Systematic Review | Open Access

Artificial Intelligence Surgery

Su *et al. Art Int Surg.* 2026;6:638-64 **DOI:**10.20517/ais.2026.59

The application of machine learning for the diagnosis and assessment of adolescent idiopathic scoliosis: a systematic review

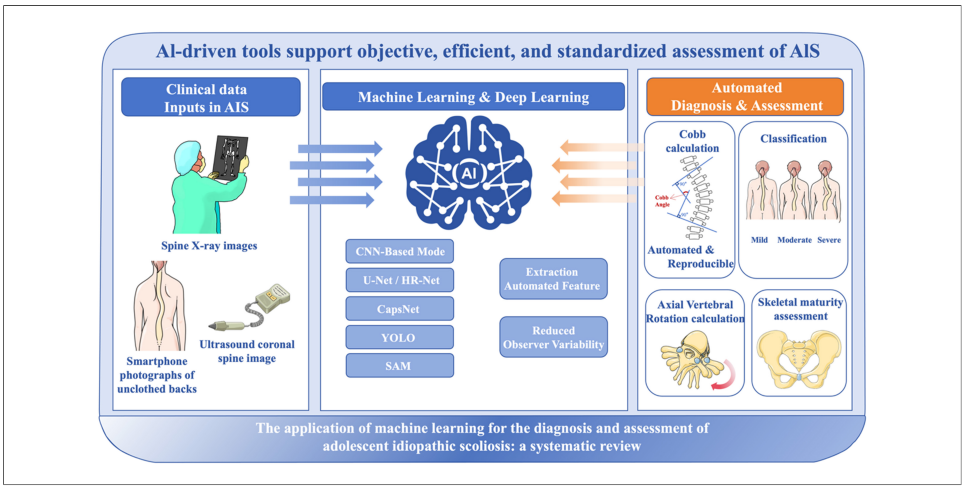
Jingxi Su[1], Chang Liu[2], Guilin Chen[1], Terry Jianguo Zhang[1], Nan Wu[1]

Keywords:
Adolescent idiopathic scoliosis (AIS), artificial intelligence, machine learning, diagnosis and assessment, systematic review, CLAIM

Citation: Su J, Liu C, Chen G, Zhang TJ, Wu N. The application of machine learning for the diagnosis and assessment of adolescent idiopathic scoliosis: a systematic review. *Art Int Surg.* 2026;6:638-64. https://dx.doi.org/10.20517/ais.2026.59

Received: 3 Jul 2026
First Decision: 18 Sep 2026
Revised: 26 Oct 2026
Accepted: 6 Nov 2026
Published: 16 Nov 2026

Academic Editor:
Peter Passias
Copy Editor:
Tong Wang
Production Editor:
Tong Wang



Abstract

Aim: This systematic review evaluated the applications, methodologies, and reporting quality of artificial intelligence (AI) models for the diagnosis and assessment of adolescent idiopathic scoliosis (AIS).

Methods: PubMed, Embase, and IEEE Xplore were searched for studies published from January 2015 to May 2026, with additional searches of Medical Image Computing and Computer Assisted Intervention (MICCAI), IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), and International Conference on Machine Learning (ICML) proceedings (2020-2026). Studies applying AI or machine learning (ML) for AIS diagnosis or assessment were included. Methodological quality was assessed using a 12-dimensional framework based on the 2024 Checklist for Artificial Intelligence in Medical Imaging (CLAIM).

Check for updates

[1]Department of Orthopedic Surgery, State Key Laboratory of Complex Severe and Rare Diseases, Beijing Key Laboratory of Big Data Innovation and Application for Skeletal Health Medical Care, Key Laboratory of Big Data for Spinal Deformities, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China.
[2]Department of Orthopedic Surgery, Yantai Affiliated Hospital of Binzhou Medical University, Yantai 264100, Shandong, China.

Correspondence to: Prof. Nan Wu, Department of Orthopedic Surgery, State Key Laboratory of Complex Severe and Rare Diseases, Beijing Key Laboratory of Big Data Innovation and Application for Skeletal Health Medical Care, Key Laboratory of Big Data for Spinal Deformities, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China. E-mail: dr.wunan@pumch.cn

Results: Of the 8,579 records initially identified, 45 studies were included. Automated Cobb angle measurement was the most common application, reported in 32 studies. The mean absolute error for Cobb angle measurement ranged from 1.07° to 5.31°. Other applications included binary scoliosis classification (8 studies), curve classification and severity grading (12 studies), vertebral rotation assessment (2 studies), and skeletal maturity assessment (2 studies). Eleven studies (24.4%) used non-radiographic inputs encompassing six modalities. Quality assessment showed that data partitioning (93.3%) and comparisons with competing models (93.3%) were well reported, whereas external validation (26.7%), code availability (13.3%), and explainable AI (35.6%) remained uncommon.

Conclusion: AI demonstrates technical accuracy in Cobb angle estimation comparable to manual measurement variability. However, limited external validation, prospective evaluation, and transparent reporting remain barriers to clinical translation. Future studies should emphasize characterization of clinical practice variation, prospective validation, and standardized reporting to facilitate clinical implementation.

INTRODUCTION

Adolescent idiopathic scoliosis (AIS) is a structural spinal deformity that develops during adolescence. It is characterized by lateral curvature of the spine in the coronal plane, often accompanied by axial vertebral rotation (AVR), resulting in a characteristic three-dimensional (3D) deformity[1]. Clinically, AIS may involve one or more spinal segments and present with S- or C-shaped curves. Although disease progression is relatively slow in most patients, some cases may progressively worsen, leading to abnormal trunk appearance and potentially adverse effects on cardiopulmonary function and quality of life. The etiology of AIS remains incompletely understood, and its underlying pathogenesis requires further clarification.

Given the potential for AIS progression, accurate assessment is essential for disease monitoring and treatment decision-making. In clinical practice, AIS is primarily evaluated using radiographic and non-radiographic methods. Radiographic assessment relies on standing full-spine radiographs as the reference standard, with the Cobb angle used to quantify the major and compensatory curves and thereby assess scoliosis severity, risk of progression, and indications for treatment[2]. In selected cases, lateral or side-bending radiographs are additionally obtained to comprehensively assess sagittal spinal parameters and curve flexibility. However, imaging-based measurements are influenced by examiner experience and differences in measurement technique, resulting in appreciable inter- and intra-observer variability.

In contrast, non-radiographic assessment primarily relies on physical examination and clinical observation, including evaluations of shoulder-height asymmetry, waistline and pelvic imbalance, the presence of a rib hump, and asymmetry of the ribs or breasts[3]. The Adams forward bend test is widely used for preliminary screening of adolescent scoliosis because it accentuates axial trunk rotation and makes surface asymmetry more apparent. During this examination, a scoliometer is often used to semi-quantitatively measure the angle of trunk rotation (ATR), thereby assisting in the detection of scoliosis and estimation of its potential severity. These non-radiographic assessment methods are easy to perform and avoid radiation exposure, making them suitable for large-scale population screening and clinical follow-up[4]. However, the results of these methods remain highly dependent on examiner experience and consistency. In addition, the derived indicators are largely indirect or semi-quantitative and therefore cannot fully characterize the 3D structural features of spinal deformity. These limitations are particularly relevant to early diagnosis, severity grading, and progression-risk prediction. In these settings, surface-based indicators may fail to capture the full extent of the underlying spinal deformity.

In recent years, rapid advances in computer vision and deep learning (DL) have highlighted the considerable potential of artificial intelligence (AI) for the diagnosis and assessment of AIS[5]. AI-based approaches can

integrate and analyze multiple imaging and clinical parameters, providing objective and efficient support for early detection, severity assessment, and treatment planning. As the volume of AI research in AIS has increased, there is a growing need for a structured synthesis that extends beyond simply cataloging individual models. Accordingly, this review organizes the available literature according to clinical tasks rather than model architectures, evaluates methodological quality against the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) 2024 reporting framework, and examines the emerging transition toward non-radiographic assessment modalities. The aim is to clarify not only the current capabilities of AI within the AIS clinical workflow but also the critical gaps in evidence quality and clinical readiness that remain.

METHODS

Search strategy

The literature search covered studies published between January 1, 2015, and May 31, 2026. The primary search was initially conducted on June 10, 2025, updated on September 22, 2025, and further updated in May 2026 to capture the most recent literature in this rapidly evolving field.

This systematic review aimed to characterize the current machine learning (ML) models used for the diagnosis and assessment of AIS, identify research trends since 2015, summarize their applications across different clinical tasks, and evaluate the methodological characteristics and reporting quality of the available literature.

PubMed, Embase, and IEEE Xplore were searched as the primary literature databases. The search terms were ("AIS" OR "Scoliosis" OR "Adolescent Idiopathic Scoliosis") AND ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Neural Network" OR "Computational Intelligence" OR "Computational Learning" OR "Convolutional Network" OR "Machine Prediction" OR "Machine Intelligence") and were required to appear in the title or abstract.

Because important methodological advances in computer vision and medical image analysis are frequently published in peer-reviewed conference proceedings rather than indexed journals, we performed a prespecified supplementary search of conference proceedings to complement the primary database search. This search covered peer-reviewed full-paper proceedings from Medical Image Computing and Computer Assisted Intervention (MICCAI), IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), and International Conference on Machine Learning (ICML) published between 2020 and 2026. Conference papers were identified through official conference repositories, the Computer Vision Foundation (CVF) Open Access archive, and the Digital Bibliography & Library Project (DBLP) bibliography database and were screened using the terms scoliosis, spine, spinal, vertebra, Cobb angle, and spinal curvature. We also cross-referenced publications associated with the Accurate Automated Spinal Curvature Estimation (AASCE) MICCAI 2019 challenge because of its role in establishing a benchmark dataset in this field. Only full-length papers that had undergone formal peer review were eligible; preprints and extended abstracts were excluded.

The primary search period began in 2015 to provide a comprehensive overview of the field since the emergence of DL in medical image analysis. In contrast, the supplementary conference search was restricted to 2020 onward to capture more recent methodological advances while minimizing duplication of earlier architectures already represented in the journal literature.

Additionally, the reference lists of eligible articles were manually screened to identify relevant studies not captured by the database searches. The detailed search strings and search dates are provided in Supplementary Table 1.

Eligibility criteria

Two reviewers independently screened and selected eligible studies, with disagreements resolved by consensus.

Studies were eligible if they were published in English, addressed the diagnosis or assessment of AIS, and focused on nongenetic forms of scoliosis. Exclusion criteria included letters to the editor, non-peer-reviewed conference abstracts and extended abstracts, book chapters, non-English publications, review articles, meta-analyses, systematic reviews, conference reviews, scoping reviews, tombstones, and notes. Full-length, peer-reviewed papers published in the proceedings of the aforementioned computer science conferences were considered eligible and were therefore not excluded under the conference-related criteria.

Data analysis

Data were systematically extracted from each included study into two standardized evidence tables. Table 1 summarizes the clinical and technical characteristics of each study, including publication details (author, year, and country), the underlying AI model and its specific application, input data modality, measurement standard, model-derived features or parameters, and reported performance metrics. Section "Methodological quality and reporting practices" presents the methodological and reporting quality of the included studies across the 12 predefined dimensions described in the following subsection.

Data extraction was performed independently by 2 reviewers using a predefined form, and discrepancies were resolved through consensus. The synthesis focused on the analytical applications of ML for the diagnosis and assessment of AIS and deliberately excluded applications involving surgical outcome prediction or longitudinal progression modeling.

Evidence quality assessment

To evaluate the methodological rigor and reporting transparency of the included studies, we developed a 12-dimensional quality assessment framework, with the corresponding results presented in Section "Methodological quality and reporting practices". The 12 dimensions were selected to capture aspects of study design, data handling, and model evaluation that most directly influence the reliability and reproducibility of AI research in AIS. These dimensions comprised study design, data partitioning (DP), dataset source, dataset size, dataset scoliosis ratio (DSR), Cobb angle range reported (CAR), dataset quality control (QC), codebase open source (COS), base model improvement (BMI), explainable AI used (XAI), testing on external data (TED), and competitive models comparison (CMC).

Each dimension was evaluated according to whether the relevant information was fully reported, absent, or partially reported in the original study. We deliberately did not combine these dimensions into a single weighted quality score because doing so would require unsupported assumptions regarding their relative importance and additivity. Such an arbitrary composite score could obscure the dimension-specific strengths and deficiencies that Section "Data sources and modalities" was designed to highlight. Two reviewers independently assessed all 45 studies, and disagreements were resolved through discussion. Section "Methodological quality and reporting practices" provides the rationale for selecting these dimensions, describes their alignment with established reporting standards, and analyzes cross-study patterns.

RESULTS

Overview of included studies

Study characteristics and distribution

The final search identified 45 eligible studies published between 2020 and 2026 [Figure 1]. Publication activity was concentrated in the most recent 3 years, with 24 of the 45 studies (53.3%) published from 2024 onward [Figure 2]. This increase reflects both the maturation of DL tools for medical imaging and growing

Table 1. Key characteristics and performance of AI models for AIS diagnosis and assessment in included studies

Ref.	Country	Base AI model	AI application	Input params	Model performance
Tınas and Esmer[6], 2024	Turkey	YOLO	Diagnose (0/1)	Spine X-ray images (AP)	Precision = 0.985, Recall = 0.982, F1 Score = 0.983
Goral and Kose[7], 2022	Turkey	CapsNet	Cobb calculation	Spine X-ray images (AP)	Accuracy = 0.98, Macro Average Precision = 0.98, Macro Average Recall = 0.98 Macro F1 Score = 0.94, Micro F1 Score = 0.98
Wang *et al.*[8], 2024	China	U-Net	Cobb calculation	Spine X-ray images (AP)	SMAPE = 11.67%
Peng *et al.*[9], 2024	China	U-Net	Cobb calculation	Spine X-ray images (AP)	(LDEP method) MAE = 1.6°, Pearson correlation = 0.983
Kim *et al.*[10], 2020	Korea	U-Net	Cobb calculation	Spine X-ray images (AP)	CMAE = 3.51° ± 3.89° (Internal), 4.11° ± 3.98° (External) SMAPE = 7.84% (Internal), 6.44% (External)
Maaliw *et al.*[11], 2022	Philippines	U-Net	Cobb calculation	Spine X-ray images (AP)	(Ground truth *vs.* Predicted Cobb angle): MAPE = 8.18%, Accuracy = 91.81%
Liu *et al.*[12], 2021	China	Faster R-CNN	Cobb calculation	Ultrasound coronal spine image	Automatic *vs.* US manual: MAD ± SD = 4.3° ± 3.3° (R = 0.89) Automatic *vs.* Cobb: MAD ± SD = 5.5° ± 4.6° (R = 0.79)
Zou *et al.*[13], 2023	China	U-Net	Cobb calculation	Spine X-ray images (AP)	SMAPE = 6.42%, CMAE = 3.01°
Shang *et al.*[14], 2023	China	SVM, XGBoost	Diagnose (0/1)	Wrist pulse signal	Accuracy = 0.7856 (SVM), 0.8046 (XGBoost) F1-score = 0.8238 (SVM), 0.8540 (XGBoost)
Liang *et al.*[15], 2023	China	YOLO	Cobb calculation	Spine X-ray images (AP)	SMAPE = 11.74% ± 12.05% (Internal), 5.90% ± 6.12% (External) CMAE = 2.49° ± 7.38° (Internal), 3.61° ± 3.04° (External)
Qiu *et al.*[16], 2024	China	ResNet, U-Net	Cobb calculation	Spine X-ray images (AP)	SMAPE = 7.28%, CMAE = 2.26°, MAE = 3.18°
Dagli *et al.*[17], 2025	USA	ResNet	Cobb calculation	Spine X-ray images (AP)	MAE = 1.78° ± 1.61° (Internal), 2.22° ± 2.04° (External) SMAPE = 3.67% (Internal), 4.29% (External) ICC = 0.99 (Internal), 0.98 (External)
Zhu *et al.*[18], 2023	China	U-Net	Cobb calculation	Spine X-ray images (AP)	SMAPE = 11.82%, ICC = 0.902, 0.915, MAD = 3.08°, 2.91°
Zhao *et al.*[19], 2024	China	HRNet	AVR calculation	Spine X-ray images (AP)	AI *vs.* Expert1 = 3.56°, 3.99°, 4.29°, AI *vs.* Expert2 = 3.69°, 3.97°, 4.31°
Zhang *et al.*[20], 2021	China	U-Net	Cobb calculation	Spine X-ray images (AP)	SMAPE = 25.99%
Logithasan *et al.*[21], 2022	Canada	U-Net	AVR calculation	Spine X-ray images (AP)	MAD = 2.9° ± 2.2°, clinical acceptance rate = 88.1%
Magnide *et al.*[22], 2021	Canada	ResNet, SVM	Skeletal maturity assessment	Spine X-ray images (AP)	Accuracy = 0.84 Humeral heads classification: Accuracy = 0.78 Femoral heads classification: Accuracy = 0.80
Zhao *et al.*[23], 2025	China	CNN, Swin Transformer	Skeletal maturity assessment	Pelvis radiographs	Accuracy = 81.53%, Precision = 80.63%, Recall = 80.80%, F1-score = 80.58%
Zhang *et al.*[24], 2023	China	CNN with attention mechanisms	Classification	Smartphone photographs of unclothed backs (standing)	[Training on Internal Data, Validation on External Data] Severity classification: AUC = 0.827 Curve type classification: AUC = 0.777 (T), 0.760 (TL/L), 0.860 (mix)

Xie *et al.*[25], 2025	China	ResNet	Diagnose (0/1), Cobb calculation, Classification	Spine X-ray images (AP, LAT)	Cobb calculation: MAD = 2.15° ± 1.69° (Cobb calculation), 2.72° ± 1.99° (Sagittal parameters) ICC = 0.985 (Cobb calculation), 0.927 (Sagittal parameters) Diagnose: Accuracy = 97.2%, Sensitivity = 0.968, Specificity = 0.976 Severity classification: Accuracy = 94.5% Sagittal abnormalities classification: Accuracy > 91.8%, Sensitivity > 93.6%, Specificity > 81.5%
Li *et al.*[26], 2024	China	ResNet, Transformer	Cobb calculation	Spine X-ray images (AP)	MAE = 2.00° ± 1.43° (photography), 2.08° ± 1.38° (photo upload) ICC > 0.97 (photography & photo upload)
Wu *et al.*[27], 2022	China	HRNet	Cobb calculation	Spine X-ray images (AP)	Accuracy = 89.36%, Recall = 97.35%, SMAPE = 5.99%, MAE = 1.07°
Maaliw *et al.*[28], 2023	Philippines	U-Net	Cobb calculation	Spine X-ray images (AP)	MAD = 2.86°, MAPE = 3.86% (accuracy = 96.13%)
Wong *et al.*[29], 2024	Canada	U-Net	Cobb calculation	Spine X-ray images (AP)	Detection rate = 98%, CMAE = 2.8° ± 2.8°, Within clinical acceptance (≤ 5°) = 91%
Liu *et al.*[30], 2021	China	Deeplab	Cobb calculation	Spine X-ray images (AP)	MAE = 2.036° (Expert1); 2.818° (Expert2) ME = -0.106° (Expert1); -0.262° (Expert2)
Du *et al.*[31], 2025	China	CNN with attention mechanisms, XGBoost	Cobb calculation, Classification	Spine X-ray images (AP, R & L-Bending)	Cobb calculation: SMAPE = 7.19%, MAE = 3.06° (internal validation) SMAPE = 8.44%, MAE = 5.31° (external validation) Lenke classification: Accuracy = 93%
Lu *et al.*[32], 2024	China	HRNet	Cobb calculation, Classification	Spine X-ray images (AP)	[Internal Set for Tuning, External Set for Performance Validation] Cobb calculation: MAE = 2.15° (ALL), MAE = 1.89° (Major curve) Classification consistency for major curve magnitude: kappa = 0.77, ICC = 0.97(ALL)
Li *et al.*[33], 2024	China	YOLO, SAM	Diagnose (0/1)	Smartphone photographs of unclothed backs (standing & bending)	Curvature type classification (Standing): Accuracy = 0.887 Severity classification: Accuracy = 0.725 Fusion method classification accuracy = 0.793 (internal validation), 0.702 (external validation) ATR calculation: MRE = 6.2%-10.3%
Zou *et al.*[34], 2023	China	HRNet, U-Net	Cobb calculation	Spine X-ray images (AP)	SMAPE = 13.9%, CMAE = 2.56°
Gadu and Potala[35], 2026	India	U-Net, Transformer	Cobb calculation	Spine X-ray images (AP)	Spineweb dataset: MAE = 1.45°, MAPE = 2.30% AASCE Challenge dataset: MAE = 3.01°, SMAPE = 6.21%
Rahmaniar *et al.*[36], 2024	Japan	ResNet	Diagnose (0/1), Cobb calculation, Classification	Spine X-ray images (AP)	(Curve type classification, Lenke classification, Severity classification): Precision = 0.966, 0.914, 0.963, Recall = 0.952, 0.935, 0.957 Cobb calculation: SMAPE = 5.27%, median accuracy = 94.15%, median difference = 0.5°, standard deviation = 1.375%
Zhang *et al.*[37], 2022	China	Faster R-CNN	Cobb calculation	Spine X-ray images (AP, LAT)	CMAE = 3.52° ± 1.07° (AP), 4.05° ± 1.36° (LAT) SMAPE = 13.71% ± 4.76% (AP), 12.60% ± 4.12% (LAT)

Yang *et al.*[38], 2025	China	HRNet	Diagnose (0/1), Cobb calculation	Spine X-ray images (AP) 2-D slices of sagittal view CT scans	SMAPE= 10.24% ± 0.29%
Yao *et al.*[39], 2022	China	CNN, Transformer	Cobb calculation	Spine X-ray images (AP)	SMAPE = 8.26%, MAD = 3.76°
Zhao *et al.*[40], 2022	China	U-Net	Cobb calculation	Spine X-ray images (AP)	CMAE = 3.38°, SMAPE = 15.35%
Feng *et al.*[41], 2025	China	CNN	Cobb calculation, Classification	3D back surface scan (Adam's forward bend)	Cobb angle: MAE = 2.9° ± 0.3°, RMSE = 6.9° ± 0.7° Curve type (thoracic, lumbar, mixed, none): Sensitivity = 73%, 89%, 89%, 75%, Specificity = 92%, 81%, 97%, 94% Curve direction (right, left, none): Sensitivity = 80%, 75%, 80%, Specificity = 96%, 97%, 84%
González-Ruiz *et al.*[42], 2026	Germany	Linear Regression Model (LRM)	Cobb calculation	3D back surface scan (standing)	MAE = 3.97° (95%CI: 2.87°-5.09°), RMSE = 4.70° (95%CI: 3.43°-5.77°)
Chen *et al.*[43], 2026	China	HRNet	Cobb calculation, Classification	Spine X-ray images (AP)	MAE = 4.03° ± 3.12° (T), 3.75° ± 2.94° (TL/L) Severity Level (internal dataset, external dataset): Accuracy = 94.17%, 87.13%-94.17% Curve Type (internal dataset, external dataset): Accuracy = 98.48%, 88.24%-98.73%
Stecher *et al.*[44], 2025	Germany	SVM	Classification	Torsobarographic pressure images	k = 0.91, Accuracy = 97.0%, Sensitivity = 92.0%, Specificity = 98.0%
Vishwakarma *et al.*[45], 2026	India	CNN	Classification	gait video	Accuracy = 89.05%, Sensitivity = 89.22%, Specificity = 92.48%
Lin *et al.*[46], 2021	China	ResNet, EfficientNet	Cobb calculation	Spine X-ray images (AP)	SMAPE=7.32%, MAE=3.73°
Huo *et al.*[47], 2021	China	ResNet	Cobb calculation	Spine X-ray images (AP)	SMAPE=20.51% (128 test);
Zhou *et al.*[48], 2024	China	ResNet	Diagnose (0/1), Classification	Gait silhouette video	Accuracy = 82.0% (ScoNet-MT), 51.3% (ScoNet) Sensitivity = 99.0% (ScoNet-MT), 100.0% (ScoNet) Specificity = 76.5% (ScoNet-MT), 33.2% (ScoNet)
Zhou *et al.*[49], 2026	China	ResNet	Classification	2D pose keypoints (from Gait video)	Accuracy = 86.0%, Macro-averaged Precision = 84.1% Macro-averaged Recall = 79.2%, Macro-averaged F1 = 80.8
Li *et al.*[50], 2026	USA	ResNet	Diagnose (0/1), Classification	Gait video + text guidance (expert + GPT-4o generated)	Accuracy = 89.9%, Sensitivity = 99.5%, Specificity = 86.4%, F1 = 90.2

AI: Artificial intelligence; AIS: adolescent idiopathic scoliosis; AUC: area under the ROC curve; ICC: intraclass correlation coefficient; AP: average precision; MRE: mean radial error/mean relative error; MAD: mean absolute difference; SD: standard deviation; ME: mean error; MAE: mean absolute error; CMAE: circular mean absolute error; MAPE: mean absolute percentage error; SMAPE: symmetric mean absolute percentage error; R: Pearson correlation coefficient; LDEP: line detection and endpoint projection; SVM: support vector machine; CNN: convolutional neural network; LAT: lateral; T: thoracic; TL: thoracolumbar; R & L-Bending: right and left bending; ALL: all curves; ATR: angle of trunk rotation; AASCE: Accurate Automated Spinal Curvature Estimation; CT: computed tomography; RMSE: root mean square error.

clinical interest in AI-assisted spinal assessment. The geographic distribution was heavily skewed toward China, which contributed 30 studies (66.7%). Canada contributed 3 studies, while Turkey, the Philippines, the United States, Germany, and India each contributed 2. Korea and Japan each contributed 1 study.

Notably, despite the diversity of input modalities across the included studies, all reported sample sizes represented counts of individual analytical units, allowing quantitative comparison of sample sizes across different input types. For gait-based studies, the sample size refers to the number of video sequences rather

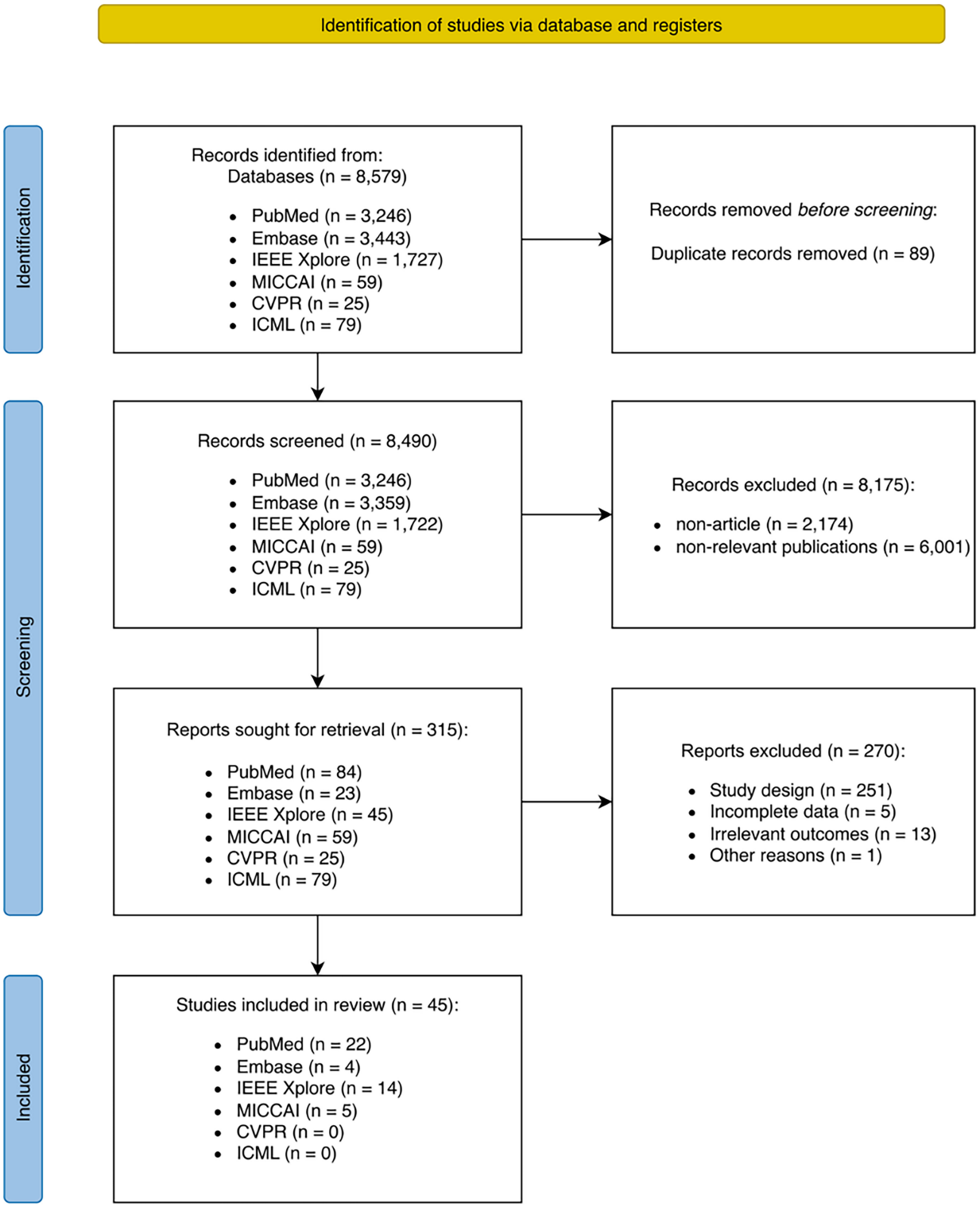


Figure 1. PRISMA flow diagram. Flowchart illustrating the study identification, screening, eligibility, and inclusion process for this systematic review. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; IEEE: Institute of Electrical and Electronics Engineers; MICCAI: Medical Image Computing and Computer Assisted Intervention; CVPR: Computer Vision and Pattern Recognition; ICML: International Conference on Machine Learning; *n*: number of records, reports, or studies, as applicable.

than individual participants or image frames. Sample sizes varied substantially, ranging from fewer than 100 images to more than 20,000. The median sample size was 609, a value strongly influenced by the widespread use of the AASCE MICCAI 2019 challenge dataset, which contains 609 anteroposterior spine radiographs

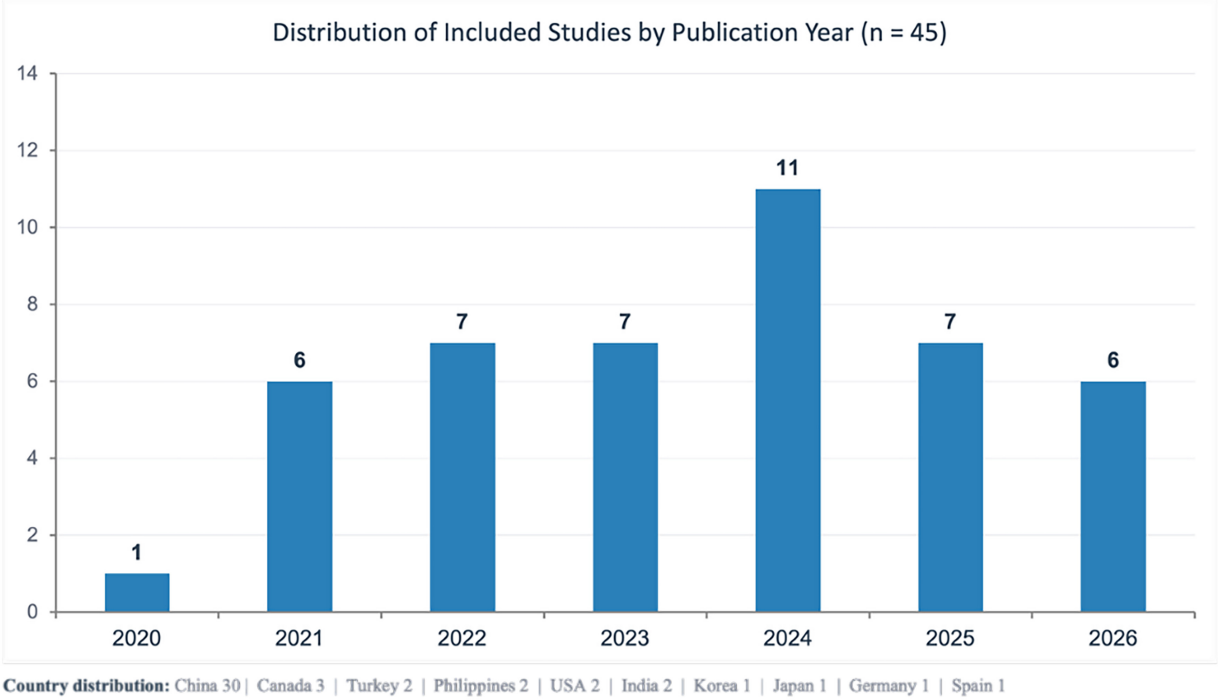


Figure 2. Distribution of included studies by publication year. Bar chart showing the number of included studies (*n* = 45) published each year from 2020 to 2026. The country distribution of the included studies is provided at the bottom.

and served as the sole or partial data source for 10 studies. The concentration of both geographic origin and benchmark data within a limited number of sources has implications for the diversity of clinical settings represented in the current evidence base.

AI architectures and methodological landscape

The 45 studies employed a diverse range of model architectures encompassing both classical ML and modern DL approaches [Table 2]. U-Net and its variants were the most frequently used architectures, appearing in 14 studies, predominantly for vertebral and spinal segmentation. ResNet served as a versatile feature-extraction backbone in 11 studies and was applied to classification, landmark detection, and direct angle regression. HRNet was used in 6 studies for tasks requiring preservation of high-resolution spatial information, particularly vertebral keypoint localization. Convolutional Neural Network (CNN)-Transformer hybrid architectures appeared in 4 studies, reflecting a recent trend toward integrating local feature extraction with global contextual modeling. Classical ML methods, including Support Vector Machine (SVM) and XGBoost, were used in 4 studies, representing a total of 5 model applications. These methods were primarily applied to classification tasks using structured feature inputs rather than raw images. Less frequently used architectures included YOLO and Faster R-CNN for object detection, CapsNet for pose-aware landmark detection, and DeepLab and Segment Anything Model (SAM) for semantic segmentation. Several studies integrated multiple model families within a single pipeline, using one architecture for feature extraction or segmentation and another for downstream classification or regression.

Data sources and modalities

The input data modalities used across the 45 studies are summarized in Table 3. Anteroposterior spine radiographs remained the predominant input modality and were used in 29 studies. An additional 5 studies expanded radiographic inputs to include lateral views, bending radiographs, pelvic radiographs, or CT slices. Overall, radiograph-based studies accounted for 34 of the 45 studies (75.6%).

Table 2. AI architectures and their roles in included AIS studies

Model family	*n*	Primary technical role
U-Net and variants	14	Vertebral/spinal semantic segmentation
ResNet	11	Feature backbone for classification, landmark detection, and direct regression
HRNet	6	High-resolution keypoint and landmark localization
CNN + Transformer hybrids	4	Joint local and global feature extraction for landmark detection
YOLO	3	Real-time vertebral object detection and landmark localization
SVM	3	Classical classification for structured feature inputs
CNN + Attention	2	Region-focused analysis with interpretable attention maps
Faster R-CNN	2	Two-stage object detection for vertebral structure localization
XGBoost	2	Gradient-boosted ensemble classification
Other*	5	CapsNet, DeepLab, SAM, Swin Transformer, EfficientNet (1 study each)

*Single-study models were grouped for readability. Several studies used multiple models; counts represent the number of studies employing each model family rather than mutually exclusive assignments. AI: Artificial intelligence; AIS: adolescent idiopathic scoliosis.

Table 3. Categorization of input data modalities in included AIS studies

Input Modality	*n*
Radiographic (*n* = 34, 75.6%)	
Spine X-ray (AP)	29
Spine X-ray (AP and LAT)	2
Spine X-ray (AP and Bending)	1
Spine X-ray (AP) and CT sagittal slices	1
Pelvis radiograph	1
Non-radiographic (*n* = 11, 24.4%)	
Gait video	4
Smartphone back photograph	2
3D back surface scan	2
Ultrasound (coronal VPI)	1
Torsobarographic pressure imaging	1
Wrist pulse signal	1

AIS: Adolescent idiopathic scoliosis; AP: anterior-posterior; LAT: lateral; VPI: volume projection imaging.

The remaining 11 studies (24.4%) used non-radiographic inputs spanning 6 distinct modalities. Gait video analysis constituted the largest non-radiographic category, with 4 studies, followed by smartphone back photography and 3D surface scanning, with 2 studies each. Ultrasound, torsobarographic pressure imaging, and wrist pulse signals were each investigated in a single study. The emergence and expansion of these non-radiographic approaches are examined in detail in Section "The shift toward radiation-free AIS assessment".

AI applications across clinical tasks in AIS

The 45 included studies addressed 5 distinct clinical tasks. Many studies contributed to more than one task, most commonly by combining Cobb angle measurement with severity classification or curve typing within a single pipeline. The following subsections organize the literature according to clinical task to identify both the areas in which substantial progress has been achieved and those that remain underexplored. Detailed performance metrics for all included studies are provided in Table 1. Studies using non-radiographic input modalities are discussed in detail in Section "The shift toward radiation-free AIS assessment" and are

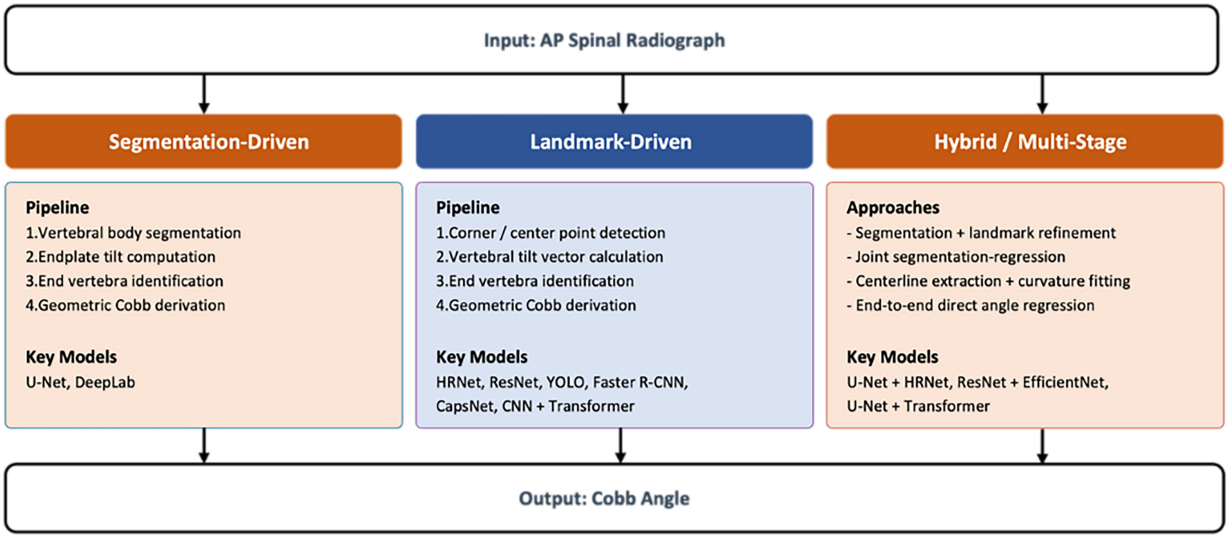


Figure 3. Methodological approaches to automated Cobb angle measurement. Schematic overview categorizing the radiograph-based studies into segmentation-driven, landmark-driven, and hybrid/multi-stage pipelines, along with the representative AI models used in each approach. AP: Anterior-posterior; AI: artificial intelligence.

referenced only briefly here. More detailed information on the clinical tasks is provided in Supplementary Table 2.

Cobb angle measurement

Automated Cobb angle measurement was the most extensively investigated task, addressed by 32 of the 45 included studies. Of these, 29 used anteroposterior spinal radiographs, whereas 3 used non-radiographic modalities (Section "The shift toward radiation-free AIS assessment"). The methodological approaches could be broadly classified into 3 categories [Figure 3]. Segmentation-driven pipelines train a network to delineate vertebral contours, estimate endplate inclination from the resulting masks, and geometrically derive the Cobb angle. Landmark-driven pipelines detect anatomical keypoints, such as vertebral corners or centroids, and calculate the angle using vector geometry. Hybrid and multistage pipelines combine elements of both approaches or employ joint segmentation-regression frameworks.

Despite substantial architectural diversity, reported performance converged within a relatively narrow range. In X-ray-based studies, mean absolute error (MAE) values typically ranged from 1.07° to 5.31°, while symmetric mean absolute percentage error (SMAPE) values ranged from 4% to 15%. These values approach the documented range of inter-observer variability in manual Cobb angle measurement[51-53], suggesting that automated methods are approaching the practical accuracy ceiling imposed by variability in the reference standard itself. Several studies reported reduced performance for mild curves below 10° and severe curves above 50°, for which end-vertebra identification is particularly challenging because of subtle differences in vertebral tilt or complex multicurve patterns.

Two extensions beyond standard anteroposterior Cobb angle measurement are noteworthy. Du *et al.* incorporated bending radiographs for Lenke classification alongside Cobb angle measurement, demonstrating that a single pipeline could support multiple components of the preoperative assessment workflow[31]. Given the 3D nature of AIS, Zhang *et al.* used lateral radiographs to measure sagittal parameters, enabling a more comprehensive evaluation of spinal deformity[37]. These extensions suggest a broader trajectory toward AI systems capable of supporting comprehensive clinical assessment rather than isolated measurements.

Binary scoliosis classification

Binary classification of scoliosis was addressed in 8 studies, 4 of which used non-radiographic inputs (Section "The shift toward radiation-free AIS assessment"). Among radiograph-based studies, binary classification was typically incorporated into broader analytical pipelines rather than treated as a standalone objective. The high classification accuracy reported in these studies likely reflects the relative simplicity of the task on radiographic images, where the visual features distinguishing scoliotic from non-scoliotic spines are relatively well defined. A more clinically relevant question is whether AI can achieve reliable detection using lower-cost and more accessible inputs in settings where radiography is unavailable, a gap that the non-radiographic approaches discussed in Section "The shift toward radiation-free AIS assessment" seek to address.

Curve classification and severity grading

Classification according to severity, curve type, or structured taxonomic systems was addressed in 12 studies. Among the X-ray-based studies, 2 distinct strategies emerged. The first derives classifications from automated Cobb angle measurements by applying established clinical thresholds[32,43], providing a transparent and clinically verifiable approach. The second directly addresses multiparameter classification systems. Du *et al.*[31] and Rahmaniar *et al.*[36] investigated automated Lenke classification, which requires simultaneous assessment of curve type, lumbar modifier, and sagittal modifier. Du *et al.* incorporated bilateral bending radiographs to assess curve flexibility and achieved an accuracy of 93%, representing one of the most clinically comprehensive single-pipeline applications identified in this review[31]. Xie *et al.* extended classification to sagittal alignment abnormalities using biplanar radiographs, reflecting increasing clinical attention to sagittal balance in AIS management[25]. Among studies using non-radiographic modalities (Section "The shift toward radiation-free AIS assessment"), classification performance was consistently lower than that achieved with radiograph-based approaches, particularly when distinguishing adjacent severity categories and mixed curve patterns.

Vertebral rotation assessment

Only 2 studies addressed automated assessment of AVR. Zhao *et al.* used HRNet for vertebral landmark and pedicle detection and subsequently applied the Stokes method to estimate AVR, achieving greater accuracy at the main thoracic curve than at the lumbar or upper thoracic segments[19]. Logithasan *et al.* used U-Net-based segmentation of the spinal column, vertebrae, and pedicle regions to derive AVR, achieving a clinical acceptance rate of 88.1%[21]. The limited number of studies addressing this task is notable given the clinical importance of vertebral rotation in surgical planning and the poor reproducibility of manual AVR measurements. The dependence of the Stokes method on precise pedicle identification, which is technically challenging on plain radiographs, may partly explain the limited research activity in this area.

Skeletal maturity assessment

Two studies investigated automated Risser sign grading. Magnide *et al.* trained a ResNet-SVM pipeline to classify Risser stages using multiple anatomical regions on AP spinal radiographs, achieving accuracies of 78%-84% depending on the region analyzed[22]. Zhao *et al.* used a CNN-Swin Transformer hybrid model on pelvic radiographs and achieved an overall accuracy of 81.5%, with Gradient-weighted Class Activation Mapping (Grad-CAM) visualization confirming that model attention was focused on the iliac apophysis[23]. Both studies reported that early Risser stages, particularly stage 0, were the most difficult to classify, likely because pre-ossification states provide fewer discriminative imaging features. Given that the primary clinical value of Risser staging lies in identifying patients who remain in periods of rapid growth and are therefore at greater risk of curve progression[54], future studies may benefit from evaluation metrics that assign greater penalties to prognostically consequential misclassifications than to errors between adjacent stages.

Table 4. Summary of evidence quality assessment in included studies

Ref.	Study Design	DP	Dataset source	Dataset size	DSR	CAR	QC	COS	BMI	XAI Used	TED	CMC
Tınas and Esmer[6], 2024	Retrospective	Yes	Public	259 participants	71:188	-	Not specified	No	No	No	No	Yes
Goral and Kose[7], 2022	Retrospective	Yes	Public	609 images	-	-	Expert-reviewed/guided	No	Yes	Yes (CAM)	No	Yes
Wang *et al.*[8], 2024	Retrospective	Yes	Public	580 images	-	-	Not specified	No	Yes	No	No	Yes
Peng *et al.*[9], 2024	Retrospective	Yes	Private	155 images	-	-	Expert-reviewed/guided	No	Yes	No	No	Yes
Kim *et al.*[10], 2020	Retrospective	Yes	Private	629 images	-	-	Expert-annotated	No	Yes	Yes (confidence map + tilt field)	Yes	Yes
Maaliw *et al.*[11], 2022	Retrospective	Yes	Public, Private	210 images	-	-	Not specified	No	Yes	No	No	Yes
Liu *et al.*[12], 2021	Retrospective	Yes	Private	200 images	0:1	10°-45°	Not specified	No	No	No	No	No
Zou *et al.*[13], 2023	Retrospective	Yes	Public	609 images	-	-	Not specified	No	Yes	No	No	Yes
Shang *et al.*[14], 2023	Retrospective	No	Private	242 pulse recordings	-	-	Not specified	No	No	No	No	No
Liang *et al.*[15], 2023	Retrospective	Yes	Public, Private	Public: 707 images Private: 36 images	-	0°-80°	Expert-annotated	No	Yes	No	Yes	Yes
Qiu *et al.*[16], 2024	Retrospective	Yes	Public	609 images	-	-	Expert-annotated	No	Yes	Yes (Grad-CAM)	No	Yes
Dagli *et al.*[17], 2025	Retrospective	No	Public, Private	Public: 609 Private: 166 images	0:1	10°-80°	Expert-annotated	No	No	No	Yes	Yes
Zhu *et al.*[18], 2023	Retrospective	Yes	Private	380 images	-	-	Expert-reviewed/guided	No	Yes	No	No	Yes
Zhao *et al.*[19], 2024	Retrospective	No	Public, Private	Public: 609 Private: 220 images	-	-	Expert-annotated	No	Yes	No	No	Yes
Zhang *et al.*[20], 2021	Retrospective	Yes	Public	707 images	-	-	Not specified	Yes	Yes	Yes (NDPAFs visualization + heatmap)	No	Yes
Logithasan *et al.*[21], 2022	Retrospective	Yes	Private	297 images	0:1	Training dataset: 25.5° ± 16.2° (range 6° to 97°) Test dataset: 26.3° ± 10.6° (range 7° to 41°)	Expert-annotated	No	Yes	No	No	Yes
Magnide *et al.*[22], 2021	Prospective	Yes	Private	514 segmented images	0:1	-	Not specified	No	Yes	Yes (Grad-CAM)	No	Yes
Zhao *et al.*[23], 2025	Retrospective	Yes	Private	634 images (for iliac region detection) 1,266 images (for Risser stage assessment)	-	-	Expert-annotated	Yes	Yes	Yes (Grad-CAM)	No	Yes

Zhang *et al.*[24], 2023	Retrospective	Yes	Private	5,152 reference radiographs	-	Cobb angle < 20° group: 3.9° ± 1.2° Cobb angle 20°-40° group: 16.1° ± 6.3° Cobb angle > 40° group: 48.7° ± 22.6°	Expert-annotated	Yes	Yes	Yes (Score-CAM)	Yes	No
Xie *et al.*[25], 2025	Retrospective	Yes	Private	600 patients	308:292	0°-77.7°	Expert-reviewed/guided	No	Yes	No	No	Yes
Li *et al.*[26], 2024	Retrospective	Yes	Private	20,802 images	-	Cobb angle < 25° group: 19.43° ± 3.35° (maximum = 24.92°) Cobb angle 25°-45° group: 32.99° ± 5.52° (maximum = 44.93°) Cobb angle > 45° group: 52.04° ± 6.51° (maximum = 75.02°)	Expert-annotated	No	Yes	No	No	Yes
Wu *et al.*[27], 2022	Retrospective	Yes	Public, Private	Public: 72 Private: 901 images	0:1	-	Expert-annotated	No	Yes	No	Yes	Yes
Maaliw *et al.*[28], 2023	Retrospective	Yes	Public	318 images	0:1	-	Expert-annotated	No	Yes	No	No	Yes
Wong *et al.*[29], 2024	Retrospective	Yes	Private	330 images	0:1	Spinal column segmentation training dataset: 24.6° ± 12.4° (range 6° to 97°) Vertebral body segmentation training dataset: 24.9° ± 10.8° (range 7° to 54°) Cobb angle measurement test dataset: 24.6° ± 9.7°(range 8° to 52°)	Expert-reviewed/guided	No	Yes	Yes (visual segmentation outputs)	No	Yes
Liu *et al.*[30], 2021	Retrospective	Yes	Public	157 images	-	-	Expert-annotated	No	Yes	No	No	Yes
Du *et al.*[31], 2025	Retrospective	Yes	Public, Private	Public: 609 Private: 372 images	0:1	PT Cobb: 24.30° ± 17.87° MT Cobb: 47.36° ± 12.27° TL Cobb: 36.22° ± 14.51°	Expert-annotated	No	Yes	Yes (visual keypoint outputs)	Yes	Yes
Lu *et al.*[32], 2024	Retrospective	Yes	Public, Private	Public: 481 Private: 118 patients	6:112	23.62° (16.54°, 30.52°) (median, Q1, Q3)	Expert-annotated	No	Yes	No	Yes	Yes
Li *et al.*[33], 2024	Retrospective	Yes	Private	3,640 images	-	-	Expert-annotated	No	Yes	Yes (visual keypoint + contour outputs)	Yes	Yes

Zou *et al.*[34], 2023	Retrospective	Yes	Public, Private	Public: 609 Private: 508 images	-	0°-90°	Not specified	No	Yes	No	No	Yes
Gadu and Potala[35], 2026	Retrospective	Yes	Public	Spine Web: 609 images; AASCE: 609 images	0:1	17°-75°	Expert-annotated	No	Yes	No	No	Yes
Rahmaniar *et al.*[36], 2024	Retrospective	Yes	Public	481 images	-	0°-90°	Not specified	No	Yes	Yes (visual UI + bounding box + labels)	No	Yes
Zhang *et al.*[37], 2022	Retrospective	Yes	Private	2,738 radiograph pairs	0:1	-	Expert-annotated	No	Yes	No	No	Yes
Yang *et al.*[38], 2025	Retrospective	Yes	Public, Private	Public: 591 Private: 819 images	-	-	Not specified	No	Yes	Yes (heatmaps + anatomy metrics)	No	Yes
Yao *et al.*[39], 2022	Retrospective	Yes	Public	609 images	-	26.64° ± 17.08°	Not specified	No	Yes	No	No	Yes
Zhao *et al.*[40], 2022	Retrospective	Yes	Public, Private	Public: 98 Private: 185 images	-	-	Not specified	No	Yes	No	Yes	Yes
Feng *et al.*[41], 2025	Prospective	Yes	Private	160 images (640 after augmentation)	0:1	3°-99°	Expert-annotated	No	Yes	No	No	Yes
González-Ruiz *et al.*[42], 2026	Retrospective	Yes	Private	106 participants	0:1	10°-60°	Not specified	No	No	No	No	Yes
Chen *et al.*[43], 2026	Retrospective	Yes	Private	3,899 images	-	0°-80°	Expert-annotated	No	Yes	No	Yes	Yes
Stecher *et al.*[44], 2025	Prospective	Yes	Private	87 participants	-	average Cobb angle 16.5° ± 7.7°	Expert-annotated	No	No	No	No	Yes
Vishwakarma *et al.*[45], 2026	Retrospective	Yes	Public	1,493 video sequences	493:800	-	Not specified	No	Yes	Yes (activation heatmap)	Yes	Yes
Lin *et al.*[46], 2021	Retrospective	Yes	Public	609 images	-	-	Not specified	No	Yes	Yes (CAM)	No	Yes
Huo *et al.*[47], 2021	Retrospective	Yes	Public	609 images	-	-	Not specified	No	Yes	No	No	Yes
Zhou *et al.*[48], 2024	Retrospective	Yes	Public	1,493 video sequences	493:800	-	Expert-annotated	Yes	Yes	Yes (attention transfer)	No	Yes
Zhou *et al.*[49], 2026	Retrospective	Yes	Public	1,493 video sequences	493:800	-	Expert-annotated	Yes	Yes	Yes (CAM)	No	Yes
Li *et al.*[50], 2026	Retrospective	Yes	Public	1,493 video sequences	493:800	-	Expert-annotated	Yes	Yes	No	No	Yes

The Scoliosis1K dataset contains 1,493 video sequences, including 800 normal, 493 scoliosis, and 200 neutral (borderline) sequences. The neutral sequences are excluded from the DSR calculation. DP: Data partitioning; DSR: dataset scoliosis ratio (Normal: Scoliosis); CAR: Cobb angle reported; QC: dataset quality control (measurement quality control); COS: codebase open source; BMI: base model improvement; XAI: explainable artificial intelligence; TED: testing on external data; CMC: competitive models comparison; CAM: class activation mapping; NDPAFs: non-directional part affinity fields; UI: user interface.

The shift toward radiation-free AIS assessment

Repeated radiographic imaging is a well-recognized concern in the management of AIS. Patients typically undergo serial full-spine radiographs over several years of growth monitoring, and cumulative radiation exposure during adolescence may have long-term health implications. This concern has motivated a growing body of research seeking to replace or supplement conventional radiography with non-radiographic input modalities. Among the 45 studies included in this review, 11 (24.4%) used input data that did not involve ionizing radiation, encompassing 6 distinct modalities. Most of these studies were published after 2022, suggesting that radiation-free AI-based assessment of AIS represents an emerging and increasingly active research direction rather than an isolated line of investigation.

Ultrasound-based assessment represents one of the earliest efforts to incorporate radiation-free imaging into AI-assisted scoliosis evaluation. Liu *et al.* developed a DL pipeline that first detected vertebral laminae in ultrasound volume projection images and then estimated spinal curvature from the identified landmarks[12]. The method demonstrated moderate agreement with manual ultrasound measurements and somewhat weaker agreement with radiographic Cobb angles, reflecting the inherent differences between ultrasound-derived curvature measurements and the radiographic reference standard. Ultrasound offers the advantages of real-time, portable imaging without ionizing radiation; however, its dependence on operator skill and acoustic window quality introduces variability that may be difficult to standardize across clinical settings.

At the other end of the input spectrum, Shang *et al.* explored a fundamentally different signal source by using radial pulse waveforms for scoliosis classification[14]. This approach was based on traditional Chinese medicine concepts linking spinal alignment with peripheral vascular dynamics. Using classical ML classifiers applied to pulse-signal features, the study achieved a classification accuracy below 81%, which remains insufficient for standalone clinical diagnosis. The primary significance of this work lies less in its current performance than in demonstrating that scoliosis may be associated with detectable physiological signals beyond the musculoskeletal system. Whether pulse-derived features could complement other modalities within a multimodal screening framework remains to be determined.

Smartphone-based back photography has attracted increasing attention as a low-cost and highly accessible screening modality. Zhang *et al.* trained a model using back photographs to perform severity classification and curve-type classification and retrospectively assessed disease progression by comparing paired baseline and follow-up photographs with corresponding radiographic changes[24]. Li *et al.* adopted a more comprehensive strategy by developing a keypoint-detection model for the back surface, followed by downstream classification of curve type and severity[33]. The study also attempted to estimate axial trunk rotation through edge detection on surface images. Both studies demonstrated that clinically relevant postural features can be extracted using a standard smartphone camera without specialized equipment. However, model performance varied substantially across classification categories, with mixed-type curves and moderate-severity scoliosis proving particularly difficult to distinguish. Smartphone-based assessment is therefore highly attractive for large-scale community screening, particularly in settings with limited access to radiography; nevertheless, neither study demonstrated sufficient accuracy to support clinical decision-making at the individual-patient level.

3D surface scanning provides a more structured alternative to smartphone photography by capturing the complete topography of the torso. Feng *et al.* used a structured-light 3D scanner to reconstruct trunk surface geometry and trained a model to estimate the Cobb angle, curve type, and curve direction directly from surface data[41]. This study was notable for its prospective design and strong correlation with radiographic Cobb angles, with an MAE of less than 3°. González-Ruiz *et al.* similarly estimated Cobb angles from 3D surface scans and reported comparable accuracy with corresponding 95% confidence intervals[42]. Together,

these studies indicate that external trunk morphology contains sufficient geometric information to approximate underlying spinal curvature with clinically meaningful precision. The principal limitations of 3D surface scanning are the cost of specialized equipment and the requirement for a controlled acquisition environment, which favor its use in clinic-based settings rather than community-level screening.

Torsobarographic pressure imaging, investigated by Stecher *et al.*, captures pressure-distribution patterns while the patient lies on a sensor mat[44]. This modality reflects asymmetric load distribution resulting from spinal and rib cage deformities. The study used a prospective design and achieved high classification accuracy, suggesting that pressure-derived features may be both discriminative and reproducible. As a passive and fully automated measurement technique that does not require specialized patient-positioning skills, torsobarography has practical advantages for integration into routine clinical workflows. However, its current evidence base is limited by the absence of replication by independent research groups and the lack of validation across diverse patient populations.

Gait video analysis represented the largest category of non-radiographic research identified in this review, accounting for 4 of the 11 studies (36.4%). Vishwakarma *et al.*[45], Zhou *et al.*[48], Zhou *et al.*[49], and Li *et al.*[50] all used video-derived gait features to classify scoliosis, based on the premise that spinal deformity produces detectable asymmetries in walking patterns. Reported classification accuracies ranged from 82% to approximately 90%, with sensitivity exceeding 99% in some model configurations. Gait analysis is particularly attractive as a screening modality because it requires only a standard video camera, involves no physical contact with the patient, and can be performed in an open environment. However, an important limitation must be considered when interpreting these findings. All 4 studies used the same Scoliosis1K dataset for model training and evaluation. Consequently, these studies should not be considered 4 independent validations of gait-based scoliosis detection. Rather, they represent 4 methodological approaches evaluated in the same population under a common data-collection protocol. The resulting evidence base is therefore narrower than the number of studies might suggest, and the generalizability of gait-based classification across different clinical settings, camera configurations, and patient populations remains uncertain.

In conclusion, the 11 non-radiographic studies exhibited several characteristics that distinguished them from the broader AI-AIS literature. They were predominantly recent, were more likely to use prospective study designs (2 of the 3 prospective studies identified in this review belonged to this group), and focused primarily on screening-oriented tasks rather than precise quantitative measurements. The wide range of modalities under investigation indicates that the field has not yet converged on a dominant non-radiographic approach. Although this diversity reflects substantial methodological innovation, it also suggests that no single modality has yet demonstrated sufficient advantages to replace radiography in clinical assessment. The shared-dataset limitation observed in gait-analysis studies further highlights the need for larger, publicly available, multicenter datasets that enable independent validation across this emerging field. For non-radiographic approaches to progress from research prototypes to clinically applicable tools, future studies should demonstrate robust performance not only against radiographic reference standards but also across the demographic, anthropometric, and environmental heterogeneity encountered in real-world adolescent populations.

Methodological quality and reporting practices

To move beyond a narrative summary of model performance toward a more structured evaluation of evidence quality, we assessed all 45 included studies using a domain-specific quality framework informed by the 2024 update of the CLAIM[55]. CLAIM provides a standardized set of reporting items for AI studies in medical imaging, from which we selected 12 assessment dimensions considered most relevant to the AI-AIS

literature. Each dimension presented in Table 4 corresponds to one or more CLAIM items. Study Design corresponds to CLAIM Item 5 (prospective or retrospective design). Dataset Source and Dataset Size correspond to CLAIM Items 7 and 21, respectively. DP corresponds to CLAIM Item 19. DSR and CAR both correspond to CLAIM Item 36, which addresses the demographic and clinical characteristics of the study population. QC encompasses CLAIM Items 14, 16, and 18, covering reference-standard definition, annotation source, and inter- or intrarater variability. BMI corresponds to CLAIM Item 22 regarding detailed model description. COS corresponds to CLAIM Item 43 on software and data availability. XAI Used corresponds to CLAIM Item 31 on interpretability methods. TED corresponds to CLAIM Item 33, and CMC corresponds to CLAIM Item 28 on performance metrics. The following sections synthesize findings across these dimensions to identify methodological strengths and critical gaps in the current AI-AIS literature.

Strengths and gaps in current AI-AIS reporting

The field demonstrated consistent strengths in several aspects of methodological reporting. Forty-two of the 45 studies (93.3%) clearly described their data-partitioning strategies, indicating that separation of training, validation, and test sets has become a well-established practice. Similarly, 42 studies (93.3%) included comparisons with competitive baseline models, and 39 (86.7%) reported meaningful improvements over these baselines. These findings suggest that the AI-AIS research community has largely adopted the expectation that newly developed methods should be benchmarked against existing approaches and that architectural innovations should be described in sufficient detail to facilitate replication.

Annotation quality is fundamental to supervised learning and was explicitly addressed in 28 of the 45 studies (62.2%). Among these, 23 relied on expert-annotated ground truth, whereas 5 used expert-reviewed or expert-guided labeling. This level of specialist involvement represents an important methodological strength because the reliability of Cobb angle measurement and vertebral landmark identification depends heavily on annotator expertise. However, the remaining 17 studies (37.8%) did not specify how annotations were generated or quality controlled. This group was not entirely homogeneous. Several studies relied on the AASCE MICCAI 2019 challenge dataset, which was released with challenge-organizer annotations and therefore provides an implicit level of annotation quality even when individual authors did not describe their own QC procedures. The concern is greater for studies that constructed private clinical datasets without reporting annotation provenance, annotator qualifications, or inter-rater agreement, because readers are then unable to assess the reliability of the ground truth underlying the reported results.

External validation remained a major limitation. Only 12 of the 45 studies (26.7%) evaluated their models using data from institutions or populations not represented in the training set. The remaining 33 studies (73.3%) relied exclusively on internal data splits, which cannot adequately account for distribution shifts arising from differences in imaging equipment, acquisition protocols, patient demographics, and annotation conventions across clinical settings. This limitation is particularly relevant to Cobb angle measurement models, for which even small systematic biases may shift a patient across clinical decision thresholds for bracing or surgery.

The predominance of retrospective study designs further limited the strength of the available evidence. Forty-two of the 45 studies (93.3%) were retrospective and relied on pre-existing image archives rather than prospectively defined cohorts. Only 3 studies adopted prospective designs. Magnide *et al.* prospectively validated an automated Risser grading system using spinal radiographs[22], Feng *et al.* prospectively collected 3D surface scans under a standardized protocol[41], and Stecher *et al.* prospectively enrolled patients for torsobarographic scoliosis classification[44]. The latter 2 studies used non-radiographic modalities, reinforcing the broader shift toward radiation-free assessment described in Section "The shift toward radiation-free AIS

assessment". The near absence of prospective validation across the remaining literature indicates that most reported performance estimates have not been evaluated under real-time clinical conditions, in which variability in image quality, patient positioning, and workflow constraints may not be adequately represented by retrospective datasets.

Interpretability methods have received increasing attention but remain relatively uncommon. Sixteen of the 45 studies (35.6%) incorporated some form of XAI, most commonly Grad-CAM, Score-CAM, or confidence-map visualizations that enable clinicians to inspect the image regions contributing to model predictions. Such approaches represent an important step toward clinical interpretability, particularly because clinicians may be reluctant to adopt measurement tools whose decision-making processes cannot be examined. Nevertheless, 29 studies (64.4%) provided no interpretability mechanism. For tasks such as Cobb angle measurement, where clinical decisions depend directly on quantitative outputs, the inability to verify whether a model is correctly identifying the relevant vertebral endplates rather than adjacent soft-tissue structures may represent a practical barrier to clinical adoption.

Code and data availability represented one of the most pronounced deficiencies. Only 6 of the 45 studies (13.3%) released their source code or provided an openly accessible platform, leaving most studies without a means for independent verification, secondary analysis, or downstream benchmarking. This limited openness is particularly problematic given the heavy reliance on a small number of shared datasets. Ten studies (22.2%) used the AASCE MICCAI 2019 challenge dataset, which contains only 609 anteroposterior spine radiographs. Because this dataset originates from a single institution and follows a fixed imaging protocol and annotation convention, models trained and evaluated predominantly on AASCE data inevitably reflect characteristics of that specific clinical environment. Multiple research groups have developed and tested models using the same 609 images, often without releasing their code. Consequently, apparently independent studies may be based on substantially overlapping evidence. This overlap complicates cross-study comparisons and represents one reason why a formal meta-analytic synthesis of current AI-AIS results remains technically challenging. A similar issue is evident in the gait-based studies discussed in Section "The shift toward radiation-free AIS assessment", all 4 of which relied on the same Scoliosis1K dataset, further illustrating how shared data dependencies can inflate the apparent volume of independent evidence.

Reporting of population characteristics was inconsistent. Eighteen studies (40.0%) reported the Cobb angle range of their study population, whereas 27 (60.0%) did not, preventing readers from determining whether model performance had been evaluated across the full clinical spectrum from mild to severe deformity. The proportion of scoliosis cases within each dataset was also reported in only a minority of studies. Without this information, it is difficult to determine whether a reported accuracy of 90% reflects robust performance across balanced classes or merely the influence of an imbalanced prevalence distribution. These omissions fall within the scope of CLAIM Item 36 and represent a fundamental reporting deficiency that limits the interpretability of headline performance metrics.

In summary, the AI-AIS literature reflects a field that has established relatively strong conventions for model benchmarking, DP, and architectural innovation but has not achieved comparable maturity in clinically oriented reporting practices. External validation, prospective study design, code availability, and population-level descriptive statistics remain among the most prominent weaknesses. Addressing these limitations will be essential before the models reviewed here can be credibly considered for integration into routine clinical workflows.

Reporting items beyond CLAIM coverage in the AI-AIS literature

Although CLAIM provides a comprehensive reporting framework for AI studies in medical imaging, several practices routinely expected in clinical research fall outside its explicit scope. These elements are not merely optional additions; rather, they are fundamental to establishing the clinical credibility of any diagnostic tool. Their inconsistent reporting in the AI-AIS literature therefore constitutes an important gap that warrants separate consideration.

Statistical inference was rarely reported across the 45 included studies. Only a small minority provided confidence intervals for their primary performance metrics, and formal hypothesis testing was similarly uncommon. Most studies reported point estimates of accuracy, MAE, or SMAPE without any accompanying measure of uncertainty. In a clinical context, the distinction between a mean Cobb angle error of 2.5° with a narrow confidence interval and the same error with a wide confidence interval has direct implications for confidence in the tool at the individual-patient level. The routine omission of measures of uncertainty makes it difficult to distinguish models with consistently stable performance from those whose average performance may conceal substantial case-to-case variability.

Sample size justification was absent from all 45 studies. None reported a power analysis, minimum detectable effect size calculation, or other formal rationale for the number of images or patients included. Although the widespread use of convenience samples and existing challenge datasets may make prospective sample size planning difficult, the absence of formal sample size justification limits readers' ability to assess whether negative or null findings reflect genuine equivalence or insufficient statistical power. This issue is particularly relevant for studies based on small private datasets, in which the distinction between a robust finding and a chance observation may be especially uncertain.

Reporting of ethical oversight and patient consent was also inconsistent. Although several studies relied on publicly available challenge datasets for which ethical approval had been obtained during the original data collection, studies using private clinical datasets did not consistently report institutional review board approval, informed consent procedures, or data de-identification protocols. Given the substantial reliance of this field on imaging data from adolescent patients, transparent reporting of ethical safeguards is not merely an administrative requirement but an essential component of responsible research conduct and dissemination.

Failure-mode analysis was largely absent. Most studies reported aggregate performance metrics across the entire test set, whereas few examined model behavior in clinically challenging subgroups, such as patients with severe curves exceeding 60°, atypical curve patterns, or images of suboptimal quality. Without such analyses, clinicians considering the adoption of an AI tool cannot determine whether the reported performance applies broadly to individual patients or primarily to a narrower subset of relatively straightforward cases. The limited number of studies that stratified performance by severity or curve type identified meaningful reductions in accuracy at the extremes of the clinical spectrum, suggesting that aggregate metrics may overestimate real-world reliability.

Calibration and alignment with clinical decision-making were not systematically evaluated. In AIS management, the Cobb angle is not simply a continuous measurement but also informs discrete treatment decisions at established clinical thresholds. A model with a favorable MAE may nevertheless systematically overestimate mild curves or underestimate severe curves, thereby shifting patients across clinically relevant decision boundaries. Only a small number of studies evaluated misclassification at clinically meaningful thresholds, and none reported calibration curves or decision-analytic outcomes. This mismatch between how models are evaluated and how their outputs are ultimately used represents one of the most clinically consequential gaps in the current literature.

Individually, each of these reporting deficiencies could be addressed through relatively straightforward additions to the methods and results sections of future studies. Collectively, however, they reveal a broader pattern in which the AI-AIS literature has prioritized technical performance metrics over the clinical and methodological context required to interpret those metrics appropriately. This pattern is not unique to AIS but reflects broader tendencies within the AI-in-medicine literature. These deficiencies should be recognized as a field-level issue rather than merely a collection of study-specific limitations. From this perspective, reporting quality is not simply a checklist-based requirement but a substantive research gap that directly affects the conclusions that can be drawn from the existing evidence. The broader implications of this gap for the future development of AI in AIS are discussed further in Section "DISCUSSION".

DISCUSSION

This systematic review examined 45 studies applying AI to the diagnosis and assessment of AIS, encompassing literature published through May 2026 in both peer-reviewed journals and leading computer science conference proceedings. Several overarching findings emerged from this synthesis. Automated Cobb angle measurement has achieved accuracy approaching the documented inter-observer variability of manual measurement, with MAE values typically ranging from 1.07° to 5.31° across the included studies. Beyond Cobb angle measurement, AI has been applied to binary scoliosis classification, curve classification, vertebral rotation assessment, and skeletal maturity grading, although research activity remains heavily concentrated on Cobb angle estimation. Another notable and rapidly expanding trend is the use of non-radiographic input modalities, with 11 of the 45 studies (24.4%) employing gait videos, smartphone photographs, 3D surface scans, or other radiation-free signals. At the same time, our methodological quality assessment identified persistent deficiencies in external validation, code availability, population-level reporting, and statistical inference, all of which limit the clinical interpretability and generalizability of the reported findings.

Several features distinguish this review from previous work. Zhu *et al.* published a systematic review and meta-analysis in 2025 that focused exclusively on DL for Cobb angle measurement from X-rays. The review included 50 studies, had a search cutoff of September 2023, and used Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) for risk-of-bias assessment[56]. Kumar *et al.* conducted a systematic review of vertebral identification and Cobb angle estimation using DL, similarly restricting their analysis to X-ray-based Cobb angle measurement and studies published from 2017 to 2023[57]. A scoping review by Goldman *et al.* mapped the broader landscape of AI applications in AIS across 40 studies but did not incorporate a structured framework for methodological quality assessment[58]. Li and Wong systematically reviewed ML approaches for predicting AIS curve progression, a task beyond the scope of the present review[59]. The present study extends this body of evidence in several important respects. First, this review covers the full spectrum of AI-assisted diagnostic and assessment tasks rather than focusing on a single measurement. This clinical-task-oriented synthesis allows readers to assess the maturity of AI support across different stages of the AIS clinical workflow. Second, it includes studies published through May 2026 and incorporates MICCAI conference proceedings from 2020 to 2026, thereby capturing methodological innovations that are often disseminated through computer science venues before appearing in clinical journals. Third, it introduces a 12-dimensional quality assessment framework aligned with the CLAIM 2024 reporting standard, allowing a more granular evaluation of reporting practices than the binary or categorical approaches used in previous reviews. Fourth, it provides a dedicated analysis of non-radiographic modalities, an area of the literature that has not been systematically examined in previous reviews of AI applications in AIS.

The clinical implications of the available evidence vary across the 5 task categories. For Cobb angle measurement, the convergence of MAE values toward the range of manual measurement variability suggests that automated systems have achieved a relatively mature level of accuracy and consistency for angle quantification. However, this finding does not imply that current AI systems can independently support comprehensive AIS assessment or clinical decision-making. AIS is a 3D deformity, and treatment decisions also depend on curve pattern, sagittal alignment, vertebral rotation, skeletal maturity, curve flexibility, and other clinical factors. For outcomes such as vertebral rotation, skeletal maturity assessment, and curve classification, AI performance remains less consistent and the supporting clinical evidence is comparatively limited. These outputs should therefore be interpreted more cautiously than automated Cobb angle measurements. Importantly, variability in manual measurements is not solely technical in origin. Although standardized measurement protocols exist and the definitions of key parameters such as the Cobb angle are well established, their implementation in routine practice varies across clinicians and institutions, resulting in meaningful variation in recorded values. Such practice-level heterogeneity is poorly represented when AI models are trained and evaluated using averaged or consensus labels. AI systems intended for clinical use should ultimately be robust to such real-world variation rather than simply averaging it away. These concerns are compounded by the near-universal reliance on retrospective datasets and the reductions in performance observed during external validation. Prospective, multicenter deployment studies are therefore needed before widespread clinical adoption can be justified. For binary scoliosis classification, current evidence supports the feasibility of AI-assisted detection on radiographic images, but the greater clinical need lies in non-radiographic approaches suitable for community-level screening, where performance has not yet reached a level sufficient for standalone use. For curve classification and Lenke typing, the successful integration of multiple assessment steps into unified AI pipelines provides an important proof of concept for end-to-end automation of selected components of the clinical workflow. In contrast, vertebral rotation assessment and skeletal maturity grading remain substantially underexplored, with only 2 studies addressing each task, leaving considerable scope for further methodological and clinical development.

AI-assisted assessment may support the clinical evaluation of AIS by improving access to screening and follow-up, measurement efficiency, and measurement consistency. For non-radiographic assessment, Zhang *et al.*[24] evaluated a smartphone-based model using back photographs for curve severity classification and progression detection. Their findings suggest that such methods may extend patient monitoring beyond specialist clinics and help identify patients who require further evaluation. For radiographic assessment, Xie *et al.* reported that automated measurement using biplanar imaging required substantially less time than manual assessment (9.1 s *vs.* 185.1 s)[25]. Li *et al.* also reported higher interobserver agreement for an AI-based mobile application than for manual measurements using the picture archiving and communication system, with intraclass correlation coefficient (ICC) values of 0.997 and 0.992, respectively[26]. These findings suggest that AI-assisted assessment may offer benefits in measurement efficiency, reproducibility, and clinical workflow. However, whether improvements in technical performance and efficiency translate into better clinical decisions and patient outcomes remains to be further investigated. For example, differences in Cobb angle measurements may affect whether a patient falls within the range in which brace treatment is considered (usually about 25° to 40°) or the range in which surgical evaluation may be considered (usually 40° to 45° or above). However, treatment decisions also depend on skeletal maturity, curve progression, and other clinical factors. At present, evidence remains insufficient to determine whether AI-assisted assessment improves timely referral, treatment decisions, or long-term outcomes. Future studies should move beyond technical performance comparisons and examine whether measurement differences affect clinical classification and treatment decisions, particularly in patients whose Cobb angles are close to clinically relevant thresholds. They should also determine whether these changes lead to better patient outcomes.

The methodological quality analysis presented in Section "Methodological quality and reporting practices" revealed a field-level pattern in which technical benchmarking practices are well established, whereas clinically oriented reporting remains underdeveloped. Although 93.3% of studies clearly described their data-partitioning strategies and included comparisons with competitive models, only 26.7% evaluated their models using external data, 13.3% released source code, and 35.6% incorporated any form of XAI. The near-complete absence of statistical inference, sample size justification, and failure-mode analysis across the reviewed literature means that even the highest-performing models cannot currently be evaluated with the rigor expected of clinical diagnostic tools. Strong internal technical performance alone does not constitute evidence of clinical readiness. Because generalizability can ultimately be established only through external validation, studies should rigorously report internally derived validity and performance metrics even when external testing is not yet feasible. However, we acknowledge that such metrics remain insufficiently developed for these models. There is currently no consensus regarding which internal metrics best predict external performance or how acceptable thresholds should be defined. Until reliable internal validity metrics are established and their relationship with external validity is better understood, these models should be interpreted cautiously in clinical contexts. Their reported performance should primarily be regarded as evidence of feasibility rather than readiness for deployment. This reporting gap is not unique to AIS research but reflects a broader challenge across the AI-in-medicine literature that the CLAIM 2024 framework was specifically designed to address.

This review has several limitations. Restriction to English-language publications may have excluded relevant studies published in other languages, particularly Chinese, given that China contributed 66.7% of the included studies and may have additional domestic publications in Chinese-language journals. The inclusion of conference papers, although a deliberate methodological decision intended to capture emerging methods, introduced heterogeneity in reporting depth because conference papers typically provide more concise methodological descriptions than journal articles. The quality assessment framework, although aligned with CLAIM 2024, was applied retrospectively to studies that were not originally designed according to CLAIM recommendations, and some reporting omissions may therefore reflect publication space constraints rather than methodological deficiencies. In addition, the exclusion of studies focused on surgical outcome prediction and curve progression, although necessary to maintain a coherent review scope, means that the present review does not encompass the full spectrum of AI applications in AIS management.

Several important challenges remain. The shared-dataset problem, exemplified by the 10 studies relying on the 609-image AASCE dataset and the 4 gait-analysis studies using the Scoliosis1K dataset, creates a situation in which the number of published studies substantially exceeds the number of genuinely independent evaluations. This overlap is a major reason why meta-analytic pooling of AI-AIS performance estimates is not currently methodologically appropriate despite the growing volume of literature. The absence of standardized evaluation protocols, including inconsistent choices of error metrics, threshold definitions, and severity stratification schemes, further limits meaningful cross-study comparisons. Existing studies vary substantially in clinical tasks, datasets, input modalities, outcome definitions, performance metrics, and validation strategies, precluding reliable direct comparisons among specific methodological combinations. The limited use of prospective designs (3 of 45 studies) and external validation (12 of 45 studies) further indicates that the evidence supporting real-world clinical deployment remains limited.

Future research should prioritize the explicit operationalization and prospective evaluation of clinical measurement protocols. Studies should also characterize variation in clinical practice and determine how such variation influences AI performance. These efforts may help ensure that AI models better reflect real-world clinical judgment rather than merely averaged data. Given the heterogeneity of the current evidence base, another important objective is to determine which combinations of input modality, model architecture,

and methodological approach provide the most robust performance and to identify the minimum input required to achieve clinically acceptable accuracy. Simply increasing data volume or imposing uniform data formats cannot eliminate practice variation and may instead preserve biases embedded within existing datasets. Accordingly, ever-larger multisite data aggregation should not be regarded as the primary solution to the field's current limitations. The principal value of multisite studies lies not merely in increasing sample size but in testing model generalizability across clinically meaningful variations in real-world practice. Prospective validation studies embedded within clinical workflows would provide critical evidence regarding the performance of automated measurements under routine conditions, including variable image quality, nonstandard patient positioning, and diverse patient populations. Expanding AI research into currently underserved tasks, particularly vertebral rotation assessment and skeletal maturity grading, could address clinically important needs that remain largely unmet. For non-radiographic modalities, multisite data collection and cross-dataset generalization studies will be essential before these approaches can be considered for deployment in community-based assessment programs. From a reporting perspective, broader adoption of CLAIM 2024 or equivalent standards would facilitate more rigorous evidence synthesis and move the field toward the level of transparency expected of clinical diagnostic tools.

DECLARATIONS

Acknowledgment

The illustrations used in the Graphical Abstract were sourced from Pixabay (https://pixabay.com/) and Servier Medical Art (https://smart.servier.com/). Images provided by Servier Medical Art are licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/); the original images have been modified for the purposes of this Graphical Abstract.

Authors' contributions

Conception and design of the study, methodology development, data acquisition and curation, data analysis and interpretation, and preparation of the original manuscript: Su J
Data analysis and contribution to manuscript preparation: Su J, Liu C
Critical review and editing of the manuscript: Su J, Liu C, Chen G, Wu N
Supervision and project administration: Wu N
Funding acquisition and resources: Zhang TJ, Wu N

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

During the preparation of this manuscript, the AI tool Claude (version 4.6, released 2026-02-05; Anthropic) was used solely for language editing. The tool did not influence the study design, data collection, analysis, interpretation, or the scientific content of the work. All authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

Financial support and sponsorship

This research was funded in part by the National Key Research and Development Program of China (2023YFC2507700 to Zhang TJ and Wu N); National Natural Science Foundation of China (82572698 to Zhang TJ); CAMS Innovation Fund for Medical Sciences (CIFMS, 2023-I2M-C&T-A-003 to Zhang TJ, 2024-I2M-TS-002 to Wu N, 2025-I2M-XHJC-002 to Wu N); National High Level Hospital Clinical Research Funding (2025-PUMCH-C-003 to Zhang TJ); Peking Union Medical College Hospital Talent Cultivation Program (Category A) No. ULJ11240 to Wu N; the National Key Clinical Specialty Construction Project and the Non-profit Central Research Institute Fund of the Chinese Academy of Medical Sciences (No. 2019PT320025).

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

Supplementary Materials

Supplementary Materials

REFERENCES

1. Choudhry MN, Ahmad Z, Verma R. Adolescent idiopathic scoliosis. *Open Orthop J.* 2016;10:143-54. DOI PubMed PMC
2. Labrom FR, Izatt MT, Claus AP, Little JP. Adolescent idiopathic scoliosis 3D vertebral morphology, progression and nomenclature: a current concepts review. *Eur Spine J.* 2021;30:1823-34. DOI PubMed
3. Kotwicki T, Negrini S, Grivas TB, et al. ; Members of the international Society on Scoliosis Orthopaedic and Rehabilitation Treatment (SOSORT). Methodology of evaluation of morphology of the spine and the trunk in idiopathic scoliosis and other spinal deformities - 6th SOSORT consensus paper. *Scoliosis.* 2009;4:26. DOI PubMed PMC
4. Dunn J, Henrikson NB, Morrison CC, Blasi PR, Nguyen M, Lin JS. Screening for adolescent idiopathic scoliosis: evidence report and systematic review for the us preventive services task force. *JAMA.* 2018;319:173-87. DOI PubMed PMC
5. Hui AT, Alvandi LM, Eleswarapu AS, Fornari ED. Artificial intelligence in modern orthopaedics: current and future applications. *JBJS Rev.* 2022;10. DOI PubMed
6. Tınas İ, Esmer GB. A comprehensive method for spine segmentation and scoliosis classification using YOLOv8 and curvature analysis. In: 2024 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES); 2024 Dec 11-13; Penang, Malaysia. Piscataway (NJ): IEEE; 2024. pp. 58-63. DOI
7. Goral S, Kose U. Development of a CapsNet and fuzzy logic decision support system for diagnosing the scoliosis and planning treatments via Schroth method. *IEEE Access.* 2022;10:129055-78. DOI
8. Wang H, Song S, Lu T, et al. Landmark detection network of adolescent idiopathic scoliosis by integrating the image enhancement module and vertebral center guidance. In: 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT); 2024 Mar 29-31; Nanjing, China. Piscataway (NJ): IEEE; 2024. pp. 945-9. DOI
9. Peng L, Hu Y, Zhang K, et al. Cobb angle measurement based on spine segmentation using ATT UNet 3+. *IEEE Access.* 2024;12:68545-55. DOI
10. Kim KC, Yun HS, Kim S, Seo JK. Automation of spine curve assessment in frontal radiographs using deep learning of vertebral-tilt vector. *IEEE Access.* 2020;8:84618-30. DOI
11. Maaliw RR, Susa JAB, Alon AS, et al. A deep learning approach for automatic scoliosis Cobb angle identification. In: 2022 IEEE World AI IoT Congress (AIIoT); 2022 Jun 6-9; Seattle, WA, USA. Piscataway (NJ): IEEE; 2022. pp. 111-7. DOI
12. Liu Z, Qian L, Jing W, et al. Automatic spinal curvature measurement on ultrasound spine images using Faster R-CNN. In: 2021 IEEE International Ultrasonics Symposium (IUS); 2021 Sep 11-16; Xi'an, China. Piscataway (NJ): IEEE; 2021. pp. 1-4. DOI
13. Zou L, Zhang R, Guo L. DDEW-net: automatic assessment of scoliosis through dual decoder enhancement network. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI); 2023 Apr 18-21; Cartagena, Colombia. Piscataway (NJ): IEEE; 2023. pp. 1-5. DOI
14. Shang Y, Li Q, Yang J, Dai Z, Wang X, Zhang X. Research on classification of scoliosis based on pulse signal under dynamic pressure. In: 2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC); 2023 Feb 24-26; Chongqing, China. Piscataway (NJ): IEEE; 2023. pp. 1466-9. DOI
15. Liang Y, Lv J, Li D, Yang X, Wang Z, Li Q. Accurate Cobb angle estimation on scoliosis X-Ray images via deeply-coupled two-stage network with differentiable cropping and random perturbation. *IEEE J. Biomed. Health Inform.* 2023;27:1488-99. DOI PubMed
16. Qiu Z, Yang J, Wang J. MMA-Net: multiple morphology-aware network for automated Cobb angle measurement. In: 2024 IEEE International Conference on Robotics and Automation (ICRA); 2024 May 13-17; Yokohama, Japan. Piscataway (NJ): IEEE; 2024. pp. 9139-45. DOI
17. Dagli MM, Sussman JH, Gujral J, et al. Development and validation of a keypoint region-based convolutional neural network to automate thoracic Cobb angle measurements using whole-spine standing radiographs. *Acta Neurochir.* 2025;167:227. DOI PubMed PMC

18. Zhu J, Zhou Z, Yao C. Automatic measurement of scoliosis based on an improved segmentation model. *IEEE Access.* 2023;11:65749-58. DOI
19. Zhao Y, Zhang J, Li H, Wang Q. Automatic axial vertebral rotation estimation on radiographs for adolescent idiopathic scoliosis by deep learning. *Biomed Signal Process Control.* 2024;88:105711. DOI
20. Zhang C, Wang J, He J, Gao P, Xie G. Automated vertebral landmarks and spinal curvature estimation using non-directional part affinity fields. *Neurocomputing.* 2021;438:280-9. DOI
21. Logithasan V, Wong J, Reformat M, Lou E. Using machine learning to automatically measure axial vertebral rotation on radiographs in adolescents with idiopathic scoliosis. *Med Eng Phys.* 2022;107:103848. DOI PubMed
22. Magnide E, Tchaha GW, Joncas J, et al. Automatic bone maturity grading from EOS radiographs in Adolescent Idiopathic Scoliosis. *Comput Biol Med.* 2021;136:104681. DOI PubMed
23. Zhao Y, Zhang J, Li H, Wang Q, Li Y, Wang Z. Automatic skeletal maturity grading from pelvis radiographs by deep learning for adolescent idiopathic scoliosis. *Med Biol Eng Comput.* 2025;63:1571-83. DOI PubMed
24. Zhang T, Zhu C, Zhao Y, et al. Deep learning model to classify and monitor idiopathic scoliosis in adolescents using a single smartphone photograph. *JAMA Netw Open.* 2023;6:e2330617. DOI PubMed PMC
25. Xie K, Zhu S, Lin J, et al. A deep learning model for radiological measurement of adolescent idiopathic scoliosis using biplanar radiographs. *J Orthop Surg.* 2025;20:236. DOI PubMed PMC
26. Li H, Qian C, Yan W, et al. Use of artificial intelligence in Cobb angle measurement for scoliosis: retrospective reliability and accuracy study of a mobile app. *J Med Internet Res.* 2024;26:e50631. DOI PubMed PMC
27. Wu C, Meng G, Lian J, et al. A multi-stage ensemble network system to diagnose adolescent idiopathic scoliosis. *Eur Radiol.* 2022;32:5880-9. DOI PubMed
28. Maaliw RR. SCOLIONET: an automated scoliosis Cobb angle quantification using enhanced X-ray images and deep learning models. *J Imaging.* 2023;9:265. DOI PubMed PMC
29. Wong J, Reformat M, Lou E. Applying machine learning and point-set registration to automatically measure the severity of spinal curvature on radiographs. *IEEE J Transl Eng Health Med.* 2024;12:151-61. DOI PubMed PMC
30. Liu J, Yuan C, Sun X, Sun L, Dong H, Peng Y. The measurement of Cobb angle based on spine X-ray images using multi-scale convolutional neural network. *Phys Eng Sci Med.* 2021;44:809-21. DOI PubMed
31. Du X, Wang H, Jiang L, Lv C, Xi Y, Yang H. Adjacent point aided vertebral landmark detection and Cobb angle measurement for automated AIS diagnosis. *Comput Med Imaging Graph.* 2025;121:102496. DOI PubMed
32. Lu Q, Ni L, Zhang Z, Zou L, Guo L, Pan Y. Superior performance of a center-point AI model over VFLDNet in automated Cobb angle estimation for scoliosis assessment. *Eur Spine J.* 2024;33:4710-9. DOI PubMed
33. Li Y, Peng X, Mao Q, et al. Intelligent evaluation method for scoliosis at home using back photos captured by mobile phones. *Bioengineering.* 2024;11:1162. DOI PubMed PMC
34. Zou L, Guo L, Zhang R, et al. VLTENet: a deep-learning-based vertebra localization and tilt estimation network for automatic Cobb angle estimation. *IEEE J Biomed Health Inform.* 2023;27:3002-13. DOI PubMed
35. Gadu SR, Potala CS. SCANet: automated Cobb angle measurement using deep learning model in X-ray images for adolescent idiopathic scoliosis patients. *Biomed Signal Process Control.* 2026;112:108750. DOI
36. Rahmaniar W, Suzuki K, Lin T. Auto-CA: automated Cobb angle measurement based on vertebrae detection for assessment of spinal curvature deformity. *IEEE Trans Biomed Eng.* 2024;71:640-9. DOI PubMed
37. Zhang K, Xu N, Guo C, Wu J. MPF-net: an effective framework for automated Cobb angle estimation. *Med Image Anal.* 2022;75:102277. DOI PubMed
38. Yang Y, Wang Y, Liu T, et al. Anatomical prior-based vertebral landmark detection for spinal disorder diagnosis. *Artif Intell Med.* 2025;159:103011. DOI PubMed
39. Yao Y, Yu W, Gao Y, et al. W-Transformer: accurate Cobb angles estimation by using a transformer-based hybrid structure. *Med Phys.* 2022;49:3246-62. DOI PubMed
40. Zhao Y, Zhang J, Li H, Gu X, Li Z, Zhang S. Automatic Cobb angle measurement method based on vertebra segmentation by deep learning. *Med Biol Eng Comput.* 2022;60:2257-69. DOI PubMed
41. Feng CK, Chen YJ, Dinh QT, Tran KT, Liu CY. Multi-task learning for classification and prediction of adolescent idiopathic scoliosis based on fringe-projection three-dimensional imaging. *Eur Spine J.* 2025;34:5321-33. DOI PubMed
42. González-Ruiz JM, Loayza AP, Rothstock S. Radiation-free Cobb angle estimation in adolescent scoliosis using surface topography and a linear regression model. *Spine Deform.* 2026;14:1331-9. DOI PubMed
43. Chen G, Meng N, Zhuang Y, et al. Scalable and robust artificial intelligence for spine alignment assessment: multicenter study enabled by real-time data transformation. *J Med Internet Res.* 2026;28:e78396. DOI PubMed PMC

44. Stecher N, Heinke A, Zurawski AL, Malberg H. Scoliosis assessment depending on curve magnitude based on torsobarography and machine learning. *Annu Int Conf IEEE Eng Med Biol Soc.* 2025;2025:1-5. DOI PubMed
45. Vishwakarma K, Nigam A. ScoliGaitX: a deep multi-modal fusion network for scoliosis assessment via gait video analysis. In: 2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV); 2026 Mar 6-10; Tucson, AZ, USA. IEEE; 2026. pp. 2404-13. DOI
46. Lin Y, Liu L, Ma K, Zheng Y. Seg4Reg+: consistency learning between spine segmentation and Cobb angle regression. In: de Bruijne M, Cattin PC, Cotin S, et al., Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2021. Cham: Springer; 2021. pp. 490-9. DOI
47. Huo L, Cai B, Liang P, et al. Joint spinal centerline extraction and curvature estimation with row-wise classification and curve graph network. In: de Bruijne M, Cattin PC, Cotin S, et al., Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2021. Cham: Springer; 2021. pp. 377-86. DOI
48. Zhou Z, Liang J, Peng Z, Fan C, An F, Yu S. Gait patterns as biomarkers: a video-based approach for classifying scoliosis. In: Linguraru MG, Dou Q, Feragen A, et al., Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2024. Cham: Springer; 2024. pp. 284-94. DOI
49. Zhou Z, Peng Z, Jin D, Fan C, An F, Yu S. Pose as clinical prior: learning dual representations for scoliosis screening. In: Gee JC, Alexander DC, Hong J, et al., Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2025. Cham: Springer; 2026. pp. 464-74. DOI
50. Li H, Guo Y, Jiang F, et al. Text-guided multi-instance learning for scoliosis screening via gait video analysis. In: Gee JC, Alexander DC, Hong J, et al., Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2025. Cham: Springer; 2026. pp. 649-59. DOI
51. Kuklo TR, Potter BK, Polly DW, O'Brien MF, Schroeder TM, Lenke LG. Reliability analysis for manual adolescent idiopathic scoliosis measurements. *Spine.* 2005;30:444-54. DOI PubMed
52. Morrissy RT, Goldsmith GS, Hall EC, Kehl D, Cowie GH. Measurement of the Cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error. *J Bone Joint Surg Am.* 1990;72:320-7. PubMed
53. Prestigiacomo FG, Hulsbosch MHHM, Bruls VEJ, Nieuwenhuis JJ. Intra- and inter-observer reliability of Cobb angle measurements in patients with adolescent idiopathic scoliosis. *Spine Deform.* 2022;10:79-86. DOI PubMed
54. Altaf F, Gibson A, Dannawi Z, Noordeen H. Adolescent idiopathic scoliosis. *BMJ.* 2013;346:f2508. DOI PubMed
55. Tejani AS, Klontzas ME, Gatti AA, et al. ; CLAIM 2024 Update Panel. Checklist for artificial intelligence in medical imaging (CLAIM): 2024 update. *Radiol Artif Intell.* 2024;6:e240300. DOI PubMed PMC
56. Zhu Y, Yin X, Chen Z, et al. Deep learning in Cobb angle automated measurement on X-rays: a systematic review and meta-analysis. *Spine Deform.* 2025;13:19-27. DOI PubMed PMC
57. Kumar R, Gupta M, Abraham A. A critical analysis on vertebra identification and Cobb angle estimation using deep learning for scoliosis detection. *IEEE Access.* 2024;12:11170-84. DOI
58. Goldman SN, Hui AT, Choi S, et al. Applications of artificial intelligence for adolescent idiopathic scoliosis: mapping the evidence. *Spine Deform.* 2024;12:1545-70. DOI PubMed PMC
59. Li L, Wong M. The application of machine learning methods for predicting the progression of adolescent idiopathic scoliosis: a systematic review. *Biomed Eng Online.* 2024;23:80. DOI PubMed PMC

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.

© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.