radiologists do not report 84% of patients with VCFs during routine CT scan examinations [49]. In a previously published study on a restricted cohort of patients (105 cases), 12 (63.2%) out of 19 true positive cases detected by the DL-based tool were not reported by radiologists [32]. This highlights a significant proportion of patients who are overlooked by specific prophylactic measures, as previously reported data indicate that 19.2% of individuals with incidental vertebral fractures are likely to develop new vertebral fractures within a year [50]. Moreover, these osteoporotic patients are at an increased risk of hip fractures, which, alongside vertebral fractures, account for the highest economic burden among all fracture types in public health [51]. Therefore, early detection of this high-risk patient group could prove to be a cost-effective strategy [5, 52, 53]. Among the grade 3 fractures not reported by the radiologist, nine would have been eligible for interventional treatment. Importantly, timely vertebroplasty can reduce pain and lower the risk of subsequent fractures [54]. Therefore, these patients could have experienced substantial clinical benefits if the DL-based tool for VCF detection had been implemented during their examination.

This study highlights that 83.5% of VCFs are not reported by radiologists in standard clinical practice in a tertiary cancer center. However, these results should be reviewed in the prism of a real-world clinical practice addressing several limitations. First, the scans are performed in asymptomatic patients with no clinical indication of VCF, thus the radiologists performed an oncological assessment rather than systematically evaluate VCFs. Factors such as work fatigue and high workload may impact radiologists' assessments, limiting the time available for thorough case review and reporting, leading to the omission of incidental findings. Second, radiologists often do not describe VCFs if they remain stable over several years. Oncological patients undergo regular CT scans, often quarterly, over multiple years, whereas the current analysis is performed at a single time point. It is therefore likely that some fractures detected by the software during the study were previously reported in an earlier scan, with no significant changes in the current examination, leading to their omission in the radiology report. Finally, the DL software is highly sensitive compared to the radiologists' eye in detecting grade 1 VCFs. However, these fractures have limited clinical or therapeutic relevance. Taken together, though the outputs of DL application for incidental VCF detection might provide useful clinical information, it becomes clinically relevant for patient outcomes only upon radiologist review.

Two clinical strategies of radiologist-DL tool interactions were previously proposed concerning incidental findings like VCF. On the one hand, DL-based applications can help incidental pathologies detection by acting as a second reader to catch missed cases. On the other hand, it can serve as a triage tool, prioritizing DL-flagged alerts, thus enhancing efficiency of incidental findings detection. However, this latter strategy risks to overlook false negative cases not flagged by DL-based application. It can be overcome by an adjustable balance between sensitivity and specificity of the application and clinical workflow optimization [55]. Regarding CINA-VCF Quantix, this application demonstrated to be a useful tool for complementary use in radiology practice, which would improve clinical workflow with the potential to reduce radiologists' workload and improve vigilance for incidental VCF findings. We consider that both proposed strategies for DL tool implementation have the potential to be effective for the detection of VCFs in clinical routine. However, their efficacy should be evaluated individually, taking into account the specific needs of each clinical center, its area of specialization, and the clinical protocols in place for management of patients with VCF findings.

This study had limitations. First, as a retrospective study, it did not assess the real-world clinical integration or the real-time impact on patient management. Additionally, the retrospective nature of the study may introduce selection bias. To balance this bias, all consecutive CT scans on a given period have been included. Second, it was a single-center study, limiting the generalizability of the results to other centers and medical practices. Moreover, the study focused exclusively on patients with oncological profiles, with specific characteristics related to age, treatment approaches, and follow-up. This population is probably over-exposed to osteoporosis and VCF,

thus increasing these findings in the final cohort. Finally, although reported, we did not make use of the osteoporosis parameters (mean HU) provided by the application, as the primary aim of the study was to assess the diagnostic performance for VCF detection and the unreported rate of VCF-positive patients in routine oncological clinical practice within the oncology department.

Conclusion

In summary, this study demonstrates the potential clinical benefit of the DL-based application integration into a clinical workflow in order to detect opportunistic VCFs on regular CT scans. Importantly, the application proved effective in detecting previously overlooked patients with severe (grade 3) VCFs in a tertiary cancer center, thereby facilitating their redirection to appropriate treatment pathways and offering the potential for improved clinical outcomes. The system can complement radiologists' assessments, improving the identification of patients with VCFs for inclusion in prophylactic treatment pathways and reducing non-reported rates, therefore contributing to enhancing patient long-term outcomes.

Declarations

Conflict of interest Vladimir Laletin, Yasmina Chaibi, Julie Kiewsky, Christophe Avare, Sarah Quenet and Angela Ayobi are employees of Avicenna.AI. The remaining authors have no relevant financial or non-financial interests to disclose.

Ethics approval This study was performed in line with the principles of the Declaration of Helsinki. This is an observational study. This study was approved by the Institutional Review Board of Gustave Roussy Cancer Campus (no.: 2024–381).

References

1. Kondo K (2008) Osteoporotic vertebral compression fractures and vertebral augmentation. Semin Interv Radiol 25:413–424. https://doi.org/10.1055/s-0028-1103000

2. Cauley JA (2013) Public health impact of osteoporosis. J Gerontol A Biol Sci Med Sci 68:1243–1251. https://doi.org/10.1093/gerona/glt093

3. Wong CC, McGirt MJ (2013) Vertebral compression fractures: a review of current management and multimodal therapy. J Multidiscip Healthc 6:205–214. https://doi.org/10.2147/JMDH.S31659

4. Nevitt MC, Ross PD, Palermo L, Musliner T, Genant HK, Thompson DE (1999) Association of prevalent vertebral fractures, bone density, and alendronate treatment with incident vertebral fractures: effect of number and spinal location of fractures. Bone 25:613–619. https://doi.org/10.1016/S8756-3282(99)00202-1

5. Lang S, Walter N, Freigang V, Neumann C, Loibl M, Alt V, Rupp M (2023) Increased incidence of vertebral fractures in German adults from 2009 to 2019 and the analysis of secondary diagnoses, treatment, costs, and in-hospital mortality. Sci Rep 13:6984. https://doi.org/10.1038/s41598-023-31654-0

6. Ballane G, Cauley JA, Luckey MM, El-Hajj Fuleihan G (2017) Worldwide prevalence and incidence of osteoporotic vertebral fractures. Osteoporos Int 28:1531–1542. https://doi.org/10.1007/s00198-017-3909-3

7. Fink HA, Milavetz DL, Palermo L, Nevitt MC, Cauley JA, Genant HK, Black DM, Ensrud KE (2005) What proportion of incident radiographic vertebral deformities is clinically diagnosed and vice versa? J Bone Miner Res 20:1216–1222. https://doi.org/10.1359/JBMR.050314

8. Svedbom A, Alvares L, Cooper C, Marsh D, Ström O (2013) Balloon kyphoplasty compared to vertebroplasty and nonsurgical management in patients hospitalised with acute osteoporotic vertebral compression fracture: a UK cost-effectiveness analysis. Osteoporos Int 24:355–367. https://doi.org/10.1007/s00198-012-2102-y

9. Yang J, Mao Y, Nieves JW (2020) Identification of prevalent vertebral fractures using Vertebral Fracture Assessment (VFA) in asymptomatic postmenopausal women: a systematic review and meta-analysis. Bone 136:115358. https://doi.org/10.1016/j.bone.2020.115358

10. Melton LJ, Atkinson EJ, Cooper C, O'Fallon WM, Riggs BL (1999) Vertebral fractures predict subsequent fractures. Osteoporos Int J Establ Result Coop Eur Found Osteoporos Natl Osteoporos Found USA 10:214–221. https://doi.org/10.1007/s001980050218

11. Cauley JA, Thompson DE, Ensrud KC, Scott JC, Black D (2000) Risk of mortality following clinical fractures. Osteoporos Int 11:556–561. https://doi.org/10.1007/s001980070075

12. Savage JW, Schroeder GD, Anderson PA (2014) Vertebroplasty and kyphoplasty for the treatment of osteoporotic vertebral

compression fractures. J Am Acad Orthop Surg 22:653–664. https://doi.org/10.5435/JAAOS-22-10-653

13. Cockerill W, Lunt M, Silman AJ, Cooper C, Lips P, Bhalla AK, Cannata JB, Eastell R, Felsenberg D, Gennari C, Johnell O, Kanis JA, Kiss C, Masaryk P, Naves M, Poor G, Raspe H, Reid DM, Reeve J, Stepan J, Todd C, Woolf AD, O'Neill TW (2004) Health-related quality of life and radiographic vertebral fracture. Osteoporos Int J Establ Result Coop Eur Found Osteoporos Natl Osteoporos Found USA 15:113–119. https://doi.org/10.1007/s00198-003-1547-4

14. Hallberg I, Ek A-C, Toss G, Bachrach-Lindström M (2010) A striving for independence: a qualitative study of women living with vertebral fracture. BMC Nurs 9:7. https://doi.org/10.1186/1472-6955-9-7

15. Guillevin R, Vallee J-N, Lafitte F, Menuel C, Duverneuil N-M, Chiras J (2007) Spine metastasis imaging: review of the literature. J Neuroradiol 34:311–321. https://doi.org/10.1016/j.neurad.2007.05.003

16. Roodman GD (2004) Mechanisms of bone metastasis. N Engl J Med 350:1655–1664. https://doi.org/10.1056/NEJMra030831

17. Pfeilschifter J, Diel IJ (2000) Osteoporosis due to cancer treatment: pathogenesis and management. J Clin Oncol 18:1570–1593. https://doi.org/10.1200/JCO.2000.18.7.1570

18. Michaud LB, Goodin S (2006) Cancer-treatment-induced bone loss, part 1. Am J Health Syst Pharm 63:419–430. https://doi.org/10.2146/ajhp050045.p1

19. Drake MT (2013) Osteoporosis and cancer. Curr Osteoporos Rep 11:163–170. https://doi.org/10.1007/s11914-013-0154-3

20. Wissing MD (2015) Chemotherapy- and irradiation-induced bone loss in adults with solid tumors. Curr Osteoporos Rep 13:140–145. https://doi.org/10.1007/s11914-015-0266-z

21. Moseley KF, Naidoo J, Bingham CO, Carducci MA, Forde PM, Gibney GT, Lipson EJ, Shah AA, Sharfman WH, Cappelli LC (2018) Immune-related adverse events with immune checkpoint inhibitors affecting the skeleton: a seminal case series. J Immunother Cancer 6:104. https://doi.org/10.1186/s40425-018-0417-8

22. Filippini DM, Gatti M, Di Martino V, Cavalieri S, Fusaroli M, Ardizzoni A, Raschi E, Licitra L (2021) Bone fracture as a novel immune-related adverse event with immune checkpoint inhibitors: case series and large-scale pharmacovigilance analysis. Int J Cancer 149:675–683. https://doi.org/10.1002/ijc.33592

23. Park SH, Jeong YM, Lee HY, Kim EY, Kim JH, Park HK, Ahn HK (2020) Opportunistic use of chest CT for screening osteoporosis and predicting the risk of incidental fracture in breast cancer patients: a retrospective longitudinal study. PLoS ONE 15:e0240084. https://doi.org/10.1371/journal.pone.0240084

24. Ebstein E, Brocard P, Soussi G, Khoury R, Forien M, Khalil A, Vauchier C, Juge PA, Léger B, Ottaviani S, Dieudé P, Zalcman G, Gounant V (2024) Burden of comorbidities: osteoporotic vertebral fracture during non-small cell lung cancer - the BONE study. Eur J Cancer 200:113604. https://doi.org/10.1016/j.ejca.2024.113604

25. Barragán-Montero A, Javaid U, Valdés G, Nguyen D, Desbordes P, Macq B, Willems S, Vandewinckele L, Holmström M, Löfman F, Michiels S, Souris K, Sterpin E, Lee JA (2021) Artificial intelligence and machine learning for medical imaging: a technology review. Phys Med 83:242–256. https://doi.org/10.1016/j.ejmp.2021.04.016

26. Bharadwaj P, Nicola L, Breau-Brunel M, Sensini F, Tanova-Yotova N, Atanasov P, Lobig F, Blankenburg M (2024) Unlocking the value: quantifying the return on investment of hospital artificial intelligence. J Am Coll Radiol JACR S1546–1440(24):00292–00298. https://doi.org/10.1016/j.jacr.2024.02.034

27. Najjar R (2023) Redefining radiology: a review of artificial intelligence integration in medical imaging. Diagn Basel Switz 13:2760. https://doi.org/10.3390/diagnostics13172760

28. Cacciamani GE, Sanford DI, Chu TN, Kaneko M, De Castro Abreu AL, Duddalwar V, Gill IS (2023) Is artificial intelligence replacing our radiology stars? Not yet! Eur Urol Open Sci 48:14–16. https://doi.org/10.1016/j.euros.2022.09.024

29. Bailey CR, Bailey AM, McKenney AS, Weiss CR (2022) Understanding and appreciating burnout in radiologists. Radiographics 42:E137–E139. https://doi.org/10.1148/rg.220037

30. Aggarwal V, Maslen C, Abel RL, Bhattacharya P, Bromiley PA, Clark EM, Compston JE, Crabtree N, Gregory JS, Kariki EP, Harvey NC, Ward KA, Poole KES (2021) 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 13:1759720X2110240. https://doi.org/10.1177/1759720X211024029

31. Roux C, Rozes A, Reizine D, Hajage D, Daniel C, Maire A, Bréant S, Taright N, Gordon R, Fechtenbaum J, Kolta S, Feydy A, Briot K, Tubach F (2022) Fully automated opportunistic screening of vertebral fractures and osteoporosis on more than 150 000 routine computed tomography scans. Rheumatology 61:3269–3278. https://doi.org/10.1093/rheumatology/keab878

32. Rueckel J, Sperl JI, Kaestle S, Hoppe BF, Fink N, Rudolph J, Schwarze V, Geyer T, Strobl FF, Ricke J, Ingrisch M, Sabel BO (2021) Reduction of missed thoracic findings in emergency whole-body computed tomography using artificial intelligence assistance. Quant Imaging Med Surg 11:2486–2498. https://doi.org/10.21037/qims-20-1037

33. Burns JE, Yao J, Summers RM (2017) Vertebral body compression fractures and bone density: automated detection and classification on CT images. Radiology 284:788–797. https://doi.org/10.1148/radiol.2017162100

34. Löffler MT, Jacob A, Scharr A, Sollmann N, Burian E, El Husseini M, Sekuboyina A, Tetteh G, Zimmer C, Gempt J, Baum T, Kirschke JS (2021) Automatic opportunistic osteoporosis screening in routine CT: improved prediction of patients with prevalent vertebral fractures compared to DXA. Eur Radiol 31:6069–6077. https://doi.org/10.1007/s00330-020-07655-2

35. Netherton TJ, Nguyen C, Cardenas CE, Chung C, Klopp AH, Colbert LE, Rhee DJ, Peterson CB, Howell R, Balter P, Court LE (2022) An automated treatment planning framework for spinal radiation therapy and vertebral-level second check. Int J Radiat Oncol 114:516–528. https://doi.org/10.1016/j.ijrobp.2022.06.083

36. Pan Y, Shi D, Wang H, Chen T, Cui D, Cheng X, Lu Y (2020) Automatic opportunistic osteoporosis screening using low-dose chest computed tomography scans obtained for lung cancer screening. Eur Radiol 30:4107–4116. https://doi.org/10.1007/s00330-020-06679-y

37. Pickhardt PJ, Lee SJ, Liu J, Yao J, Lay N, Graffy PM, Summers RM (2019) Population-based opportunistic osteoporosis screening: validation of a fully automated CT tool for assessing longitudinal BMD changes. Br J Radiol 92:20180726. https://doi.org/10.1259/bjr.20180726

38. Tomita N, Cheung YY, Hassanpour S (2018) Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. Comput Biol Med 98:8–15. https://doi.org/10.1016/j.compbiomed.2018.05.011

39. Sekuboyina A, Husseini ME, Bayat A, Löffler M, Liebl H, Li H, Tetteh G, Kukačka J, Payer C, Štern D, Urschler M, Chen M, Cheng D, Lessmann N, Hu Y, Wang T, Yang D, Xu D, Ambellan F, Amiranashvili T, Ehlke M, Lamecker H, Lehnert S, Lirio M, de Olaguer NP, Ramm H, Sahu M, Tack A, Zachow S, Jiang T, Ma X, Angerman C, Wang X, Brown K, Kirszenberg A, Puybareau É, Chen D, Bai Y, Rapazzo BH, Yeah T, Zhang A, Xu S, Hou F, He Z, Zeng C, Xiangshang Z, Liming X, Netherton TJ, Mumme RP, Court LE, Huang Z, He C, Wang L-W, Ling SH, Huynh LD,

Boutry N, Jakubicek R, Chmelik J, Mulay S, Sivaprakasam M, Paetzold JC, Shit S, Ezhov I, Wiestler B, Glocker B, Valentinitsch A, Rempfler M, Menze BH, Kirschke JS (2021) VerSe: a vertebrae labelling and segmentation benchmark for multi-detector CT images. Med Image Anal 73:102166. https://doi.org/10.1016/j.media.2021.102166

40. Marques Dos Santos F, Senra C, Mota C, Antunes F, Matos I, Sousa PT, Marques ID (2024) Comparing artificial intelligence ai-rad companion chest CT vs experts on emphysema, lung nodules, thoracic aorta and thoracic spine. Acta Radiol Port. https://doi.org/10.25748/ARP.29177

41. Genant HK, Wu CY, van Kuijk C, Nevitt MC (1993) Vertebral fracture assessment using a semiquantitative technique. J Bone Miner Res Off J Am Soc Bone Miner Res 8:1137–1148. https://doi.org/10.1002/jbmr.5650080915

42. Jang S, Graffy PM, Ziemlewicz TJ, Lee SJ, Summers RM, Pickhardt PJ (2019) Opportunistic osteoporosis screening at routine abdominal and thoracic CT: normative l1 trabecular attenuation values in more than 20 000 adults. Radiology 291:360–367. https://doi.org/10.1148/radiol.2019181648

43. Guenoun D, Quemeneur MS, Ayobi A, Castineira C, Quenet S, Kiewsky J, Mahfoud M, Avare C, Chaibi Y, Champsaur P (2025) Automated vertebral compression fracture detection and quantification on opportunistic CT scans: a performance evaluation. Clin Radiol 83:106831. https://doi.org/10.1016/j.crad.2025.106831

44. Seo JW, Lim SH, Jeong JG, Kim YJ, Kim KG, Jeon JY (2021) A deep learning algorithm for automated measurement of vertebral body compression from X-ray images. Sci Rep 11:13732. https://doi.org/10.1038/s41598-021-93017-x

45. Zhang H, Yuan G, Zhang Z, Guo X, Xu R, Xu T, Zhong X, Kong M, Zhu K, Ma X (2024) A multi-scene deep learning model for automated segmentation of acute vertebral compression fractures from radiographs: a multicenter cohort study. Insights Imaging 15:290. https://doi.org/10.1186/s13244-024-01861-y

46. Curl PK, Jacob A, Bresnahan B, Cross NM, Jarvik JG (2024) Cost-effectiveness of artificial intelligence-based opportunistic compression fracture screening of existing radiographs. J Am Coll Radiol 21:1489–1496. https://doi.org/10.1016/j.jacr.2023.11.029

47. Yoda T, Maki S, Furuya T, Yokota H, Matsumoto K, Takaoka H, Miyamoto T, Okimatsu S, Shiga Y, Inage K, Orita S, Eguchi Y, Yamashita T, Masuda Y, Uno T, Ohtori S (2022) Automated differentiation between osteoporotic vertebral fracture and malignant vertebral fracture on MRI using a deep convolutional neural network. Spine 47:E347–E352. https://doi.org/10.1097/BRS.0000000000004307

48. Yeh L-R, Zhang Y, Chen J-H, Liu Y-L, Wang A-C, Yang J-Y, Yeh W-C, Cheng C-S, Chen L-K, Su M-Y (2022) A deep learning-based method for the diagnosis of vertebral fractures on spine MRI: retrospective training and validation of ResNet. Eur Spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc 31:2022–2030. https://doi.org/10.1007/s00586-022-07121-1

49. Carberry GA, Pooler BD, Binkley N, Lauder TB, Bruce RJ, Pickhardt PJ (2013) Unreported vertebral body compression fractures at abdominal multidetector CT. Radiology 268:120–126. https://doi.org/10.1148/radiol.13121632

50. Lindsay R (2001) Risk of new vertebral fracture in the year following a fracture. JAMA 285:320. https://doi.org/10.1001/jama.285.3.320

51. Shi N, Foley K, Lenhart G, Badamgarav E (2009) Direct healthcare costs of hip, vertebral, and non-hip, non-vertebral fractures. Bone 45:1084–1090. https://doi.org/10.1016/j.bone.2009.07.086

52. Tatangelo G, Watts J, Lim K, Connaughton C, Abimanyi-Ochom J, Borgström F, Nicholson GC, Shore-Lorenti C, Stuart AL, Iuliano-Burns S, Seeman E, Prince R, March L, Cross M, Winzenberg T, Laslett LL, Duque G, Ebeling PR, Sanders KM (2019) The cost of osteoporosis, osteopenia, and associated fractures in Australia in 2017. J Bone Miner Res 34:616–625. https://doi.org/10.1002/jbmr.3640

53. Patel N, Dahl K, O'Rourke R, Williamson A, Chatfield MD, Fong KM, Yang IA, Marshall HM (2023) Vertebral CT attenuation outperforms standard clinical fracture risk prediction tools in detecting osteoporotic disease in lung cancer screening participants. Br J Radiol 96:20220992. https://doi.org/10.1259/bjr.20220992

54. Grafe IA, Da Fonseca K, Hillmeier J, Meeder P-J, Libicher M, Nöldge G, Bardenheuer H, Pyerin W, Basler L, Weiss C, Taylor RS, Nawroth P, Kasperk C (2005) Reduction of pain and fracture incidence after kyphoplasty: 1-year outcomes of a prospective controlled trial of patients with primary osteoporosis. Osteoporos Int 16:2005–2012. https://doi.org/10.1007/s00198-005-1982-5

55. Ammari S, Camez AO, Ayobi A, Quenet S, Zemmouri A, Mniai EM, Chaibi Y, Franciosini A, Clavel L, Bidault F, Muller S, Lassau N, Balleyguier C, Assi T (2024) Contribution of an artificial intelligence tool in the detection of incidental pulmonary embolism on oncology assessment scans. Life Basel Switz 14:1347. https://doi.org/10.3390/life14111347