TRIDHA
Journal of Clinical Cases & Reports
Volume 2023, Issue S12, 7 pages
https://doi.org/10.46619/joccr.2023.6-S12.1069
www.tridhascholars.org

SHORT COMMUNICATION

Evaluation of AI Applications in Thoracic Imaging: Performance in the Detection of Vertebral Body Alterations

Carlos Senra[1*], Cristina Mota[1], Francisco Antunes[1], Inês Matos[1], Filipa Santos[1], Inês Dias Marques and Pedro Sousa[1]

Imaging Service of the Vila Nova de Gaia Hospital Center/Espinho, Portugal

Correspondence should be addressed to Carlos Senra, Imaging Service of the Vila Nova de Gaia Hospital Center/Espinho, Portugal

Received: 13 December 2022; Accepted: 20 December 2022; Published: 27 December 2022

Copyright © Carlos Senra. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

PURPOSE

Fractures of the vertebral bodies of the dorsal spine are an increasingly frequent pathology, especially in the elderly; however, these are often unreported by radiologists. This study aims to evaluate the performance of an artificial intelligence (AI) algorithm assessing spinal alterations in the detection of fractures of the dorsal spine.

MATERIALS AND METHOD

350 patients who underwent chest CT scans were retrospectively evaluated by the AI algorithm (Siemens AI RAD Companion) that assigned a score (MSK range) ranging from 1 to 4; score 1 is considered normal and score 2 to 4 abnormal. Subsequently, the scans were reviewed by 2 experienced radiologists which evaluated the clinical relevance of the spinal alterations in each study and whether these alterations were previously reported by the reporting radiologist.

RESULTS

348 scans were successfully processed and 2 were excluded. The AI algorithm detected abnormal spinal findings in 130 patients (37.4% of cases). Of those, only 18 (13.8%) were considered to have clinically relevant findings by the reviewing radiologist. Among these 18 cases, 15 were previously reported by the radiologist reporting the study (83.3%), while in 3 patients the spinal alteration was not reported (16,67%). Sensitivity and specificity of the AI algorithm were estimated at 47,37% and 63,87%, respectively.

CONCLUSION

This study demonstrates that this AI tool can be useful as an auxiliary diagnostic method for dorsal spine fractures.

KEYWORDS

Artificial intelligence; Thoracic imaging; Spine; Fractures

INTRODUCTION

Dorsal Spine Fractures

Fractures of the vertebral bodies of the dorsal spine are an increasingly frequent pathology, especially in the elderly (mainly due to osteoporosis, medications and other chronic conditions), becoming more common as Western populations age; however, these are often unreported by radiologists [1,2]. This is mainly because the spine is a "hidden area" in routine thoracic studies, especially if reconstructions on the coronal and sagittal plane are not available. In addition, radiologists are not always aware of spinal fractures associated morbidity and treatment options.

With an increasing workload for the radiologist, artificial intelligence (AI) algorithms may reduce repetitive tasks, standardize reports and increase diagnostic precision when interpreting imaging of the dorsal spine.

AI-Rad Companion Chest Computed Tomography

AI-Rad Companion Chest Computed Tomography (AIRC-cCT) (Siemens Healthineers®, Germany) is AI-based automated post-processing, augmented workflow solution. It automatically detects, highlights, characterizes and quantifies relevant anatomies and abnormalities. Results are displayed as structured tables and as 2D/3D reconstructed images series with colored contours. Additionally, a traffic-light color scheme is used to highlight/classify measurements outside the defined thresholds. By integrating this solution into clinical workflows, both images, and supporting information can be accessed on any PACS for reporting. AIRC-cCT detects and classifies pulmonary emphysema and pulmonary nodules and segments and evaluates the dorsal spine and the diameter of the aorta.

AIRC-cCT is CE marked and has been cleared by Food and Drug Administration (FDA) for clinical use.

Dorsal Spine Evaluation

The main parameter of bone health is bone density, which is defined as the quantity of mineral bone contained within total bone tissue and is usually quantified using dual-energy x-ray absorptiometry (DEXA) [3,4].

This bone health marker is frequently ignored, particularly in low-risk/asymptomatic patients which in conjunction with the low availability of DEXA scanners in many countries makes the diagnosis and follow-up of vertebral body collapse challenging. This, however many times undervalued, is associated with many co-morbidities, reduces the quality of life, and increases healthcare costs, justifying the importance of early diagnosis and appropriate treatment.

On the other hand, CT scanners are universal and, even though they don't provide a precise quantitative measurement of bone mineral density, the average Hounsfield Units (HU) values of each vertebral body, in routine CT scans, may be used as an opportunistic and convenient screening tool [4]. However, reading time of each CT scan may possibly be increased significantly, and therefore the cost would rise.

AIRC-cCT algorithm segments the dorsal spine with representation on 2D and 3D, measuring in each dorsal vertebral body the average HU and their heights (anterior, middle and posterior), highlighting the values outside the defined thresholds, offering an easy and simple way to diagnose dorsal spine fractures.

This study aims to evaluate the performance of an artificial intelligence algorithm (AIRC-cCT, version VA22B, Siemens Healthineers®, Germany) assessing spinal alterations in the detection of fractures of the dorsal spine against Radiologists and understand if this tool has positive clinical impact in imaging departments.

MATERIALS AND METHOD

All relevant Ethical approval for this study was obtained from our institution's local Ethics Commission.

350 chest CT scans, executed with or without intravenous iodine contrast agent, were retrospectively and randomly selected from the pool of exams performed in the month of January of 2021 at our institution. We included outpatient adults ≥22 years old to meet with software requirements. Medium (soft tissue) and low kernel (lung parenchyma) axial image series from selected CT studies were originally stored on our institution's PACS (SECTRA IDS7®). These series were uploaded to AIRC-cCT via the teamplay digital health platform (Teamplay, Siemens®, Germany), which automatically anonymizes, encrypts and transfers the data. AIRC-cCT results were assessed using AI-Rad Companion online platform and stored within teamplay Images.

The variables evaluated by the AI algorithm were emphysema, lung nodules, thoracic aorta and thoracic spine. For this study we only consider the data from the thoracic spine.

AIRC-cCT uses a classifying score that outlines bone alterations (MSK RANGE 1-4), where score 1 equals no changes/normal and score 2-4 mean spine alterations, with 4 representing the most severe changes. Although this score is user-configurable, we used the default configuration (Figure 1).

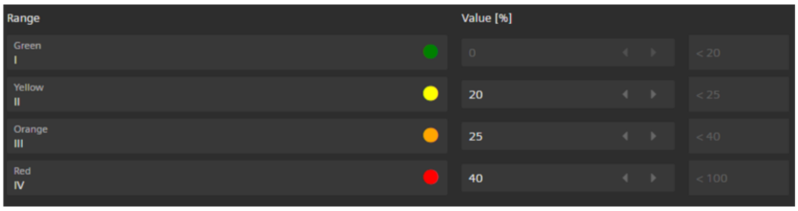

Figure 1: AIRC-cCT score ranges for overall classification of the thoracic spine changes.

Expert Panel included two radiologists from our hospital with over 10 years of experience who evaluated the exams blinded to the report and analysis. For thoracic vertebrae assessment, the Expert Panel's opinion was considered the ground truth.

Subsequently, another two investigators reviewed the reports from the selected studies and verified if some spine anomaly had been reported.

Statistical analysis of the final data was done with IBM® SPSS® Statistics 28. Sensitivity, specificity, positive and negative predictive values (PPV and NPV respectively), negative and positive likelihood ratio were calculated to evaluate the effectiveness in detecting thoracic spine fractures. Cohen's Kappa coefficient was used to assess the findings' agreement between the relevant groups. P-value was calculated with Fisher method.

RESULTS

Population Selection

The 350 patients enrolled in this study have a median age of 65 years (Range: 24 years to 96 years), 199 patients (56.9%) were males and 151 females.

Out of 350 randomly selected patients, the program was able to successfully process 348; 2 of them were excluded due to slice thickness >3 mm (processing error of 0.6%).

Thoracic Spine

The AI algorithm automatically classified 130 patients as abnormal (Attributing a score between 2-4) - 37.4% of scans ranked as having abnormal spine findings. On the other hand, the remaining 218 patients were classified as normal (score 1). Among the 130 patients classified as abnormal by the AIRC-cCT only 18 (13.8%) were considered to have clinically relevant findings by the Expert Panel. From these 18 patients, 15 of them had changes reported by the reporting radiologist (83.3%), in contrast to the other 3 patients in which no spine change had been referred in the report (16.7%). Table 1 summarizes the MSK score attributed to the AIRC-cCT to 348 patients and the corresponding classification attributed by the Expert Panel and the data from the reporting.

Table 1: Summary of MSK score attributed by the algorithm and corresponding data from Expert panel and reporting radiologist.

MSK Range	AI-RAD Companion Chest CT	Expert Panels (Scans with Clinically Relevant Findings)	Reporting Radiologist (Findings Reported)
1	218	20	3
2	56	8	1
3	46	8	6
4	28	2	5
2 to 4	130	18	15

From another point of view, the Expert panel considered 20 patients to have spine alterations who were not detected by the AI algorithm - and so classified as score 1. Table 2 compares the classification made for the AIRC-cCT and the classification attributed for the Expert panel.

Table 2: Comparison between AIRC-cCT and expert panel; MSK RANGE 1 considered normal (No) and MSK RANGE 2-4 considered abnormal (Yes).

			Expert Panels	
			Yes	**No**
AI-RAD Companion Chest CT	**Yes**	130	18	112
	No	218	20	198
		348	38	310

Overall, the spine assessment and spine segmentation were considered reasonable. The program was able to detect spine alterations such as vertebral body fractures that were both considered relevant by Expert panels and reported by the radiologist (Figure 2) and, on the other hand, the program was able to detect spine alteration not reported by the radiologist (Figure 3). Despite that, there is still a high number of false positives, in part due to wrong segmentation/measurement by the software. We noticed a systematic wrong segmentation of D1 (Figure 4) and patients with scoliosis (Figure 5).

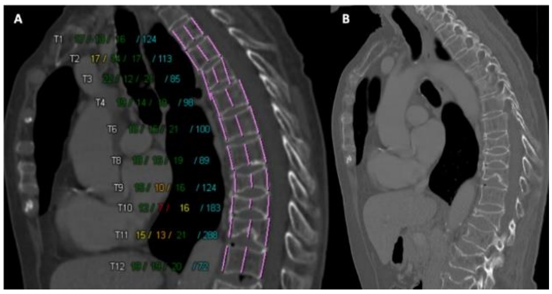

Figure 2: **(A)** Siemens® AI-RAD Companion Chest CT MSK results (MSK RANGE 4), and corresponding **(B)** sagittal chest CT. The algorithm classified this study as a score 4, which was in agreement with the expert panel (Clinically relevant) and was equally referred by the reporting radiologist. On image **(B)**, we can see several compressive fractures of the vertebral bodies of the dorsal spine.

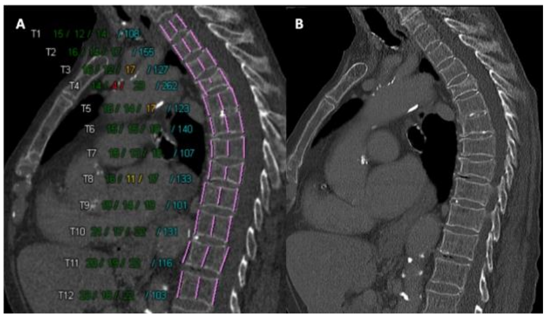

Figure 3: **(A)** Siemens® AI-RAD Companion Chest CT MSK results (MSK RANGE 4), and corresponding **(B)** sagittal chest CT. The algorithm classified this study as a score 4, which was in agreement with the expert panel (clinically relevant), however, in this case it was not reported by the reporting radiologist. The segmentation of the middle portion of the D4 vertebral body was inaccurate. On image **(B)**, we can see a compressive fracture of D4 (arrow), correctly identified in the segmentation by the algorithm in **(A)**.

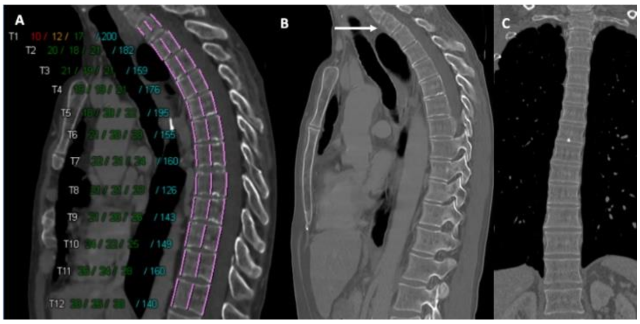

Figure 4: **(A)** Siemens® AI-RAD Companion Chest CT MSK results (MSK RANGE 4), and corresponding **(B)** sagittal and **(C)** coronal chest CT. The algorithm classified this study as score 4, based on a change in the height of the vertebral body of D1 as we can see in the segmentation in **(A)**, however the expert panel in this case concluded that the changes were not clinically relevant (images **(B)** and **(C)**) probably resulting from a segmentation error.

Statistically, sensitivity was 47.4% and specificity 63.9%. The negative predictive value was 90.8% and the positive predictive value was of only 13.9%. The concordance was only mild in stretch due to the rate of false positive (Cohen's kappa of 62.1%, p <0.0001) (table 3).

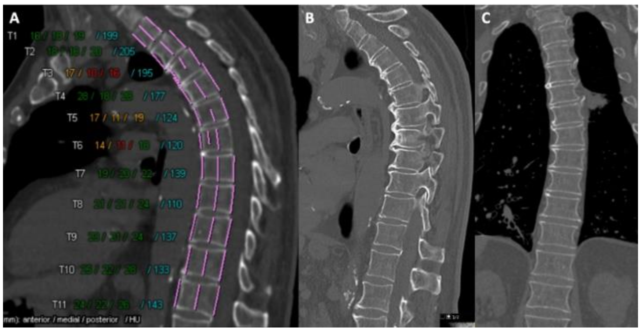

Figure 5: **(A)** Siemens® AI-RAD Companion Chest CT MSK results (MSK RANGE 4), and corresponding **(B)** sagittal and **(C)** coronal chest CT. Despite MSK RANGE 4, the expert panel concluded that there was no clinical relevance, corresponding only to degenerative changes in the spine. A possible explanation for the error in the segmentation of the algorithm could be due to mild scoliosis as shown in image **(C)**.

Table 3: Statistic results of AIRC-cCT.

Statistic	Value	95% CI
Sensitivity	47.37%	30.98% to 64.18%
Specificity	63.87%	58.25% to 69.22%
Positive Predictive Value	13.85%	10.02% to 18.82%
Negative Predictive Value	90.83%	87.86% to 93.12%
Accuracy - Cohen's Kappa	62.07%	56.74% to 67.19%

DISCUSSION/CONCLUSIONS

This study aimed to assess the advantage of the application of artificial intelligence algorithms as an auxiliary method for the diagnosis of dorsal spine fractures. Our results demonstrated that AIRC-cCT is still substantially redundant with a high number of false positives (positive predictive value of only 13.9%), suggesting that this application is not yet fit as a screening tool for a radiology department.

The authors considered the spine segmentation globally well achieved; however, a few spine segmentation errors still occur, mainly in segmentation of the first dorsal vertebra and in patients with scoliosis. These findings open an opportunity to improve software applications. We still considered the delivered results as annotated 3D spine a clean and easy data-driven way of presenting the results obtained.

Our findings suggest that even though AIRC-cCT is still not suitable as a diagnostic tool for dorsal spine changes, it is a very promising and innovative AI application. Further improvements and software upgrades are needed to be more useful in clinical practice.

Previous studies evaluated the use of AI applications for estimation of bone density in chest CT. However, to our knowledge, this is the first study that compared the efficacy of an AI algorithm in the detection of vertebral body fractures against reporting radiologists and expert panel opinion [5].

Additional studies are needed focusing on avoiding false-positive results and improving accuracy.

CONFLICT OF INTEREST

No conflict of interest.

REFERENCES

1. Weerink LBM, Folbert EC, Kraai M et al. (2014) Thoracolumbar spine fractures in the geriatric fracture center: Early ambulation leads to good results on short term and is a successful and safe alternative compared to immobilization in elderly patients with two-column vertebral fractures. Geriatric Orthopaedic Surgery & Rehabilitation 5(2): 43-49.
2. Savage RH, van Assen M, Martin SS et al. (2020) Utilizing artificial intelligence to determine bone mineral density via chest computed tomography. Journal of Thoracic Imaging 35: S35-S39.
3. Lewiecki EM, Binkley N, Morgan SL et al. (2016) Best practices for dual-energy X-ray absorptiometry measurement and reporting: International Society for Clinical Densitometry Guidance. Journal of Clinical Densitometry 19(2): 127-140.
4. Gausden EB, Nwachukwu BU, Schreiber JJ et al. (2017) Opportunistic use of CT imaging for osteoporosis screening and bone density assessment: A qualitative systematic review. Journal Of Bone and Joint Surgery 99(18): 1580-1590.
5. Fischer AM, Yacoub B, Savage RH et al. (2020) Machine learning/deep neuronal network: Routine application in chest computed tomography and workflow considerations. Journal of Thoracic Imaging 35: S21-S27.