



Artificial Intelligence in Scoliosis Screening: Clinical evidence and implementation in adolescent idiopathic scoliosis

Nishant

M.S. Ortho, D-Ortho, F.A.G.E., Assistant Professor in Orthopaedic Surgery RDGM Medical College, Muzaffarpur Bihar India; Email: nishantspine@gmail.com, ORCID: 0000-0002-6206-7732

Received:- 28/06/2026, Revised:- 25/07/2026, Accepted:- 17/08/2026, Published:- 28/08/2026

Abstract

Introduction: Artificial intelligence is capable of carrying out scoliosis measurements and assisting with referral decisions by means of radiographs, photographs, surface scans, depth sensors, or ultrasound. The value of this approach in screening will depend on whether these methods are able to identify curves that are clinically relevant without leading to an excessive number of referrals or providing false reassurance.

Methods: A search of PubMed from January 2000 to 17 September 2026 was carried out to find out about the use of artificial intelligence and machine learning in the screening, diagnosis, measurement of the Cobb angle, and prediction of progression of adolescent idiopathic scoliosis. Thirty clinical studies and reviews were chosen for a structured narrative evaluation and were grouped according to clinical task, imaging modality, and validation design.

Results: The method involving automated radiographic Cobb angle measurement is supported by the most substantial body of evidence; a meta-analysis found the pooled mean absolute error to be approximately 3 degrees. Radiation-free methods have demonstrated high sensitivity in certain groups, but their performance depends on the quality of acquisition and the patients selected. A school-based study involving 5,026 pupils found that an artificial intelligence-assisted structured-light method—which included radiographic verification—had high specificity and a high negative predictive value. Although ultrasound and progression models may be useful for monitoring, concerns still exist regarding missed curves, incomplete calibration, and the lack of external validation.

Conclusion: The existing evidence indicates that task-specific artificial intelligence tools should be used under supervision. It is necessary to carry out checks on image quality, provide the option of withholding results that are uncertain, have a clinician review them, and carry out selective radiographic confirmation. However, the evidence does not support replacing clinical assessment or radiographs when treatment decisions are based on the Cobb angle.

Keywords: adolescent idiopathic scoliosis, artificial intelligence, screening, Cobb angle, surface topography

1. Introduction

Adolescent idiopathic scoliosis (AIS) is defined as a lateral curve of the spine of 10 degrees or more accompanied by vertebral rotation, having ruled out any other possible causes. The purpose of screening is to find adolescents who may gain from either being closely monitored or receiving treatment while they are growing. However, the overall benefit at the population level is still in question since both the low prevalence of the condition and the imperfection of the tests can result in a large number of unnecessary referrals and radiographs. While the US Preventive Services Task Force has stated that there is not enough evidence to evaluate the balance of benefits and harms, the Scoliosis Research Society advocates for screening in programmes that have trained personnel, clearly defined referral criteria, and access to treatment [1,2]. AI should be judged in light of these practical objectives. The fact that accurate classification can be achieved in a specialist clinic does not prove that the screening programme improves the decisions regarding referrals, reduces radiation exposure, or leads to earlier detection of curves that are actionable. The use of AI in AIS involves image analysis, diagnosis, and prediction [3]. This review looks at the application of AI for population screening, for estimating nonradiographic severity, for carrying out radiographic measurements, for predicting progression, and for monitoring.

2. Materials and Methods

The search on PubMed combined the term AIS with those relating to artificial intelligence, machine learning, deep learning, screening, diagnosis, Cobb angle, surface topography, ultrasound, photography, and progression; the time frame covered was from January 2000 up to 17 September 2026 and the reference lists of recent scoping and systematic reviews were also looked at. Thirty reports were chosen to address the main clinical tasks, modalities, and validation designs. It should be noted that this is a structured clinical review, not a systematic review or a new meta-analysis.

The appraisal of the study took into account the recruitment procedure, the reference standard, the separation of the training and test data, the sample size, external validation, and the way in which poor-quality inputs were handled. The diagnostic results were compared with both the prevalence in the population and the decision threshold. With regard to the progression models, the

assessment also looked at when the predictors became available, calibration, the length of follow-up, and the effects of treatment. Technical accuracy and actual clinical benefit were evaluated separately.

3. Results

3.1 The clinical tasks involved in artificial intelligence screening

3.1.1 The detection of population cases and the process of referring them

Screening models can identify adolescents who might need to be examined or have radiography carried out; these models make use of various inputs such as clinical variables, back photographs, or three-dimensional surface scans and then give a recommendation for referral, a probability of disease, or a severity category. Since a curve might be missed and this could delay the assessment during growth, sensitivity is important when selecting a threshold. Specificity is also relevant: in a population where the prevalence is low, even a small false-positive rate can result in a substantial number of unnecessary referrals and clinical-variable models have also been assessed for the detection and severity evaluation of scoliosis [4,5].

A scoping review conducted in 2026 summarized 26 studies on AI screening involving over 330,000 participants and found a median area under the receiver operating characteristic curve (AUC) of 0.94, an average accuracy of 0.91, and an average Cobb angle error of 3.1 degrees. Eight of the studies were multicenter and seven were prospective; however, none of them directly compared an AI approach with the established Scoliosis Research Society screening method, and none evaluated the cost-effectiveness [6]. While the results provide information on model performance, they do not address the benefits of deploying the method in a population setting.

A study carried out in a school assessed an AI-assisted structured-light system among 5,026 pupils. Of the 185 pupils who tested positive upon rescreening, 160 were found to have scoliosis, and two additional cases were detected in 300 pupils who had initially tested negative and then had their results verified. Using inverse probability weighting, the sensitivity was found to be 0.832, the specificity 0.995, the positive predictive value 0.865, and the negative predictive value 0.993 [7]. While the analysis took into account the issue of partial verification, further research is still required in schools

comprising different populations and under different acquisition conditions.

3.1.2 Nonradiographic severity estimation

Spinal deformity is estimated in this way by looking at the shape of the trunk. It allows for repeated imaging without the use of ionizing radiation and might therefore enable evaluations to be carried out outside specialist clinics. Yet the surface appearance also depends on body build, posture, the shape of the muscles, pelvic position, and vertebral rotation. Because of this the link between surface asymmetry and the Cobb angle is not perfect. Models which have been developed for patients with already established deformities must be tested separately in healthy adolescents and in those who have postural asymmetry.

A model based on a single photograph, which had been trained on 1,780 patients and was then tested prospectively in 378 cases, obtained AUC values of 0.839 for giving a recommendation regarding followup and 0.902 for suggesting surgery. The sensitivity relating to follow-up was 84.9%, this figure being higher than that of the two surgeon comparators; however, the prediction of progression was less accurate [8]. An earlier study using back images also found classification performance to be above that of specialists, although data on its implementation in a general population were limited [9]. A more recent network using bare-back images reached an AUC of 0.91, its performance lying between that of senior and less experienced doctors [10].

Three-dimensional methods give more surface information than a single photograph. In a study involving five centres and using depth sensors, the estimates correlated with the radiographic Cobb angle at 0.91; the mean absolute error was 4.0 degrees and accuracy for curves of 10 degrees or more was 94% [11]. A prospective study that used smartphone surface topography in two countries had 236 participants and found 100% sensitivity and 89% specificity at the 10-degree threshold, although there were limitations due to selection and reproducibility [12]. In another application study, 20 of the 125 scans were excluded because of poor quality. Of the 51 test scans, the mean absolute error was 5.9 degrees. Sensitivity and specificity were about 0.96 at 10 degrees, but sensitivity dropped to 0.71 at 25 degrees and to 0.67 at 40 degrees [13]. Other surface-based studies have also found lower performance. In a study involving 900 participants, rasterstereographic prediction yielded a mean absolute error of 6.1 degrees

and an accuracy rate of 59% for severity classification [14]. An ensemble model built from 654 scans showed a test error of 6.13 degrees, even though it had good precision and specificity in certain severity groups [15]. In a group that included individuals with normal posture and other deformities, a symmetry-sensitive videorasterstereographic network obtained an accuracy of 0.744, a sensitivity of 0.726, and a positive predictive value of 0.417 [16]. Therefore, patient selection and acquisition failure rates are important factors when interpreting the reported accuracy.

3.1.3 The automatic measurement of the Cobb angle

Radiographic Cobb measurements are more advanced than radiation-free screening methods, since both the algorithm and the reference reader examine the same underlying radiograph. In most cases, the systems detect vertebral landmarks or divide up the vertebrae, choose the end vertebrae, and then compute the curve angles. Clinical failures that are significant include failing to detect curves, selecting the wrong end vertebrae, and a decrease in accuracy when dealing with severe deformities, atypical anatomy, or unfamiliar imaging protocols. Earlier research had also involved the automation of radiological analysis of spinal deformity by using predicted anatomical landmarks [17].

The systematic review involved 50 studies, 17 of which were subjected to a meta-analysis. The pooled circular mean absolute error was 2.99 degrees and showed substantial heterogeneity; the pooled errors were 2.40 degrees for segmentation-based methods and 3.31 degrees for landmark-based methods [18]. A study involving four centres and comprising 484 radiographs with 1,054 consensus curves found a curve detection rate of 94.0% and a mean absolute error of 3.01 degrees [19]. In the seven-hospital study, realtime image transformation was employed to deal with the variations between centres; the mean error stayed below 4 degrees and the R^2 value went above 0.90 when assessed on external datasets [20].

The validation results also vary from one system to another. In one study involving children, 346 out of 352 curves were detected, 91% of the measurements fell within the study's clinical acceptance range, and an intraclass correlation coefficient (ICC) of 0.92 was reported [21]. A network based on keypoint regions obtained a mean absolute error of 2.22 degrees and an ICC of 0.98 [22]. In an external comparison, the

mean absolute difference was 2.36 degrees for one system and 6.51 degrees for the other [23]. Previous segmentation research had found errors less than 3 degrees and correlations greater than 0.95 [24]. Local validation is therefore necessary even if the published results seem clinically acceptable.

A measurement system must refrain from producing a result if the image quality or the anatomy makes it unreliable. The coronal-balance system of 2026 included such a safety feature and gave errors in the range of 0.86 to 2.45 mm for shoulder height, pelvic obliquity, and C7 offset [25]. The number of results that are withheld and the review these necessitate should be taken into account together with the accuracy of the measurements.

3.1.4 The use of ultrasound in combination with other assessment methods

Three-dimensional ultrasound is able to show the posterior spinal anatomy without the use of ionizing radiation, but both the acquisition and the interpretation require training. An automated method was evaluated on 144

Ultrasonographs were used and 125 out of the 144 curves that had been manually identified were detected. The method obtained an ICC of 0.81, and 75 per cent of the measurements were within 5 degrees of the manual result [26]. Since it fails to detect some of the curves, it cannot be used as a replacement for radiographs. However, it could be used to aid in monitoring within a specific acquisition protocol which defines when further imaging is needed.

Multimodal models use clinical information together with global radiographs, regional vertebral characteristics, or skeletal maturity. A composite model based on 1,870 patients reached an AUC of 0.84, with a sensitivity of 80.9% and a specificity of 83.6% for detecting progression, a result that outperformed the performance of its individual modality components [27]. Although using a number of different data sources can improve the accuracy of the prediction, it also means that the model's performance depends on the availability of the input data and on local acquisition practices. The rules regarding missing data and the role of each modality must be clearly reported.

3.2 Predicting and monitoring progression

Progression prediction involves estimating future changes once a curve has been identified. In order to carry out an evaluation it is necessary to specify the prediction time, obtain adequate follow-up, and have

consistent documentation of the treatment. Calibration must be carried out in order to check whether the predicted risks correspond to the actual outcomes. Bracing makes interpretation more difficult since it might affect progression and is more frequently prescribed to patients who are considered to be at higher risk.

In a group of 1,119 female patients, the curve-specific models used to predict progression beyond 25 degrees gave AUC values of 0.94 for the proximal thoracic curve, 0.89 for the main thoracic curve, and 0.84 for thoracolumbar or lumbar curves [28]. A different multicenter model had an external-validation AUC of 0.896, but it included as important predictors the changes between visits; as a result, its use depends on having follow-up information and it cannot be applied at the first screening appointment [29]. Regression models developed from data on 201 girls had mean absolute errors of 4.22 degrees after one year, 5.24 degrees after two years, and 5.99 degrees at the last visit [30].

Simple prognostic models are still useful as points of comparison. In a prospective cohort consisting of 127 patients, a model that took into account the Risser stage, the Cobb angle, the patient's reported appearance, and treatment exposure had a concordance of 0.79. Machine-learning approaches did not clearly exceed the performance of the Cox model [31]. Extra complexity should only be accepted if it improves prediction or decision-making sufficiently to compensate for the additional data and the need for implementation.

4. Discussion

4.1 The clinical interpretation of the evidence

Table 1 sets out the main clinical applications, the performance that has been reported, and the implications for supervised use.

Image-language systems that are general in their application need to be evaluated separately. A prospective study involving 97 adolescents revealed a high number of false-positive results and Cobb angle estimates that were clinically unacceptable; in one case, the system had a specificity of zero when it came to distinguishing between postural asymmetry and scoliosis [32]. The fact that someone is fluent in talking about scoliosis does not mean that they are capable of assessing a clinical photograph.

4.2 Clinical implementation

4.2.1 A supervised pathway

A supervised pathway begins with

standardized acquisition and a check of image quality. The model should be able to withhold an uncertain result. Positive findings, disagreement with clinical assessment, and unusable images require clinician review. Radiographs should be obtained selectively. Thresholds should match a specific clinical action. The threshold used to rule out a condition should have high sensitivity, and the negative predictive value should be understood in the context of the expected prevalence. When it comes to radiography, the threshold must also take into account the possibility of unnecessary imaging. Separate thresholds might be required for repeat acquisition, examination, radiography, expedited referral, and routine follow-up. Accuracy as a whole is not sufficient to describe these consequences. Expedited referral, and routine follow-up. Overall, failure reports ought to specify the acquisition or analytical problem involved, for example, obscuration caused by clothing or hair, poor positioning, lighting conditions, motion, an incomplete field of view, uncertain segmentation, or incompatibility of the device. Eliminating low-quality images from the analysis may lead to an overstatement of performance. In reality, for each failed acquisition either another scan has to be carried out or a manual review is needed, or the patient remains unevaluated. Studies should include information about these outcomes and the proportion of cases in which the system chooses not to act. When AI is used to carry out measurements or triage, it is still up to the clinicians to make the diagnosis and manage the case. The review interfaces should show the source image, the landmarks or segmentation, the uncertainty, the previous measurements, and any reason why a result was withheld. Providing this information enables the reviewer to identify any errors that might have incorrectly reassured a family or led to a change in treatment. The supervised method proposed and the responses needed at each stage are set out in Table 2.

4.3 Bias generalizability and reporting

Cohorts from tertiary centres tend to have a higher number of cases of scoliosis and more severe curves than those from school or primary care settings. The findings from such cohorts may not be applicable to screening, especially with regard to positive predictive value. When carrying out external validation, consecutive patients or a population-based sample should be recruited from the setting in question and at least a random portion of

the negative screenings should be verified. Simply recruiting from multiple centres is not enough if data from the same patients or the same sites are included in both the training and the test sets; evaluation should instead involve separating the data at the patient level and, where appropriate, using held-out centres or prospective cohorts.

The reference measurement is also subject to uncertainty. The consensus readings and the quantified interobserver variation aid in the interpretation of algorithm error. Studies must keep all relevant curves and should report the missed curves separately from the angular error. Surface scans and reference radiographs

They should be acquired close enough to one another so as to avoid the disagreements that would result from interval progression. Furthermore, prognostic studies also require calibration, clearly defined time horizons, information on treatment exposure, decision thresholds, and a description of the missing data.

Authors must adhere to the relevant reporting guidelines, such as CLAIM and TRIPOD+AI [33,34]. The reports have to include enough detail to allow the recruitment, acquisition, reference labelling, data separation, preprocessing, and threshold selection to be reproduced. Information should also be provided regarding the model versions, the results from subgroups, how failures were handled, and the confidence intervals.

Table 3 sets out the minimum reporting requirements needed for assessing these systems.

4.4 Considerations from an ethical and practical point of view

Photographs of minors and surface scans could be identified, so it is necessary for programmes to offer age-appropriate assent, secure transfer, parental permission when needed, restricted access, and clearly defined retention periods. The policies must state whether the images can be used for further training of models. Cropping or de-identification should be carried out as early as possible and consumer applications should not provide a default retention period or allow for unrelated reuse.

The overall accuracy does not mean that the performance can be considered comparable among different groups. Photographic methods can be influenced by lighting and skin tone, as well as by body habitus, clothing, and cultural acceptability when it comes to acquisition. For home monitoring, access to suitable devices is a necessary factor. And for pre-specified subgroup

analyses, it is necessary to have adequate sample sizes as well as a plan to deal with clinically important differences.

Economic assessments must take into account the cost of equipment, the time taken by staff, repeat imaging, licensing, data infrastructure, confirmatory investigations, and referrals. They should also consider the possibility of missed or delayed diagnoses. Relying solely on the cost per screened child is not enough; it is better to assess the value of a pathway in terms of the cost per correctly identified actionable curve and the cost per unnecessary radiograph avoided.

4.5 Research priorities

Compare trained conventional screening with AI-assisted screening in prospective cluster studies carried out in schools or in primary care.

Check a random sample of the negative screens and take the sampling method into consideration when estimating diagnostic performance.

Set the referral thresholds in advance and report the number of radiographs avoided, the number of cases in which false reassurance was given, the time taken to reach specialist assessment, and the cost.

Before it is more widely deployed, assess its performance across different devices, regions, body types, skin tones, and acquisition operators.

Report all decisions regarding quality control, any outputs that were withheld, and the clinical consequences of each failed acquisition.

For progression models it is necessary to provide the report calibration, treatment exposure, prediction horizon, and external validation together with discrimination.

Find out if having AI assistance in monitoring will allow the radiographic frequency to be reduced without delaying bracing or getting a referral to a specialist.

4.6 The limitations of this review

The review has chosen representative studies from PubMed and did not carry out duplicate systematic screening across multiple databases, and as a result engineering papers, publications in languages other than English, preprints, and commercial validation reports may be underrepresented. The bibliographic information for reports which were initially indexed before publication might change. Since there were differences among the studies with respect to the populations, imaging modalities, thresholds, and outcomes, a new pooled estimate could not be produced.

Table 4 includes selected study designs,
European Journal of Molecular Pharmacy

performance estimates, and limitations.

5. Conclusion

AI can be used to carry out radiographic measurements and might help with scoliosis triage provided that the acquisition is controlled. However, the evidence supporting its use in independent population screening is not well established. Its adoption should be based on local validation, the ability to reliably deal with poor-quality inputs, and having clinicians confirm the findings whenever they have an impact on patient management. Prospective comparative studies need to determine whether these systems detect actionable curves earlier, decrease the number of unnecessary imaging procedures and referrals, and continue to be effective among the different populations for which they were intended.

Statements Acknowledgement None.

Conflicts of Interest None.

Funding/Support Sources

This review was not financed by any external sources.

Author Contribution

Dr. Nishant was the sole and corresponding author and carried out the drafting, preparation and proofreading.

Statement of Ethics None.

Ethical Approval

Does not apply. This article is a review of the literature that has already been published.

Informed Consent Statement

This does not apply; no participants were recruited and no individual patient data are given in this review.

Artificial Intelligence (AI) Disclosure Statement

ChatGPT from OpenAI was employed to help with editing the manuscript, renumbering the references, and formatting the document.

References

1. US Preventive Services Task Force. Screening for Adolescent Idiopathic Scoliosis US Preventive Services Task Force Recommendation Statement. JAMA. 2018;319(2):165-172. doi:10.1001/jama.2017.19342. PMID:29318284.
2. Labelle H, Richards SB, De Kleuver M, Grivas TB, Luk KDK, Wong HK, et al. Screening for adolescent idiopathic scoliosis: an information statement by the Scoliosis Research Society international task force. Scoliosis. 2013;8:17. doi:10.1186/1748-7161-8-17. PMID:24171910.

3. Goldman SN, Hui AT, Choi S, Mbamalu EK, Tirabady P, Eleswarapu AS, et al. Applications of artificial intelligence for adolescent idiopathic scoliosis: mapping the evidence. *Spine Deform.* 2024;12(6):1545-1570. doi:10.1007/s43390-024-00940-w. PMID:39153073.
4. Negrini F, Cina A, Ferrario I, Zaina F, Donzelli S, Galbusera F, et al. Developing a new tool for scoliosis screening in a tertiary specialistic setting using artificial intelligence: a retrospective study on 10,813 patients: 2023 SOSORT award winner. *Eur Spine J.* 2023;32(11):3836-3845. doi:10.1007/s00586-023-07892-1. PMID:37650978.
5. Lv Z, Lv W, Wang L, Ou J. Development and validation of machine learning-based models for prediction of adolescent idiopathic scoliosis: A retrospective study. *Medicine (Baltimore).* 2023;102(14):e33441. doi:10.1097/MD.00000000000033441. PMID:37026913.
6. Zhou Z, He X. Use of Artificial Intelligence in Screening for Adolescent Idiopathic Scoliosis: A Scoping Review. *Stud Health Technol Inform.* 2026;337:246-250. doi:10.3233/SHTI260781. PMID:42290458.
7. Cao Y, Li F, Zhang Y, Gu W, Zheng X, Li M, et al. Prevalence and factors associated with adolescent idiopathic scoliosis in Shanghai, China: an artificial intelligence-assisted screening approach. *Front Public Health.* 2026;14:1832421. doi:10.3389/fpubh.2026.1832421. PMID:42422709.
8. Zhang T, Zhu C, Zhao Y, Zhao M, Wang Z, Song R, et al. Deep Learning Model to Classify and Monitor Idiopathic Scoliosis in Adolescents Using a Single Smartphone Photograph. *JAMA Netw Open.* 2023;6(8):e2330617. doi:10.1001/jamanetworkopen.2023.30617. PMID:37610748.
9. Yang J, Zhang K, Fan H, Huang Z, Xiang Y, Yang J, et al. Development and validation of deep learning algorithms for scoliosis screening using back images. *Commun Biol.* 2019;2:390. doi:10.1038/s42003-019-0635-8. PMID:31667364.
10. Duan X, Wang L, Tao Z, Zhu M, You D, Gao Y, et al. EDRNet-based adolescent idiopathic scoliosis screening method using bare-back images. *BMC Med Inform Decis Mak.* 2026;26(1):328. doi:10.1186/s12911-026-03639-2. PMID:42304329.
11. Kokabu T, Kanai S, Kawakami N, Uno K, Kotani T, Suzuki T, et al. An algorithm for using deep learning convolutional neural networks with three dimensional depth sensor imaging in scoliosis detection. *Spine J. European Journal of Molecular Pharmacy* 2021;21(6):980-987. doi:10.1016/j.spinee.2021.01.022. PMID:33540125.
12. Martrenchar P, Caille A, Dina A, Odent T. A Preliminary Report: Screening Idiopathic Scoliosis (IS) With a Smartphone Application. *J Pediatr Orthop.* 2026;46(9):e842-e846. doi:10.1097/BPO.0000000000003228. PMID:41747211.
13. Parent S, Roy-Beaudry M, Dufresne J, Imbeault R, Barchi S, Beauséjour M. Artificial intelligence-driven 3D surface-topography app for screening and monitoring adolescent scoliosis: early results from a single institution. *Spine Deform.* 2026;14(3):851-858. doi:10.1007/s43390-026-01282-5. PMID:41615635.
14. Minotti M, Negrini S, Cina A, Galbusera F, Zaina F, Bassani T. Deep learning prediction of curve severity from rasterstereographic back images in adolescent idiopathic scoliosis. *Eur Spine J.* 2024;33(11):4164-4170. doi:10.1007/s00586-02308052-1. PMID:38055037.
15. Hassan M, Gonzalez Ruiz JM, Mohamed N, Burke TN, Mei Q, Westover L. Ensemble learning of deep CNN models and two stage level prediction of Cobb angle on surface topography in adolescents with idiopathic scoliosis. *Med Eng Phys.* 2025;140:104332. doi:10.1016/j.medengphy.2025.104332. PMID:40436505.
16. Jochim T, Wilhelm P, Tschirschky A, Żurawski A, Heinke A, Stecher N, et al. Extending convolutional neural networks to detect differences in symmetry in videorasterstereographic back scans with the aim to improve screening for adolescent idiopathic scoliosis. *Eur Spine J.* 2026;35(2):414-421. doi:10.1007/s00586-02509520-6. PMID:41207964.
17. Galbusera F, Niemeyer F, Wilke HJ, Bassani T, Casaroli G, Anania C, et al. Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach. *Eur Spine J.* 2019;28(5):951-960. doi:10.1007/s00586-019-05944-z. PMID:30864061.
18. Zhu Y, Yin X, Chen Z, Zhang H, Xu K, Zhang J, et al. Deep learning in Cobb angle automated measurement on X-rays: a systematic review and meta-analysis. *Spine Deform.* 2025;13(1):19-27. doi:10.1007/s43390-024-00954-4. PMID:39320698.
19. Casabó-Vallés G, Aldecoa R, Pérez-Machado G, Egea-Gámez RM, Sánchez-Raya J, Vilalta-Vidal I, et al. A fully autonomous AI system for accurate and reproducible Cobb angle measurement in adolescent idiopathic

scoliosis: a multicenter study.

Spine J. 2026;26(7):1391-1403.
doi:10.1016/j.spinee.2026.01.010.
PMID:41506453.

20. Chen G, Meng N, Zhuang Y, Chen Z, Bian Z, Gong Z, 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:10.2196/78396. PMID:41861366.

21. Wong JC, Reformat MZ, Parent EC, Stampe KP, Southon Hryniuk SC, Lou EH. Validation of an artificial intelligence-based method to automate Cobb angle measurement on spinal radiographs of children with adolescent idiopathic scoliosis. Eur J Phys Rehabil Med. 2023;59(4):535-542.
doi:10.23736/S1973-9087.23.08091-7.
PMID:37746786.

22. Dagli MM, Sussman JH, Gujral J, Budihai BR, Kerr M, Yoon JW, 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 (Wien). 2025;167(1):227.
doi:10.1007/s00701-025-06645-x.
PMID:40847001.

23. Ni L, Zhang Z, Zou L, Wang J, Guo L, Qian W, et al. Clinical Application of Automatic Assessment of Scoliosis Cobb Angle Based on Deep Learning. Curr Med Imaging. 2024;20:e15734056278130.
doi:10.2174/0115734056278130231218073650. PMID:38415463.

24. 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(3):809-821.
doi:10.1007/s13246-021-01032-z.
PMID:34251603.

25. Wang Z, Zhang Y, Xue F, Maimaiti A, Li Y, Zhang J, et al. Automated quantification of coronal balance in spinal deformity: a safety-aware clinical workflow. Sci Rep. 2026;16(1):16665. doi:10.1038/s41598-026-47402-z. PMID:41951866.

26. Wong J, Reformat M, Parent E, Lou E. Validity and accuracy of automatic cobb angle measurement on 3D spinal ultrasonographs for children with adolescent idiopathic scoliosis: SOSORT 2024 award winner. Eur Spine J. 2025;34(5):16221630.
doi:10.1007/s00586-024-08376-6.
PMID:38987512.

27. Wang H, Zhang T, Zhang C, Shi L, Ng SY, Yan HC, et al. An intelligent composite model

incorporating global / regional X-rays and clinical parameters to predict progressive adolescent idiopathic scoliosis curvatures and facilitate population screening. EBioMedicine. 2023;95:104768.

doi:10.1016/j.ebiom.2023.104768.
PMID:37619449.

28. Ohyama S, Maki S, Kotani T, Ogata Y, Sakuma T, Iijima Y, et al. Machine Learning Algorithms for Predicting Cobb Angle Beyond 25° in Female Adolescent Idiopathic Scoliosis Patients. Spine (Phila Pa 1976). 2026;51(18):E455-E464.
doi:10.1097/BRS.0000000000004986.
PMID:38475972.

29. Fang X, Weng T, Zhang Z, Gong W, Zhang Y, Wang M, et al. Progression risk of adolescent idiopathic scoliosis based on SHAP-Explained machine learning models: a multicenter retrospective study. BMC Musculoskelet Disord. 2025;26(1):598.
doi:10.1186/s12891-025-08841-3.
PMID:40615838.

30. Kubota S, Kawaguchi K, Yanagida H, Yamaguchi T. Prediction of Cobb angle progression in female patients with mild-to-moderate adolescent idiopathic scoliosis. Eur Spine J. 2026;35(3):1061-1067.
doi:10.1007/s00586-025-09546-w.
PMID:41207960.

31. Dufvenberg M, Charalampidis A, Diarbakerli E, Öberg B, Tropp H, Ahl AA, et al. Prognostic model development for risk of curve progression in adolescent idiopathic scoliosis: a prospective cohort study of 127 patients. Acta Orthop. 2024;95:536-544.
doi:10.2340/17453674.2024.41911.
PMID:39287215.

32. Aydin C, Duygu OB, Karakas AB, Er E, Gokmen G, Ozturk AM, et al. Clinical Failure of General-Purpose AI in Photographic Scoliosis Assessment: A Diagnostic Accuracy Study. Medicina (Kaunas). 2025;61(8):1342.
doi:10.3390/medicina61081342.
PMID:40870387.

33. Mongan J, Moy L, Kahn CE Jr. Checklist for Artificial Intelligence in Medical Imaging CLAIM a guide for authors and reviewers. Radiol Artif Intell. 2020;2(2):e200029.
doi:10.1148/ryai.2020200029.

34. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378. PMID:38626948.

Table 1. Summary of clinical applications

Application	Reported performance	Clinical interpretation
Radiographic Cobb measurement	Typical error near 2 to 4 degrees in stronger studies	Appropriate for assisted measurement with clinician confirmation and local validation
Photographic screening	Good discrimination in selected datasets	Potential triage tool; sensitive to posture, acquisition conditions, and referral population
Application	Reported performance	Clinical interpretation
Three-dimensional surface screening	High sensitivity in some prospective studies with variable angle error	Potentially useful for community screening and monitoring when quality control is enforced
Ultrasound automation	Good reliability but missed curves remain	Adjunct for radiation reduction rather than a stand-alone diagnostic replacement
Progression prediction	Moderate to high discrimination in several cohorts	Requires a defined prediction horizon, calibration, treatment accounting, and external validation
General-purpose multimodal models	Prospective photographic testing showed clinically unacceptable failure	Do not use for diagnosis or reassurance without task-specific development and validation

Summary of the applications discussed in the manuscript. Reported performance comes from different populations and validation designs and should not be interpreted as a pooled or head-to-head comparison.

Table2: Proposed supervised pathway for AI-assisted scoliosis screening

Stage	Information to review	Required response
Acquisition	Standard view, acquisition metadata, and image quality score	Reject or repeat incomplete, obscured, or poorly positioned scans
AI screening	Probability, category, threshold, and uncertainty	Allow abstention and preserve the original image
Clinical review	Symptoms, examination, trunk rotation, and agreement with the model	Escalate discordance or neurologic findings
Radiographic confirmation	Curves, end vertebrae, Cobb angle, and skeletal maturity	Clinician verifies landmarks and missed curves
Longitudinal comparison	Change from prior study with acquisition compatibility	Distinguish measurement variation from true progression
Program monitoring	False negatives, false positives, failures, subgroup performance, and drift	Recalibrate or suspend use when predefined limits are exceeded

Proposed clinical workflow synthesized from the review; this pathway has not been prospectively validated as a complete screening program. AI, artificial intelligence.

Table3: Minimum reporting requirements

Domain	Reporting requirements
Population	Setting, recruitment dates, eligibility criteria, prevalence, curve distribution, age, sex, skeletal maturity, and treatment status

Acquisition	Device, software, view, positioning, clothing protocol, operator training, and quality failures
Reference standard	Number and expertise of readers, consensus process, timing, and handling of multiple curves
Dataset structure	Patient-level separation, center-level separation, external validation, and handling of repeat images
Performance	Sensitivity, specificity, predictive values, likelihood ratios, angular error, calibration, confidence intervals, and failure rate
Clinical utility	Referral yield, radiographs avoided, time, cost, acceptability, delayed diagnoses, and treatment consequences
Equity and safety	Performance by skin tone, body habitus, sex, age, curve type, device, and site; monitoring for drift

Reporting domains synthesized for studies of artificial intelligence in scoliosis screening and monitoring.

Table4: Selected studies and their limitations

Study	Design	Application	Reported result	Main limitation
Zhu et al 2025 [18]	Systematic review and meta-analysis	Radiographic Cobb measurement	Pooled error 2.99 degrees with high heterogeneity	Model architecture and validation methods varied
Casabó-Vallés et al 2026 [19]	Multicenter diagnostic study	Autonomous radiographic measurement	94.0% curve detection and 3.01 degree error	Retrospective tertiarycenter images
Martrenchar et al 2026 [12]	Prospective multicenter study	Smartphone surface screening	100% sensitivity and 89% specificity at 10 degrees	Selected cases and controls with no reproducibility assessment
Parent et al 2026 [13]	Single-center validation	Smartphone 3D surface topography	5.9 degree error with high performance at 10 degrees	Twenty poor-quality scans excluded and small test set
Cao et al 2026 [7]	School crosssectional study	Structured-light 3D screening	Adjusted sensitivity 0.832 and specificity 0.995	Only a sample of screen-negative pupils had radiographs
Chen et al 2026 [20]	Multicenter study	Cross-site radiographic robustness	Mean error under 4 degrees and R squared above 0.90	Retrospective dataset and transformationspecific workflow
Wong et al 2025 [26]	Ultrasound validation	Automated 3D ultrasound angle	ICC 0.81 and 75% within 5 degrees	Nineteen of 144 curves were not detected
Zhang et al 2023 [8]	Prospective diagnostic testing	Singlephotograph classification	AUC 0.839 for followup and 0.902 for surgery consideration	Clinic population and lower progression performance

Kokabu et al 2021 [11]	Five-center cohort	Depth-sensor imaging	4.0 degree error and 94% accuracy at 10 degrees	Nonrandomized suspected-scoliosis sample
Minotti et al 2024 [14]	Train-test study	Rasterstereo-graphic severity	6.1 degree error and 59% severity accuracy	Limited readiness for replacement of radiographs
Negrini et al 2023 [4]	Retrospective tertiary cohort	Clinical parameter triage	Accuracy 74% to 84% across thresholds	Referral-center spectrum and internal split
Aydin et al 2025 [32]	Prospective diagnostic study	General-purpose image models	Clinically unacceptable estimates and extreme false positives	Small cohort but direct evidence against unvalidated consumer use

Study	Design	Application	Reported result	Main limitation
Wang et al 2023 [27]	Multimodal validation	Progression prediction	AUC 0.84 with 80.9% sensitivity	Requires radiographs and complex inputs
Fang et al 2025 [29]	Multicenter retrospective study	Interpretable progression model	External-validation AUC 0.896	Important predictors included interval change after followup
Dufvenberg et al 2024 [31]	Prospective prognostic cohort	Progression prediction	Concordance 0.79 with conventional Cox model	Small cohort and treatment-related complexity
Zhou and He 2026 [6]	Scoping review	AI screening evidence	High median performance across 26 studies	Few prospective multicenter comparisons and no cost-effectiveness studies

Selected studies discussed in the review. Reference numbers correspond to the manuscript reference list. AI, artificial intelligence; AUC, area under the receiver operating characteristic curve; ICC, intraclass correlation coefficient; 3D, three-dimensional.