Automatic deep learning-based assessment of spinopelvic coronal and sagittal alignment

Mohamed Zerouali, Alexandre Parpaleix, Mansour Benbakoura, Caroline Rigault, Pierre Champsaur, Daphné Guenoun

Automatic deep learning-based assessment of spinopelvic coronal and sagittal alignment

Mohamed Zerouali[a], Alexandre Parpaleix[b], Mansour Benbakoura[b], Caroline Rigault[b], Pierre Champsaur[a,c], Daphné Guenoun[a,c,*]

[a] Department of Radiology, Institute for Locomotion, Sainte-Marguerite Hospital, APHM, 13009 Marseille, France
[b] Milvue, 75014 Paris, France
[c] Institute of Movement Sciences (ISM), CNRS, Aix Marseille University, 13005 Marseille, France

ARTICLE INFO

Keywords:
Artificial intelligence
Automated analysis
Uniplanar whole-spine radiographs
Deep learning
Spine deformities

ABSTRACT

Purpose: The purpose of this study was to evaluate an artificial intelligence (AI) solution for estimating coronal and sagittal spinopelvic alignment on conventional uniplanar two-dimensional whole-spine radiograph.
Material and methods: This retrospective observational study included 100 patients (35 men, 65 women) with a median age of 14 years (IQR: 13, 15.25; age range: 3–64 years) who underwent conventional uniplanar two-dimensional whole-spine radiograph in standing position between January and July 2022. Ten most commonly used spinopelvic coronal and sagittal parameters were retrospectively measured without AI by a junior radiologist and approved or adjusted by a senior musculoskeletal radiologist to reach final measurements. Final measurements were used as the ground truth to assess AI performance for each parameter. AI performances were estimated using mean absolute errors (MAE), intraclass correlation coefficient (ICCs), and accuracy for selected clinically relevant thresholds. Readers visually classified AI outputs to assess reliability at a patient-level.
Results: AI solution showed excellent consistency without bias in coronal (ICCs $\geq$ 0.95; MAE $\leq$ 2.9° or 1.97 mm) and sagittal (ICCs $\geq$ 0.85; MAE $\leq$ 4.4° or 2.7 mm) spinopelvic evaluation, except for kyphosis (ICC = 0.58; MAE = 8.7°). AI accuracy to classify low Cobb angle, severe scoliosis or frontal pelvic asymmetry was 91% (95% CI: 85–96), 99% (95% CI: 97–100) and 94% (95% CI: 89–98), respectively. Overall, AI provided reliable measurements in 72/100 patients (72%).
Conclusion: The AI solution used in this study for combined coronal and sagittal spinopelvic balance assessment provides results consistent with those of a senior musculoskeletal radiologist, and shows potential benefit for reducing workload in future routine implementation.

1. Introduction

Spine and/or pelvis deformity refers to morphological abnormalities in the coronal, sagittal, or axial position as a deviation from the normal position. The prevalence of such deformities can be up to 2% of adults for scoliosis and consequences on quality of life are one current public health issue [1,2]. The assessment and follow-up of spinopelvic alignment relies on radiographic measurements via either conventional uniplanar x-ray systems or orthogonal biplanar acquisitions at low dose stereoradiography [3], allowing to better understand postural adaptation and the close relationship between pelvis and spine [4]. Several radiographic parameters have been developed and validated on conventional uniplanar two-dimensional whole-spine radiographs to evaluate spinopelvic harmony and the sagittal balance of the spine [5]. Cobb angle (CA) is the most widely used parameter to evaluate frontal deviation of the spine and the reference standard to diagnose and monitor scoliosis [6]. Frontal pelvic asymmetry (FPA), sacral slope (SS), pelvic tilt (PT) and pelvic incidence (PI) are the most established parameters to evaluate spinopelvic balance.

Although manual measurements are found precise and accurate by expert readers [7,8], they remain burdensome and time-consuming in the routine radiologists' workflow. Computer-aided semi-automated techniques have been developed for the last ten years but the recent improvement of deep learning techniques offers new perspectives, with promising results on automated deep learning-based assessment of

Abbreviations: AI, Artificial intelligence; CA, Cobb angle; CI, Confidence interval; CVA, Coronal vertical axis; FN, False negative; FP, False positive; FPA, Frontal pelvic asymmetry; GT, Ground truth; ICC, Intraclass correlation coefficient; IQR, Interquartile range; MAE, Mean absolute error; NPV, Negative predictive value; PI, Pelvic incidence; PPV, Positive predictive value; PT, Pelvic tilt; SS, Sacral slope; SD, Standard deviation; SVA, Sagittal vertical axis; TN, True negative; TP, True positive

* Corresponding author:
E-mail address: daphne.guenoun@ap-hm.fr (D. Guenoun).

angular deviations in the coronal or sagittal plane [9–12]. However, none has assessed the performance and potential impact of a full solution for coronal and sagittal measurements of spine and pelvis parameters for routine radiologic implementation.

The main objective of this study was to assess, in a pediatric and adult population, the consistency of an AI solution for coronal and sagittal spinopelvic alignment assessment compared to the gold standard of musculoskeletal radiologist measurements. Secondary objectives were: (*i*) to evaluate the performance of the AI solution to classify measurements based on admitted threshold values for CA, FPA and SS; (*ii*) to report the potential errors of young/non expert readers; and (*iii*), to evaluate the potential impact of the AI solution on an overall workflow.

2. Materials and methods

2.1. Study design

This retrospective, non-interventional study was conducted in our institution and received nonfinancial support by Milvue, which provided the AI model (SmartXpert®). All authors had full control of the data and information submitted for publication. This study was approved by our institutional review board (CRM-2208–299). This study follows the CLAIM Checklist for AI in medical imaging [13].

2.2. Data source and processing

The picture archiving and communication system of our institution was queried to retrieve adult and pediatric patients who underwent whole-spine coronal and sagittal conventional radiographs in standing position from January to July 2022 by two radiologists (M.Z., a four-year resident and D.G., a senior radiologist with 12 years of experience in musculoskeletal imaging). The two radiologists selected the first 100 individual patients fulfilling the inclusion and exclusion criteria of the study. Patients had their imaging examinations in four different imaging departments of our institution using DigitalDiagnost™ (Philips Healthcare) or Clisis™ (Primax) tables. Inclusion criteria were: (*i*), whole-spine conventional radiographs acquired in standing position; and (*ii*), adequate full thoracolumbar spine visualization from external auditory meatus to femoral heads. Exclusion criteria were: (*i*), refusal to participate in the study; (*ii*), congenital spine abnormality; (*iii*), history of spine or pelvic surgery; (*iv*), motion artifact during image acquisition; or (*v*), patient had more than three spinal curvatures. DICOM metadata were parsed by the resident radiologist to extract age and gender for data collection purpose.

2.3. Study sample junior radiologist measurements

The radiology resident (M.Z.) blinded to AI measurements, performed measurements on the selected sagittal and coronal views of the study sample to provide the 10 following spinopelvic coronal and sagittal values according to standard practice of the radiology department. The 10 values were defined as follows: (*i*), Cobb angle (CA - in degrees) was defined as the angle between upper endplate of the upper vertebra and the lower endplate of the lower vertebra; (*ii*), Coronal vertical axis (CVA - in millimeters) was measured in the coronal plane and corresponded to the horizontal distance of the vertical lowered on the center of C7 (coronal C7 plumb line) and the middle of the S1 sacral endplate; (*iii*), Sagittal vertical axis (SVA - in millimeters) was measured in the sagittal plane and corresponded to the horizontal distance of the vertical lowered on the center of C7 (sagittal C7 plumb line) and the postero-superior end of S1 sacral endplate; (*iv*), Thoracic kyphosis T1-T12 (in degrees) was measured between the upper T1 endplate and the lower T12 endplate and expressed in degrees; (*v*), Lumbar lordosis L1-S1 was measured between the upper L1 endplate and the upper S1 endplate and expressed in degrees; (*vi*), T9 spinopelvic inclination (T9-SPI) was defined as the angular value between the vertical lowered at the center of the hip heads and the line joining this point to the center of the T9 vertebra and expressed in degrees; (*vii*), Frontal pelvic asymmetry (FPA) was defined as the vertical height difference between the two sides of the pelvis using the upper iliac crest point as the reference for each pelvis side and expressed in millimeters; (*viii*), Sacral slope (SS) was defined as the angle between a line tangent to the upper S1 endplate and horizontal line and expressed in degrees; (*ix*), Pelvic tilt (PT) was defined as the angle between the vertical lowered on the center of the hip heads and the line connecting the center of the hip heads to the middle of the S1 sacral endplate and expressed in degrees; (*x*), Pelvic incidence (PI), which is a key parameter to pelvis sagittal balance, was defined as the angle between the perpendicular to the upper S1 endplate passing through its center and the line connecting this point to the axis of the femoral heads and expressed in degrees.

All examinations were visualized and measured on the institutional local DICOM viewer used in routine by the readers (Centricity™ Universal Viewer, GE Healthcare). Resident radiologist's assessment time was recorded for each patient.

2.4. Ground truth definition

The senior musculoskeletal radiologist (D.G.) reviewed all measurements made by the resident radiologist and, if needed, corrected the inaccurate ones. Based on this review, the ground truth was defined for subsequent data analysis.

2.5. Deep learning algorithm description and study sample AI measurements

2.5.1. AI solution description

The development dataset consisted of equally represented full body and full spine radiographs with both the coronal and sagittal views collected from radiology departments of seven French centers and annotated by radiographers and senior musculoskeletal radiologists, both with more than 10 years of experience, on a dedicated platform internal to Milvue and after proper training. This dataset was split into three subsets, used for the training (80%), validation (20%), and internal testing of the model (10%). The deep learning algorithm code was based on the "TensorFlow Model Garden" framework [14] further revised and adjusted by Milvue for spinopelvic measurements. As part of the model-training process, data augmentation techniques were applied, including random rotations, blurring, translation, flipping, cropping, and resizing. The measurements were performed in three steps as described in Fig. 1. First, a classification neural network determined the body part and the view of the input image and allowed for further processing if the input image was recognized as a full spine, otherwise the measurement analysis was not performed. Second, a landmark-detection neural network localized the anatomic points useful for the measurements. Third, the detected landmarks were post-processed using geometric formulas to provide values of all required measurements and angles. For every analyzed view, reference lines for measurements and/or angles were outputted as burnt in pixels in the original input image and referred to the radiological output for visual assessment of the AI analysis.

2.5.2. Study sample assessment of AI solution performance

All included examinations were locally anonymized, uploaded to a local server for AI processing (MilvueSuite v1.21, Milvue), and processed for automatic measurements as described in section 2.5.1. Each whole-spine radiographic projection was independently analyzed and AI-based radiological outputs were given to the two radiologists for consensus visual evaluation of the measurement's reliability using a three categories classification as follows: (1), reliable, (*i.e.*, all

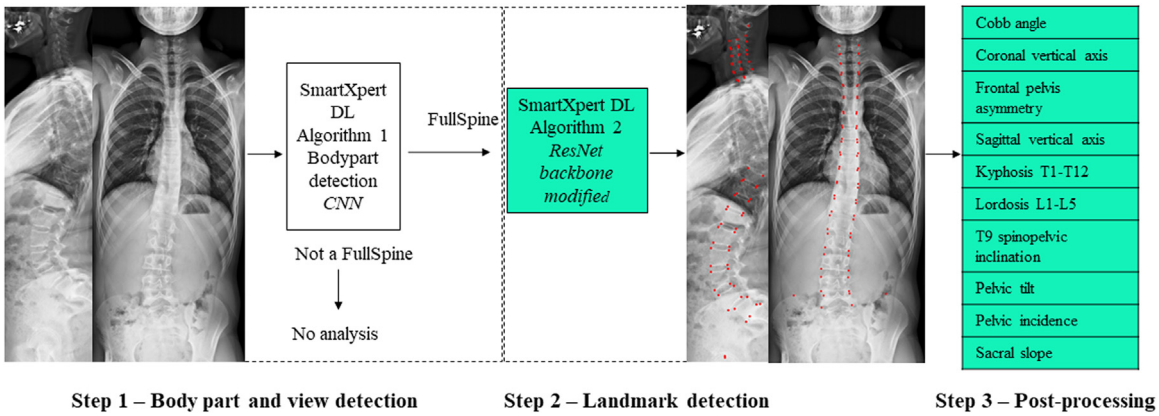

Fig. 1. SmartXpert® (Milvue, France) architecture for fully automated spinopelvic coronal and sagittal assessment. DL indicates deep learning; CNN indicates convolutional neural network.

measurements were accepted without the need of correction); (2), moderately reliable, (*i.e.*, 1 wrong measurement that needed correction; and (3), unreliable, (*i.e.*, two or more wrong measurements).

2.6. Statistical analyses

Quantitative variables were expressed as mean and standard deviation (SD) or median and interquartile range (IQR; Q1, Q3) depending on data distribution, and categorical variables were expressed as raw numbers and percentages [15]. As compared to the ground truth, AI performance metrics for each spinopelvic alignment parameter were evaluated by the mean absolute error (MAE) and its 95% confidence interval (CI), concordance with mean bias and limits of agreement with Bland Altman plots, and consistency with intra-class correlation coefficient (ICC) and its 95% CI. ICCs calculations were based on a two-way random effect model and absolute agreement. Four levels of agreement were defined as follows: excellent, ICC ≥ 0.75; good, 0.60 ≤ ICC< 0.75; fair, 0.40 ≤ ICC < 0.6; poor, ICC < 0.4 [16].

Thresholds were defined as clinically relevant for CA, FPA and SS with subsequent confusion matrices generated to evaluate AI classification's performance. The number of true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN) findings were counted, and the corresponding sensitivity, specificity, accuracy, positive predictive value (PPV) and negative predictive value (NPV) were calculated and reported with their corresponding 95% CIs. All statistical analyses were performed using R software (version 3.6.3, R Foundation for Statistical Computing).

3. Results

3.1. General results

One hundred individual patients were included. There were 35 men and 65 women with a median age of 14 years (IQR: 13, 15.25 years; age range: 3–64 years) with nine adult patients. Seventy-four patients (74%) had a Cobb angle ≥ 10° and seven patients (7%) had severe scoliosis as defined as a Cobb angle ≥40° (Table 1). One coronal view was not analyzed by the AI solution, resulting in further AI *vs.* ground truth comparison over 100 patients in sagittal view and 99 patients in coronal view (Fig. 2).

3.2. Radiology resident analysis and senior radiologist's corrections of inaccurate measurements

One thousand measurements (10 parameters measured on 100 patients) were performed by the resident radiologist with mean measurement time of 6.5 ± 2.7 (SD) min (range: 3–14 min) per patient. Mean measurement time for the first 10 patients was 12.1 ± 1.6 (SD) min, and 4.2 ± 0.8 (SD) min for the last 10 patients. A total of 45 measurements (4.5%) were corrected by the senior musculoskeletal radiologist. Among them, 24 (52.2%) were sign errors on SVA (13 errors) and PT (11 errors). Fourteen corrections (30.4%) were related to Cobb angle with an MAE of 8.1° (95% CI: 4.5–11.7). Other discrepancies were related to kyphosis (six corrections; MAE = 13° (95% CI: 5.3–20.7), lordosis (two corrections; MAE = 11.5°) and CVA (one correction; MAE = 2 mm).

3.3. Comparison between AI solution and ground truth

For coronal parameters, MAE between AI and ground truth were 2.90° (95% CI: 2.41–3.39) for Cobb angle, 1.97 mm (95% CI: 1.66 –2.29) and 0.89 mm (95% CI: 0.76–1.03) for CVA and FPA, respectively. ICCs were 0.95 (95% CI: 0.92–0.97), 0.95 (95% CI: 0.93–0.97) and 0.97 (95% CI: 0.96–0.98) for Cobb angle, CVA and FPA, respectively (Table 2).

Bland Altman plots showed mean bias as −0.17° for CA, −0.02 mm and −0.25 mm for CVA and FPA, respectively. No proportional bias was demonstrated (Fig. 3).

Based on a threshold of 10° [17] for building the confusion matrix of CA measurements, PPV and NPV for scoliosis detection were 86% (95% CI: 80–92) and 94% (95% CI: 81–100), respectively. Analysis of severe scoliosis (≥ 40°) found PPV and NPV of 88% (95% CI: 60–100) and 100%, respectively. Considering a threshold of 10 mm as clinically

Table 1
Demographic characteristics of the study population and the subgroup of patients with a Cobb angle superior or equal to 10°

	Study population (100 patients)		Cobb angle ≥ 10° subgroup (74 patients)	
Age group	Age (year)	Sex ratio (M/F)	Age (year)	Sex ratio (M/F)
All patients	14 (13, 15.25) [3–64]	35/65	14 (14, 15) [3–57]	24/50
Pediatric patients (<18 years)	14 (13, 15) [3–17]	29/62	14 (13, 15) [3–17]	21/48
Adults patients (18 to 65 years)	26 (20–31) [18–64]	6/3	20 (18, 31) [18–57]	3/2

Ages are expressed as medians, with interquartile ranges (Q1, Q3) in parentheses and ranges in brackets. M indicates male. F indicates female

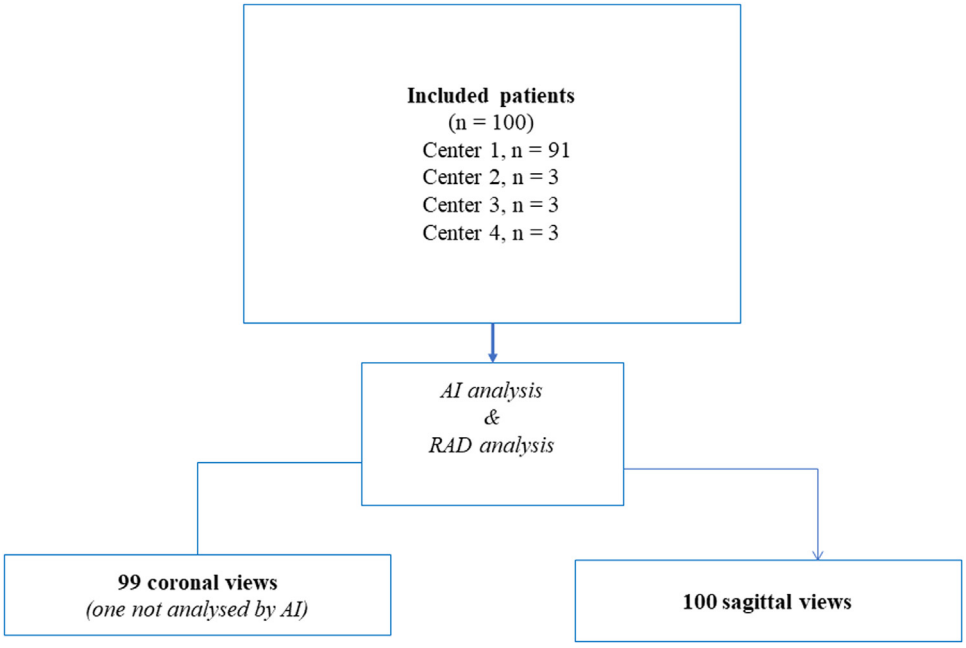

Fig. 2. Flowchart summarizes the study design.
AI indicates artificial intelligence solution; n indicates number; RAD indicates radiologist.

meaningful for detecting the presence *vs*. absence of FPA, PPV and NPV were 74% (95% CI: 53–94) and 99% (95% CI: 96–100), respectively (Fig. 4).

For sagittal parameters, MAE were 2.82° (95% CI: 2.39–3.26) for SS, 4.37° (95% CI: 3.45–5.29) for PI, 1.77° (95% CI: 1.20–2.34) for PT, 3.82° (95% CI: 3.17–4.47) for lordosis, 8.73° (95% CI: 6.58-–10.87) for kyphosis, 0.56° (95% CI: 0.35–0.76) for T9-SPI, and 2.70 mm (95% CI: 2.12–3.29) for SVA. The interrater reliability of sagittal spinal curvature characteristics was ≥ 0.85 for all sagittal parameters (SS, PI, PT, lordosis, SVA and T9-SPI) except for kyphosis for which reliability was only fair (ICC = 0.56; 95% CI: 0.40–0.69) (Table 2). Mean bias was −1.50 mm for SVA and between −1.84° and +1.17° for angle measurements except for kyphosis where the mean bias was −4.57° with limits of agreements of +13.5 and −11.2. No proportional bias was demonstrated (Fig. 5 and 6). Based on two thresholds of 35° and 45° for building the confusion matrix of SS measurements [18], the accuracy for correct classification by the AI solution was 81% (95% CI: 73–89) (Fig. 4).

3.4. Radiologist rating of AI performance

Patient-wise consensus visual evaluation of measurements reliability using a three-categories classification found AI measurements "reliable" in 72 (72%) patients, "moderately reliable" in eight (8%) patients and "unreliable" in 20 (20%) patients.

4. Discussion

Conventional uniplanar two-dimensional whole-spine radiographs are the most frequent radiological procedures prescribed to assess spinopelvic alignment, and are burdensome and time-consuming in the routine radiologists' and/or radiographer's workflow [19,20]. Accurate immediate and automatic measurements could alleviate workflow constraints and optimize time for healthcare professionals. In our study, we assessed the performance of a deep learning-based solution on an external study sample from four radiology sites and two different X-ray machine types of our institution proving the robust generalization capacity of the model.

In our study, the performance of the AI model compared favorably with those of previously published deep learning algorithms [21,22]. Recently, an AI solution demonstrated MAEs for Cobb angle ranging between 1.02° and 1.84°, and ICCs ranging between 0.83 and 0.99 [22]. Adolescent idiopathic scoliosis is a common spinal deformity in the pediatric population, affecting up to 2.2% of boys and 4.8% of girls [23]. We reported high classification accuracies for low Cobb angle (<10°) as well as for severe

Table 2
Artificial intelligence consistency compared to ground truth

Parameter	Mean absolute error (95% CI) AI *vs*. GT	Intraclass correlation coefficient (95% CI) AI *vs*. GT	Bias (95% CI) AI *vs*. GT
Cobb angle (°)	2.90 (2.41–3.39)	0.95 (0.92–0.97)	−0.17 (−0.95–0.60)
CVA (mm)	1.97 (1.66–2.29)	0.95 (0.93–0.97)	−0.02 (−0.52–0.48)
Frontal pelvic asymmetry (mm)	0.89 (0.76–1.03)	0.97 (0.96–0.98)	−0.25 (−0.46– −0.03)
Sacral slope (°)	2.82 (2.39–3.26)	0.92 (0.88–0.94)	−0.21 (−0.91–0.49)
Pelvic incidence (°)	4.37 (3.45–5.29)	0.85 (0.79–0.90)	1.17 (−0.07–2.40)
Pelvic tilt (°)	1.77 (1.20–2.34)	0.97 (0.96–0.98)	0.11 (−0.56–0.78)
Lordosis L1–S1 (°)	3.82 (3.17–4.47)	0.90 (0.84–0.94)	−1.84 (−2.76– −0.92)
Kyphosis T1–T12 (°)	8.73 (6.58–10.87)	0.56 (0.40–0.69)	−4.57 (−7.15– −1.99)
SVA (mm)	2.70 (2.12–3.29)	0.98 (0.97–0.99)	−1.50 (−2.23– −0.78)
T9–SPI (°)	0.56 (0.35–0.76)	0.96 (0.94–0.97)	−0.12 (−0.35–0.11)

AI indicates artificial intelligence; CVA indicates coronal vertical axis; GT indicates ground truth; SVA indicates sagittal vertical axis. T9-SPI indicates T9 spinopelvic inclination.

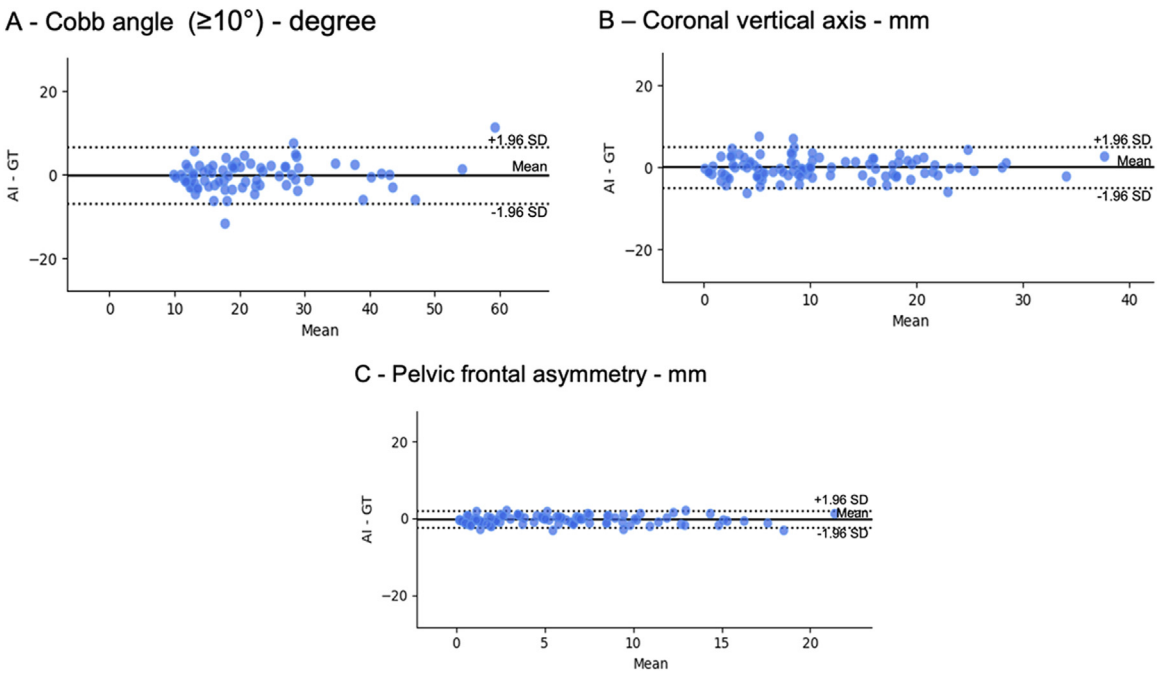

Fig. 3. Bland Altman plots assessing the agreement of coronal alignments between AI model predictions and ground truth (GT).

The Y-axis indicates the parameter difference between predicted results and the GT (*i.e.*, AI-GT). The X-axis represents the average degree of them (*i.e.*, (AI+GT)/2 or mean). The mean, and the upper and lower 95% Limits of agreement (LoAs) values as defined by 1.96 standard deviations from the mean of differences are reported in each subfigure. (A-C) Bland-Altman plots for Cobb angle superior or equal to 10°, coronal vertical axis (CVA), and frontal pelvic asymmetry (PFA), respectively.

scoliosis (≥ 40°), a result of importance for the reliability of the AI solution in screening and follow-up of these patients. More importantly, our study allowed us to analyze pelvic parameters in both coronal and sagittal views, demonstrating consistency and high classification accuracies on frontal pelvis asymmetry and sacral slope, the former being used in the Roussouly classification. Although sagittal spinopelvic alignment is of importance in adult and geriatric populations, it is of growing interest in the pediatric field [24].

To our knowledge, most prior studies on conventional whole-spine radiographs focused on either coronal or sagittal measurements [9,22,25,26] without assessment of the frontal pelvis asymmetry, although it is an important measurement to consider when reporting a coronal spinal angulation [27,28,29]. One recent algorithm combined coronal and sagittal views on biplanar radiographs using the medical machine EOS™ [9,30], although lacking pelvis parameters on the coronal and the sagittal views. A strength was therefore to combine spinal and pelvis measurements on coronal and sagittal views to provide a comprehensive analysis of spinopelvic alignment at the patient level. As a result, 72% of patients were assessed as reliably measured by AI (*i.e.*, without any measurement adjustments needed). AI tended to miscalculate the kyphosis angle with only fair reliability. Interestingly, this parameter was also most often erroneous in the sagittal plane when measured by the junior radiologist. In our study, the superimposition of the shoulders and clavicles, associated with the positioning of the hands on the clavicles

A - Cobb angle - 10° threshold

	AI <10°	AI ≥10°
GT <10°	17	8
GT ≥10°	1	73

C - Frontal pelvic asymmetry

	AI <10 mm	AI ≥10 mm
GT <10 mm	79	5
GT ≥10 mm	1	14

B - Cobb angle - 40° threshold

Cobb Angle	AI <40°	AI ≥40°
GT <40°	91	1
GT ≥40°	0	7

D - Sacral slope

SS (°)	AI <35°	AI [35°-45°]	AI >45°
GT <35°	34	4	0
GT [35°-45°]	4	33	6
GT >45°	0	5	14

Fig. 4. Confusion matrices demonstrating performance of the artificial intelligence solution

Thresholds for Cobb angle of (A) 10° and (B) 40° was defined and the confusion matrix computed. (C) Thresholds for frontal pelvic asymmetry (PFA) of 10 mm was defined. (D) For sacral slope, two thresholds of 35° and 45° was defined as per the Roussouly classification. Data are number of patients per category.

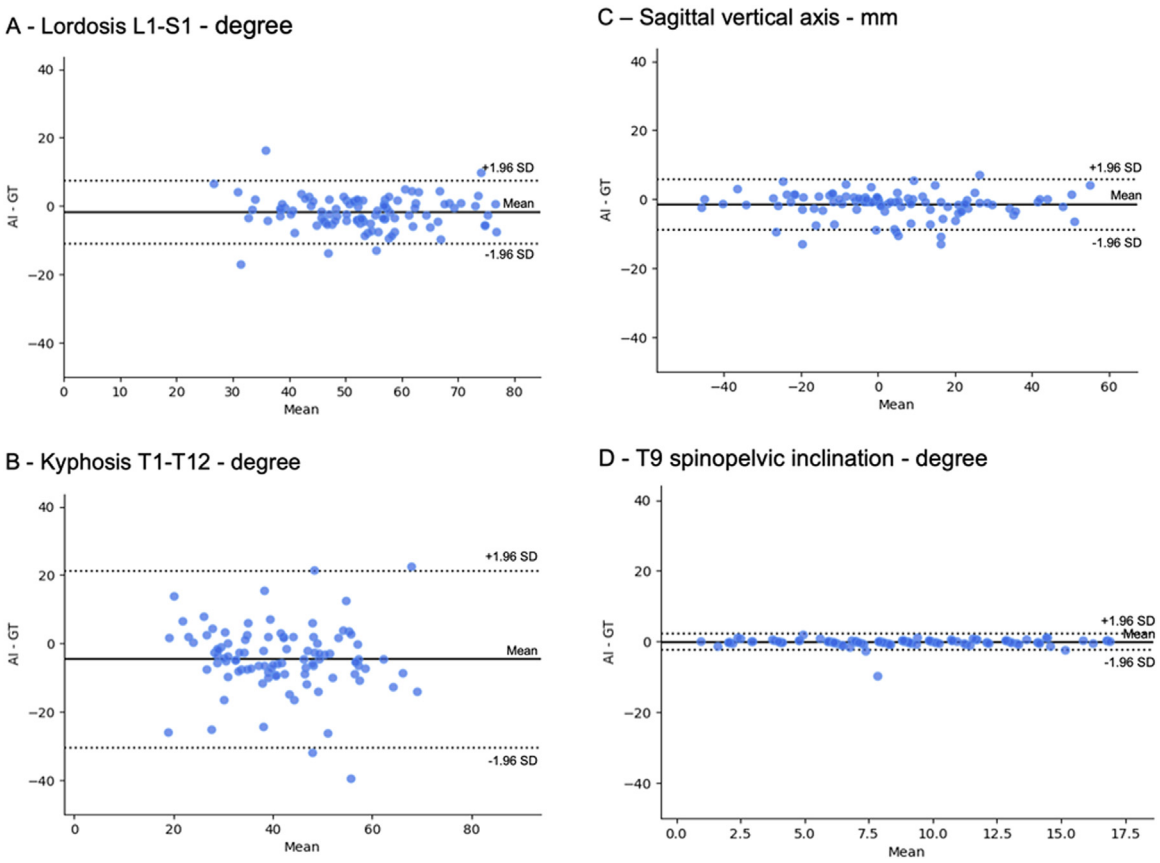

Fig. 5. Graphs show Bland Altman plots for assessing agreement of sagittal spinal parameters between the AI model and the ground truth (GT).

The Y-axis indicates the parameter difference between predicted results and the GT (*i.e.*, AI-GT). The X-axis represents the average degree of them (*i.e.*, (AI+GT)/2 or mean). The mean, and the upper and lower 95% Limits of agreement (LoAs) values as defined by 1.96 standard deviations from the mean of differences are reported in each subfigure. (A-D) Bland Altman plots for L1-S1 lordosis, T1-T12 kyphosis, SVA, and T9-SPI, respectively.

as per the standard position used in our institution, made difficult the visualization of the cervico-thoracic region and more particularly the upper endplate of T1. In such difficult situations, human reading is able to extrapolate the slope of the upper endplate of T1 from the all spinal alignment, as well as the above and below vertebra. The AI model is able of such behavior, but it remains limited. Altogether, increased measurement variability was responsible for errors of a few degrees, and a low reliability. However, and interestingly, we found that most errors by the junior radiologist were due to an error sign on SVA and pelvic tilt, which can be misleading for readers with low experience. This type of error is not possible with AI, anticipating the added value of such a solution to limit such inadvertent errors. Additionally, as our study design monitored the time for the junior radiologist to complete measurements for each patient without any learning period, we could estimate the heterogeneity of measuring times according to the reader's experience. Readers with low experience or activity of spinopelvic measurements per day or week will complete all measurements within 10 min or more on average. Experienced readers with a larger volume of cases per day will reach a measuring time limit of three to four minutes. The AI solution can provide a complete radiological output in less than two minutes. Altogether this is of importance for radiology routine where multiple sagittal and coronal measurements are done by radiographers or radiologists with limited time resources. We foresee the implementation of such an AI solution in the workflow without replacing radiological interpretation and/or radiographer expertise but rather speeding up patient turnaround time and standardizing a set of key parameters, similar to other applications [31]. Examinations not raising any measurement concern could be validated by the radiologist, while a smaller proportion of cases will be measured again on a limited number of inaccurate AI measurements. Under this hypothesis, AI could reduce turnaround time and alleviate workflow constraints while securing diagnosis accuracy, even more in the actual context of shortage of radiographers and the ever-increasing medical imaging workload.

Our study had limitations. First, the included population was mainly pediatric, limiting specific evaluations on sagittal misalignments that are encountered in adult and geriatric populations. Second, exclusion criteria were numerous and restricted the analysis on pre-operative patients without congenital anomalies, although patients with braces were not excluded as it is commonly done in scoliosis follow-up. Third, as per the study design, performance assessment of the junior radiologist was limited to a single reading and impact assessment could only be estimated.

In conclusion, a combined coronal and sagittal automatic deep learning-based evaluation of spinopelvic alignment parameters was proved consistent compared to a senior musculoskeletal radiologist, as well as potentially of benefit in reducing workload in routine implementation. Further studies with more case materials are needed to fully determine the capabilities of this AI software in the real life.

Human rights

The authors declare that the work described has been performed in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans.

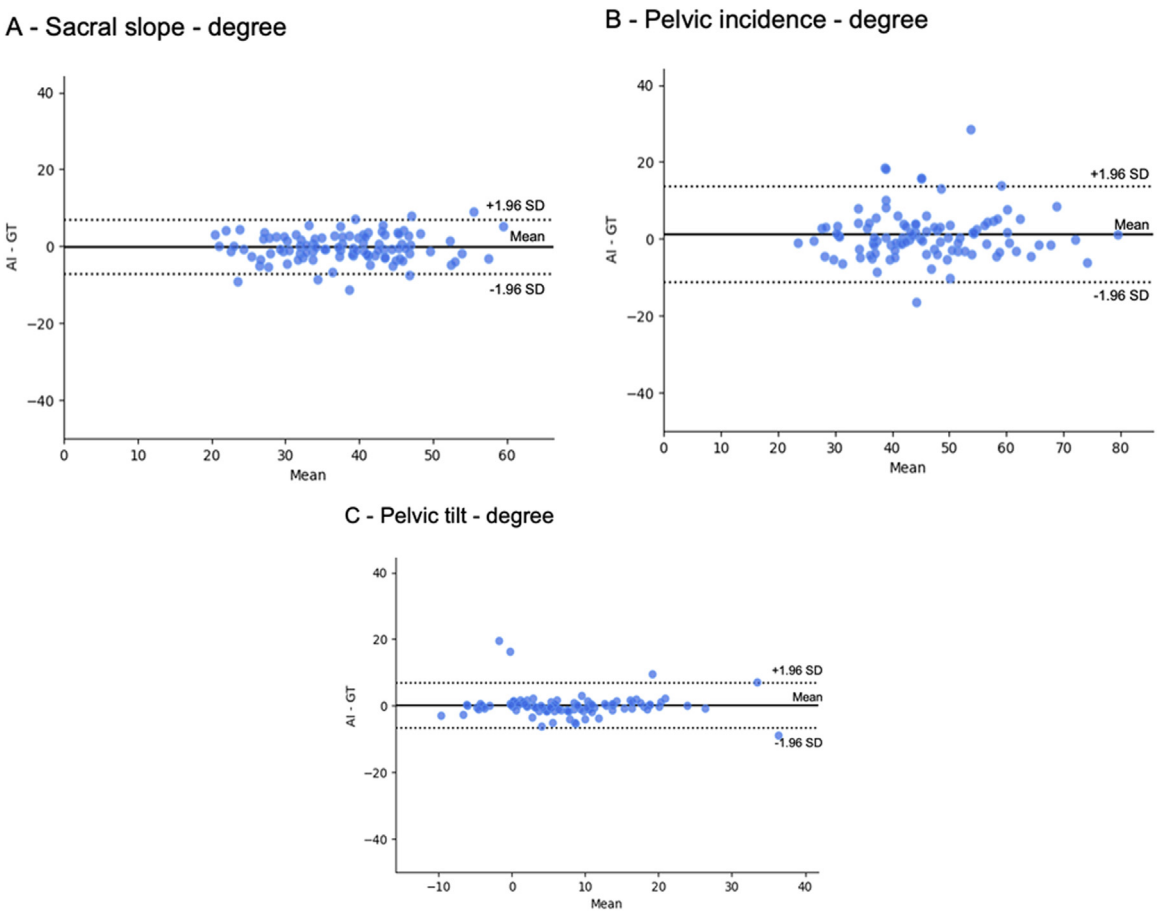

Fig. 6. Graph show Bland Altman plots for assessing agreement of sagittal pelvic parameters between the artificial intelligence model and the ground truth (GT).
The Y-axis indicates the parameter difference between predicted results and the GT (*i.e.*, AI-GT). The X-axis represents the average degree of them (*i.e.*, (AI+GT)/2 or mean). The mean, and the upper and lower 95% limits of agreement (LoAs) values as defined by 1.96 standard deviations from the mean of differences are reported in each subfigure. (A-C) Bland Altman plots for SS, PI, and PT, respectively.

Informed consent and patient details

The institutional review board of our institution approved this study (CRM-2208−299). The authors declare that this report does not contain any personal information that could lead to the identification

Funding information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. This research received nonfinancial support by Milvue, which provided the AI model free of charge

Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

Data sharing statement

Data generated by the authors are available from the corresponding author upon request. As part of a commercial software, the development dataset, the proprietary code and the weights of the AI model are not publicly available. Clinical datasets are not publicly available, proprietary to the author's institution organization.

Declaration of Competing Interest

A.P. is the cofounder and chairman of Milvue. M.B and C.R. are employees at Milvue. M.Z., P.C and D.G. has no conflict of interest to disclose in relation with this article.

CRediT authorship contribution statement

Mohamed Zerouali: Writing – original draft, Investigation. **Alexandre Parpaleix:** Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Software. **Mansour Benbakoura:** Data curation, Software. **Caroline Rigault:** Data curation, Validation. **Pierre Champsaur:** Conceptualization, Supervision. **Daphné Guenoun:** Methodology, Investigation, Writing – review & editing, Supervision.

References

[1] Rogala EJ, Drummond DS, Gurr J. Scoliosis: incidence and natural history. A prospective epidemiological study. J Bone Joint Surg Am 1978;60:173–6.
[2] Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. Lancet 2018;391:2356–67.
[3] Dietrich TJ, Pfirrmann CWA, Schwab A, Pankalla K, Buck FM. Comparison of radiation dose, workflow, patient comfort and financial break-even of standard digital radiography and a novel biplanar low-dose X-ray system for upright full-length lower limb and whole spine radiography. Skeletal Radiol 2013;42:959–67.
[4] Legaye J, Duval-Beaupère G, Marty C, Hecquet J. Pelvic incidence: a fundamental pelvic parameter for three-dimensional regulation of spinal sagittal curves. Eur Spine J 1998;7:99–103.
[5] Le Huec JC, Thompson W, Mohsinaly Y, Barrey C, Faundez A. Sagittal balance of the spine. Eur Spine J 2019;28:1889–905.

[6] Cobb JR. Outlines for the study of scoliosis measurements from spinal roentgenograms. Phys Ther 1948;59:764–5.
[7] Carman DL, Browne RH, Birch JG. Measurement of scoliosis and kyphosis radiographs. Intraobserver and interobserver variation. J Bone Joint Surg Am 1990;72:328–33.
[8] Shrader MW, Andrisevic EM, Belthur MV, White GR, Boan C, Wood W. Inter- and intraobserver reliability of pelvic obliquity measurement methods in patients with cerebral palsy. Spine Deform 2018;6:257–62.
[9] Galbusera F, Niemeyer F, Wilke HJ, Bassani T, Casaroli G, Anania C, et al. Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach. Eur Spine J 2019;28:951–60.
[10] Wu H, Bailey C, Rasoulinejad P, Li S. Automated comprehensive adolescent idiopathic scoliosis assessment using mvc-net. Med Image Anal 2018;48:1–11.
[11] Wang L, Xu Q, Leung S, Chung J, Chen B, Li S. Accurate automated Cobb angles estimation using multi-view extrapolation net. Med Image Anal 2019;58:101542.
[12] Lacroix M, Aouad T, Feydy J, Biau D, Larousserie F, Fournier L, Feydy A. Artificial intelligence in musculoskeletal oncology imaging: a critical review of current applications. Diagn Interv Imaging 2023;104:18–23.
[13] Gong B, Soyer P, McInnes MDF, Patlas MN. Elements of a good radiology artificial intelligence paper. Can Assoc Radiol J 2022. doi: 10.1177/08465371221101195.
[14] Yu H., Chen C., Du X., Li Y., Rashwan A., Hou L., Jin P., Yang F., Liu F., Kim J., Li J. Tensorflow model garden. https://github.com/tensorflow/models. Published 2020. Accessed September 20, 2022
[15] Barat M, Jannot AS, Dohan A, Soyer P. How to report and compare quantitative variables in a radiology article. Diagn Interv Imaging 2022;103:571–3.
[16] Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6:284–90.
[17] Kane WJ. Scoliosis prevalence: a call for a statement of terms. Clin Orthop Relat Res 1977;126:43–6.
[18] Roussouly P, Gollogly S, Berthonnaud E, Dimnet J. Classification of the normal variation in the sagittal alignment of the human lumbar spine and pelvis in the standing position. Spine 2005;30:346–53.
[19] Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Kim J. Long radiology workdays reduce detection and accommodation accuracy. J Am Coll Radiol 2010;7:698–704.
[20] Pomero V, Mitton D, Laporte S, de Guise JA, Skalli W. Fast accurate stereoradiographic 3D-reconstruction of the spine using a combined geometric and statistic model. Clin Biomech 2004;19:240–7.
[21] Zhang J, Li H, Lv L, Zhang Y. Computer-aided Cobb measurement based on automatic detection of vertebral slopes using deep neural network. Int J Biomed Imaging 2017;2017:1–6.
[22] Yeh YC, Weng CH, Huang YJ, Fu CJ, Tsai TT, Yeh CY. Deep learning approach for automatic landmark detection and alignment analysis in whole-spine lateral radiographs. Sci Rep 2021;11:7618.
[23] Fong DYT, Cheung KMC, Wong YW, Wan YY, Lee CF, Lam TP, et al. A population-based cohort study of 394,401 children followed for 10 years exhibits sustained effectiveness of scoliosis screening. Spine J 2015;15:825–33.
[24] El-Hawary R, Sturm PF, Cahill PJ, Samdani AF, Vitale MG, Gabos PG, et al. Sagittal spinopelvic parameters of young children with scoliosis. Spine Deform 2013;1:343–7.
[25] Zhang J, Lou E, Le LH, Hill DL, Raso JV, Wang Y. Automatic cobb measurement of scoliosis based on fuzzy hough transform with vertebral shape prior. J Digit Imaging 2009;22:463–72.
[26] Sun H, Zhen X, Bailey C, Rasoulinejad P, Yin Y, Li S. Direct estimation of spinal cobb angles by structured multi-output regression. International Conference on Information Processing in Medical Imaging. Springer; 2017. p. 529–40.
[27] Winter RB, Pinto WC. Pelvic obliquity: its causes and its treatment. Spine 1986;11:225–34.
[28] Papaioannou T, Stokes I, Kenwright J. Scoliosis associated with limb-length inequality. J Bone Joint Surg Am 1982;64:59–62.
[29] Giles LG, Taylor JR. Lumbar spine structural changes associated with leg length inequality. Spine 1982;7:159–62.
[30] Meng N, Cheung JPY, Wong KYK, Dokos S, Li S, Choy RW, et al. An artificial intelligence powered platform for auto-analyses of spine alignment irrespective of image quality with prospective validation. EClinicalMedicine 2022;43:101252.
[31] Soyer P, Fishman EK, Rowe SP, Patlas MN, Chassagnon G. Does artificial intelligence surpass the radiologist? Diagn Interv Imaging 2022;103:445–7.