DEEP NEURAL NETWORKS FOR SCOLIOSIS DETECTION AND ANALYSIS: A LITERATURE REVIEW

Dijana Stojić, Vanja Luković, Dejan Vujičić

University of Kragujevac, Faculty of Technical Sciences in Čačak, Svetog Save 65, 32000 Čačak,Serbia, dijana.stojic@ftn.kg.ac.rs,

REVIEW SCIENTIFIC PAPER

Received: 16 July 2026.
Accepted: 01 November 2026.
Published: 16 November 2026.
DOI: https://doi.org/10.63395/STEDConf15022026146S317

Corresponding Author: Dijana Stojić, University of Kragujevac, Faculty of Technical Sciences in Čačak, Svetog Save 65, 32000 Čačak,Serbia, dijana.stojic@ftn.kg.ac.rs,

Copyright © 2026 Dijana Stojić, et al.; published by UNIVERSITY PIM. This work licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.

Citation: Stojić, J., et al. (2026). Deep neural networks for scoliosis detection and analysis: A literature review. In Kojić, D. & Sredić. D. (eds.) *Conference proceedings, XV International Conference on Social and Technological Development: STED 2026* (pp. 317-329). Trebinje, B&H: University PIM.

ABSTRACT

Scoliosis affects 2–3% of the population, and manual measurement using the Cobb angle — the gold standard since 1948 — is subject to inter-observer variability and too slow for use in mass screening programs.

This paper presents a systematic review of the application of deep neural networks (DNNs) in the detection and analysis of scoliosis through an analysis of the relevant literature published between 2015 and 2024, using the PubMed, IEEE Xplore, Scopus, and Web of Science databases. A range of architectures is considered — from classical convolutional neural networks and encoder–decoder models to Transformer and hybrid CNN–Transformer approaches — together with the key datasets, evaluation metrics, and open problems in the field.

Technical progress is clearly visible and measurable: recent literature shows that hybrid CNN–Transformer models achieve a Cobb angle estimation error that approaches the level of disagreement between two experienced radiologists. Nevertheless, most published models have been tested exclusively on data from the institution in which they originated, which raises serious concerns regarding external validity — it remains unclear whether these models would achieve the same performance in other hospitals, with different imaging equipment and different patient populations. This is one of the key open questions in the field and an important direction for future research.

Keywords: deep learning, convolutional neural networks, scoliosis, Cobb angle, medical image analysis, literature review

INTRODUCTION

Scoliosis is a three-dimensional deformity of the spinal column in which the spine bends laterally by more than 10 degrees, as measured by the Cobb angle on a radiograph. In addition to the curvature itself, vertebral rotation is present, as well as changes in the natural curves of the

spine when viewed from the side. This condition occurs in approximately 2–3% of the population, and its most common form is adolescent idiopathic scoliosis (AIS), which accounts for about 80% of all scoliosis cases [1]. If not recognized and treated in time, progressive scoliosis can lead to functional and health complications, including chronic pain, reduced respiratory function, and a significant deterioration in quality of life.

The gold standard in scoliosis diagnosis, manual measurement of the Cobb angle, has been used since 1948. Although widely accepted in clinical practice, this method has certain limitations: inter-observer variability can reach up to 7 degrees [2], and the procedure itself is time-consuming and impractical for use in mass screening programs. For this reason, automating Cobb angle measurement is a clinically and methodologically justified goal.

Over the past decade, deep learning methods have undergone significant development and found wide application in medical image analysis [3]. In certain well-defined tasks and under controlled research conditions, studies have shown that deep learning–based systems can achieve performance comparable to clinicians, including the diagnosis of diabetic retinopathy [4], the classification of skin cancer [5], and the detection of pulmonary nodules on CT scans [6]. The application of DNNs in spinal analysis has grown significantly since 2017. The aim of this paper is to review, classify, and critically analyze the key approaches and results published in this area, to identify existing gaps in the literature, and to propose directions for future research. The review is based on an analysis of relevant papers published between 2015 and 2024, retrieved from the PubMed, IEEE Xplore, Scopus, and Web of Science databases.

LITERATURE REVIEW METHODOLOGY

The review was conducted by searching the PubMed, IEEE Xplore, Scopus, and Web of Science databases for papers published between 2015 and 2024. Combinations of keywords were used that cover scoliosis-related terms (scoliosis, Cobb angle, adolescent idiopathic scoliosis, spinal deformity) and deep learning methods (deep learning, convolutional neural network, transformer, segmentation, landmark detection). The review included peer-reviewed papers in English that apply deep neural networks to the tasks of detection, segmentation, classification, or Cobb angle estimation in scoliosis. Papers that do not use deep learning methods, that do not concern the spinal column, and short conference abstracts without sufficient methodological detail were excluded. After removing duplicates and checking eligibility based on title, abstract, and full text, 15 papers were included in the qualitative synthesis. Given the pronounced heterogeneity of datasets, evaluation metrics, and annotation protocols, a narrative (qualitative) synthesis was applied instead of a formal meta-analysis.

CLINICAL CONTEXT

To understand the role and potential of deep neural networks in scoliosis analysis, it is first necessary to consider the clinical context of the diagnosis itself — which decisions the physician makes, on the basis of which data, and at which stages of the process errors arise that automated analysis can help mitigate.

Classification and diagnosis of scoliosis

Scoliosis is classified according to etiology, anatomical region, and degree of curvature. The Cobb angle remains the universally accepted measure for assessing deformity severity and making treatment decisions. Conservative treatment, including the use of an orthopedic brace, is most commonly used for mild and moderate deformities, while surgical correction is considered for more severe and progressive cases [7].

Early detection of scoliosis is considered important for the timely initiation of conservative treatment and the prevention of deformity progression. The effectiveness of orthopedic brace treatment was confirmed by the randomized controlled BRAIST study, which showed that 72% of adolescents treated with a brace reached skeletal maturity without curve progression to the surgical threshold (≥50°), compared with 48% in the control group [8]. These results highlight the

importance of early diagnosis and regular monitoring, which is an important motivation for developing automated systems that enable more precise and more frequent monitoring with a reduced need for radiographic examinations and therefore lower exposure to ionizing radiation.

In clinical practice, the initial assessment of scoliosis is based on physical examination and clinical evaluation, after which, in suspected or confirmed cases, radiographic analysis is used to precisely determine the degree of deformity. The cumulative radiation exposure from repeated radiographic examinations in children and adolescents is an important clinical aspect that is taken into account during monitoring and treatment decision-making. Epidemiological data show that AIS is significantly more frequent and more severe in girls than in boys — the female-to-male ratio ranges from 1.5:1 to 3:1 and increases with age [9]. The curve magnitude at the time of diagnosis and the degree of skeletal maturity are the most significant predictors of progression, with younger patients who have larger initial curves at a considerably higher risk of deformity progression if left untreated [10].

Limitations of manual diagnosis

Inter-observer variability in Cobb angle measurement is particularly pronounced in borderline cases. A difference of only a few degrees can be clinically significant and influence the treatment decision — for example, crossing the 25-degree threshold, which is often used as an indication for bracing, or the 45–50-degree threshold, which is considered an indication for surgical treatment. Automated systems that enable consistent and rapid angle measurement, free from the influence of observer fatigue and with reduced inter-observer variability, represent a clinically justified approach, provided that they are validated on representative populations and reliably identify borderline cases.

Beyond manual measurements, increasing attention is being paid to the development of AI applications that automate Cobb angle measurement directly from radiographic images. Li et al. [11] developed a deep learning–based mobile application (CobbAngle Pro) and evaluated it on 802 radiographs from 601 patients. The mean absolute error was 2.00° when analyzing photographed images and 2.08° when analyzing uploaded images, with both methods remaining within the clinically acceptable threshold of 5°. Similarly, Russo et al. [12] analyzed 197 radiographic images and compared manual and automated Cobb angle measurements, indicating that the use of image processing techniques, such as contrast correction, can significantly reduce the discrepancy between automated and manual estimates.

OVERVIEW OF DEEP NEURAL NETWORK ARCHITECTURES

The application of deep learning in medical image analysis has grown rapidly over the past decade, and an overview of the methods is given in Litjens et al. [3]. Spinal analysis follows this trend — the number of papers applying deep neural networks to the detection and analysis of scoliosis has grown significantly since 2017, when CNN approaches first achieved performance comparable to manual methods.

Based on the reviewed literature, we identify eight classes of architectures applied in scoliosis analysis. The classification is based on the dominant architectural principle, since the same architectures are used across different tasks — e.g., encoder–decoder models for segmentation and detection, and CNN classifiers for severity assessment and screening.

The boundaries between classes are not strict, particularly in more recent work where hybrid CNN–Transformer approaches blur the clear distinctions between architectures. Models such as ConvNeXt and CvT combine CNN and Transformer principles, while some Transformer models integrate convolutional layers for better local representation, which makes the classification partly arbitrary.

Each class is presented through typical representatives, primary tasks, and indicative performance, with the caveat that direct comparisons should be interpreted with caution due to differences in datasets and evaluation protocols. The reviewed studies suggest a general trend of decreasing Cobb angle estimation error from classical CNN models toward Transformer and

hybrid CNN–Transformer approaches [20], [21], reflecting the general evolution of the field. The values given in Table 1 are approximate, rounded values summarizing the results reported in the reviewed studies; they are not the outcome of a pooled statistical analysis, and the way they were compiled is explained in the section on evaluation metrics (Table 1).

Table 1 presents a classification of the deep learning architectures most commonly used in scoliosis analysis, together with representative models, primary application areas, and typical results reported in the literature, thereby illustrating the dominant development directions and the performance levels achieved in this field.

Table 1. Architecture classes and typical results in scoliosis analysis

Architecture class	Typical examples	Primary task	Typical results
CNN (classical)	ResNet, DenseNet, EfficientNet	Classification, regression	MAE ~3–4° / AUC ~0.92–0.97
Encoder–decoder	U-Net, SegNet, Attention U-Net	Segmentation	Dice 0.88–0.93
Recurrent models	CNN–LSTM hybrid	Sequential vertebra analysis	Limited data availability
Generative models (GAN)	GAN, DCGAN, GAN inversion	Augmentation, data representation	Improved generalization (indirect effect)
Transformer models	ViT-B, Swin, DeiT	Modeling global dependencies	MAE ~2.5–3.5° / AUC ~0.95–0.98
Hybrid CNN–Transformer	ConvNeXt, CvT, ViT–ResNet	Classification, regression	MAE ~2–3° / AUC ~0.96–0.99
3D convolutional models	3D U-Net, nnU-Net, VerSe	CT/MRI segmentation	Dice 0.91–0.94
Graph networks (GNN)	SpineGCN	Inter-vertebral relations	Research stage

Classical CNN architectures

Classical CNN architectures (ResNet, DenseNet, EfficientNet) dominated the literature until 2021 and are still frequently used as backbone or hybrid components in many studies. ResNet-50 and ResNet-101 are the most common choices due to their good balance between accuracy and computational efficiency. The use of transfer learning has significantly improved performance in scoliosis analysis, particularly in automated Cobb angle measurement, where some studies report errors below 2° and a high ICC relative to clinical measurements [13], although the results were mostly obtained on smaller, single-institution datasets. DenseNet [14] has proven particularly useful in tasks where fine vertebral features are key for discrimination, thanks to its dense connections that enable the preservation and reuse of low-level features across all network layers.

Beyond diagnostic tasks, CNN architectures are also applied in specific clinical applications. Vergari et al. [15] proposed a CNN model combined with discriminant analysis for the automatic identification of the type of scoliosis treatment (brace, spinal implant, or no treatment) on posteroanterior radiographs, achieving an accuracy of 98.3% and high brace detection (99.7%). Wang et al. [16] developed a cascaded CNN architecture with a PCA shape constraint for vertebral landmark localization, decomposing the task into a two-stage process (localization of centers and vertebral angles) and achieving a significant reduction in error on the AASCE dataset. Zhang et al. [17] developed ScolioNets, which uses smartphone back photographs to assess AIS without exposure to ionizing radiation, achieving an AUC of 0.902 for surgical treatment recommendation and a sensitivity comparable to that of specialists.

Encoder–decoder architectures

U-Net and its variants dominate vertebra segmentation tasks. Skip connections between the encoder and decoder preserve the key spatial information needed for accurate reconstruction of vertebral anatomy. The most common variants include Attention U-Net, U-Net++, and 3D U-Net for volumetric segmentation of CT scans. VerSeNet [18], a specialized 3D variant, enables simultaneous vertebra localization and labeling and serves as a reference model for 3D spinal segmentation, with a Dice coefficient of 0.932 on an international benchmark. These architectures are also applied in non-invasive approaches. Duan et al. [19] developed a Dual Attention U-Net that segments back photographs instead of radiographs, for scoliosis detection in 350 patients and 108 controls. Channel and spatial attention mechanisms improved segmentation accuracy, indicating the potential of AI screening without ionizing radiation.

Transformer architectures

Transformer architectures, including the Vision Transformer (ViT) and Swin Transformer models, began to be applied in scoliosis analysis around 2021–2022 and represent one of the most significant development directions of modern medical image analysis methods. Unlike classical CNN architectures, Transformer models enable more efficient modeling of long-range spatial dependencies between anatomically distant vertebrae through attention mechanisms, which can be significant in spinal deformity assessment tasks. A review of the current literature indicates that hybrid CNN–Transformer approaches achieve competitive or improved performance compared with classical CNN models, particularly in AIS classification and automated Cobb angle measurement tasks [20], [21].

Zhao et al. [20] developed SpineHRformer, an architecture that combines HRNet for multi-scale feature extraction with a Transformer encoder for modeling global dependencies between vertebrae. The model was evaluated on 1,934 radiographs with prospective validation and achieved a localization error of 2.47 pixels and a correlation with manual measurements of $r = 0.86$. In addition, recent review papers on automated Cobb angle measurement point to the growing use of Transformer and attention-based architectures in spinal analysis, particularly because of their ability to integrate local and global anatomical information [21].

Hybrid CNN–Transformer approaches

Hybrid CNN–Transformer approaches are among the most active research directions in medical image analysis. These models combine local CNN features and global Transformer representations, enabling more efficient modeling of anatomical relationships between vertebrae. Architectures such as the Convolutional Vision Transformer (CvT) integrate convolutional layers into Transformer models for better local representation and more stable learning on smaller datasets [22]. Such approaches show significant potential for improving automated AIS classification and Cobb angle estimation. In addition, the ability to simultaneously model local and global features allows these architectures to generalize better across different datasets and more complex spinal deformities.

Generative networks and augmentation

Generative adversarial networks (GANs) are applied in scoliosis analysis primarily for data augmentation and learning representations in latent space. Lee et al. [23] proposed a GAN inversion that maps radiographs into latent space and enables their reconstruction, thereby yielding a compact representation of the features relevant for scoliosis detection. This approach shows good performance in AIS screening tasks and stability even on external datasets. Nevertheless, the use of GANs in direct augmentation remains limited, although they have the potential to improve model generalization. Their importance is particularly pronounced in medical domains with a limited number of annotated images, where synthetically generated data can help reduce the problems of class imbalance and insufficient training samples.

Three-dimensional and graph-based approaches

3D convolutional models, particularly nnU-Net [24], as well as methods based on the VerSe benchmark [18], are used for volumetric segmentation of CT scans and achieve Dice coefficients in the range of 0.91–0.94.

Graph neural networks (GNNs) model the spinal column as a graph structure, where the vertebrae are nodes and the spatial relationships are edges. This approach is intuitive from an anatomical standpoint, but it is still in an early research stage compared with CNN and Transformer methods.

DATASETS

One of the key findings of the available literature review is the pronounced fragmentation of available datasets. The dominant problem is the fact that most published studies use private datasets that are not available to the wider research community, which in practice makes direct comparison of methods and reproduction of results difficult. As a consequence, research groups often develop their own databases and evaluation protocols, which further contributes to the heterogeneity of results and complicates the verification of the progress achieved.

Table 2 provides an overview of the most significant datasets identified in the reviewed literature, with notes on availability, modality, and annotation type.

Table 2. Overview of key datasets for scoliosis analysis using DNNs

Dataset	N (approx.)	Modality	Availability	Annotations	Origin	Note
AASCE 2019 (MICCAI Challenge)	481	PA X-ray	Public	Vertebral landmarks, Cobb angles	International	Frequently used benchmark for Cobb regression [25]
VerSe Challenge	~374	CT (3D)	Public	Vertebral labels and segmentation	International	Standard benchmark for 3D vertebral analysis [18]
UK Biobank	22,000+	PA X-ray	Restricted	General musculoskeletal and clinical data	UK	Large population-based biomedical dataset
NHANES (X-ray subset)	~7,000+	PA X-ray	Public	Epidemiological and clinical variables	USA	Diverse demographics; epidemiological context
Private clinical datasets	Variable	PA X-ray/CT	Not available	Various (Cobb, segmentation, classification)	Multiple centers	Dominant in the existing literature; limited reproducibility

The AASCE 2019 Challenge dataset, with 481 annotated radiographs from the MICCAI competition — used, among others, by Wu et al. [25] — is one of the few widely used benchmark datasets for Cobb angle regression on posteroanterior radiographs. Although it is often used as a reference dataset for evaluating algorithms, its size is limited for training modern deep models without the use of transfer learning. A similar role in the area of 3D vertebra segmentation and labeling on CT scans is played by the VerSe Challenge dataset [18], which has become one of the standard reference frameworks for evaluating spinal segmentation methods.

Despite the importance of these datasets, the field of spinal analysis still lags behind other domains of medical image analysis in the availability of large public datasets. Unlike the analysis of fundus photographs or mammography, where publicly available datasets with tens of thousands of annotated samples exist, spinal analysis still lacks an equivalent large, multicenter public

resource. The limited availability of standardized datasets represents one of the key obstacles to the development of robust and generalizable deep learning models in this field.

An additional problem is the insufficient standardization of evaluation protocols and annotation procedures across different studies. Annotation protocols are documented in detail in a relatively small number of the reviewed papers, while the inter-observer reliability of annotations is quantified even more rarely. This is particularly significant given that the Cobb angle inherently exhibits inter-observer variability of approximately 3–7 degrees [2]. This variability represents a realistic limit on the expected accuracy of automated systems compared with manual measurements, while the insufficient documentation of annotation procedures further complicates a reliable comparison of performance between different methods and studies.

Key studies by category

Table 3 presents representative studies covering the main classes of deep neural networks and the most common tasks in scoliosis analysis. The selected papers illustrate different approaches based on CNN and hybrid CNN–Transformer architectures, with differences in imaging modalities, dataset sizes, and evaluation metrics.

The studies presented cover typical clinical tasks such as anatomical landmark localization, screening, deformity classification, and automated Cobb angle estimation. The aim of the presentation is not an exhaustive review of the literature, but rather an illustration of representative approaches and typical model performance under different experimental conditions.

Particular emphasis is placed on studies with clearly defined evaluation protocols and quantitative metrics, so that the results are more transparent and more suitable for mutual comparison. Due to differences in datasets, annotation protocols, and clinical criteria, the reported performance should be interpreted as indicative rather than directly comparable across all studies.

Table 3. Overview of key studies on the application of DNNs in scoliosis analysis (2017–2024)

Authors (ref.)	Year	Architecture	Task	Modality	N	Metric	Key result
Wu et al. [25]	2017	BoostNet	Anatomical landmark localization	PA X-ray	481	MSE	0.0046 (test)
Yang et al. [26]	2019	DenseNet-121	AIS screening	Back photographs	1,780	Sensitivity	89.5%
Zhao et al. [20]	2023	SpineHRformer (HRNet + Transformer)	Landmark localization + Cobb regression	PA X-ray	1,934	MED / Pearson	2.47 px / 0.86
Vergari et al. [15]	2020	CNN	Scoliosis treatment detection	X-ray	1,013	Accuracy	98.3%
Wang et al. [16]	2024	Multi-stage cascaded CNN	Vertebral landmark localization	X-ray	—	Localization error	High precision
Zhang et al. [17]	2023	Deep CNN	AIS classification and monitoring	Smartphone photographs	1,726	AUC	0.902
Tao et al.[27]	2021	Spine-Transformers (3D Transformer)	Vertebra segmentation and labeling	CT (3D)	1005 CT volumes	Dice	High segmentation precision

EVALUATION METRICS AND QUANTITATIVE ANALYSIS

The heterogeneity of evaluation metrics in the reviewed literature is a significant obstacle to the direct comparison of studies. Differences in datasets, validation protocols, and annotation criteria make it difficult to draw fully comparable conclusions across different models and architectures.

For Cobb angle regression, most recent studies achieve an error within the clinically acceptable threshold of approximately 5° [2]. Modern hybrid CNN–Transformer approaches report lower MAE values compared with earlier CNN models, while classification tasks often record high AUC-ROC values. For segmentation tasks, Dice coefficients in most papers range approximately between 0.90 and 0.94. Examples of individual values reported in the reviewed studies include a mean absolute error of 2.00–2.08° for a deep learning–based mobile application [11], Cobb angle errors below 2° for a CNN-based measurement algorithm evaluated on a single-institution dataset [13], a Dice coefficient of 0.932 for 3D vertebra segmentation on the VerSe benchmark [18], and an AUC of 0.902 for smartphone-based AIS assessment [17].

Table 4 presents the typical performance ranges by architecture class in the analyzed studies.

Table 4. Typical performance ranges by architecture class in the analyzed studies

Architecture class	Cobb angle MAE	AUC-ROC (classification)	Dice (segmentation)
CNN classical	~3–4°	~0.92–0.97	~0.89
Encoder–decoder	~3–5°	~0.90–0.96	~0.91
Transformer	~2.5–3.5°	~0.95–0.98	~0.92
Hybrid CNN–Transformer	~2–3°	~0.96–0.99	~0.93
3D CNN / volumetric models	N/A	N/A	~0.92–0.94
Recurrent models	~2.7–4°	~0.93–0.97	N/A

It is important to emphasize that the values presented represent indicative performance ranges and cannot be directly compared across all studies. Differences in deformity complexity, radiograph quality, dataset size and heterogeneity, as well as in evaluation protocols, significantly affect the reported results. The ranges in Table 4 were compiled by grouping the reviewed studies according to their dominant architecture class and taking the approximate (rounded) interval of the values reported for each metric; N/A indicates that the metric is not typically reported for that class. Since the number of studies per class is small and the studies differ in datasets, annotation protocols, and validation schemes, no pooled mean or meta-analytic estimate was computed. Individual studies on smaller, single-institution datasets may report values outside these ranges (e.g., errors below 2° in [13]), which further illustrates the limited comparability of the reported results.

Studies on automated Cobb angle measurement, such as the work of Russo et al. [12], have shown that the quality of the input radiographs can significantly affect model accuracy. The authors demonstrated that simple contrast-enhancement techniques can reduce deviations relative to manual measurements, indicating the importance of standardizing input data in the evaluation of automated systems.

For segmentation tasks, specialized architectures such as nnU-Net consistently achieve better results than generic U-Net variants, particularly with 3D volumetric data. Nevertheless, most studies remain limited to relatively small, single-center datasets, often collected under controlled research conditions. The lack of external validation and heterogeneous clinical datasets remains one of the main obstacles to the wider clinical application of these models.

DISCUSSION

The results of the reviewed literature indicate significant development of the field over the past decade — from early attempts at automated Cobb angle measurement based on classical CNN models to modern hybrid architectures that, under controlled research conditions, achieve accuracy comparable to that of experienced radiologists. However, technical progress does not necessarily imply clinical applicability, and it is precisely this discrepancy that is one of the central issues in contemporary research on the application of artificial intelligence in medicine.

It is important to emphasize that most reported results are based on retrospective studies conducted under experimental conditions, often with relatively small, single-center datasets, without adequate external validation on independent clinical populations. Such limitations significantly reduce the generalizability of the reported performance and represent one of the main obstacles to the wider clinical application of these systems [28], [29].

An additional problem is the heterogeneity of imaging protocols, radiograph quality, and annotation criteria across different institutions, which can lead to reduced model performance when deployed under real clinical conditions.

In this context, artificial intelligence–based systems should be viewed primarily as tools for supporting clinical decision-making, rather than as a replacement for the expert judgment of radiologists or orthopedists [30].

Key trends and progress

The literature review points to several clear development trends in the application of deep learning to scoliosis analysis. First, there is a shift from classical CNN classifiers toward more complex hybrid architectures that combine local feature extraction via convolutional networks with global modeling of spatial dependencies using attention mechanisms and Transformer components. This development has been accompanied by improved performance, particularly in Cobb angle regression and anatomical structure localization tasks.

A second important trend is the increasingly wide use of multi-task learning, in which models simultaneously perform several interrelated tasks, such as segmentation, anatomical landmark localization, and Cobb angle estimation. The reviewed studies indicate that joint training often leads to better performance and more stable generalization compared with separately trained models, which is consistent with the regularization effect of this paradigm.

A third trend concerns the increasing transparency and reproducibility of research. Whereas studies published before 2020 rarely provided access to source code and experimental protocols, more recent work increasingly uses platforms such as GitHub to share implementations, models, and datasets. Although reproducibility in deep learning for medical imaging remains a significant challenge, this trend contributes to easier comparison of methods and more reliable evaluation of reported results.

Methodological weaknesses

Despite significant technical progress, the literature review points to several systemic methodological limitations that complicate the interpretation and clinical application of the reported results. The dominance of single-institution and private datasets indicates that the true generalizability of most published models remains insufficiently examined [32], [33]. Models trained and evaluated within the same institution, using identical radiographic systems and standardized imaging protocols, may not achieve the same performance in different clinical environments. Small single-center datasets also increase the risk of model overfitting, while external validation on independent datasets is rarely performed [33].

An additional problem is small test sets, often with fewer than 100 images, which can lead to unstable and highly variable evaluation metrics. Statistical comparisons between models are rarely reported, while effect sizes, confidence intervals, and statistical significance analyses are largely absent from the reviewed literature. In addition, annotation protocols and inter-observer reliability

are often not adequately documented, even though they directly affect the assessment of automated system performance.

The regulatory and organizational aspects of clinical application also remain insufficiently addressed. Algorithms intended for medical use must undergo rigorous validation and regulatory approval processes, including FDA and CE procedures, which require multicenter studies and proof of measurable clinical benefit. In addition, questions of liability for potential diagnostic errors, model interpretability, and integration into existing clinical workflows have not yet been fully resolved.

Because of these limitations, despite impressive research results, the widespread application of artificial intelligence–based systems in routine orthopedic and radiological practice remains limited for now. Future research should therefore focus on multicenter datasets, the standardization of evaluation protocols, and prospective clinical validation of models (see also Table 5).

Open problems and recommendations

Prediction of disease progression is one of the clinically most important, yet at the same time least explored, areas of artificial intelligence application in scoliosis analysis. Reliable progression prediction could enable timely intervention before the surgical indication is reached, but the development of such models is hindered by the lack of longitudinal datasets with long-term patient follow-up.

One of the few papers that directly addresses this problem is the pilot study by Yahara et al. [31], in which a DCNN model was developed to predict the progression of adolescent idiopathic scoliosis in patients with an initial Cobb angle of 10–25 degrees. The model was trained on 58 patients and achieved an accuracy of 69% and an AUC of 0.70, with performance comparable to or slightly better than the average assessment of spine surgeons within the same study. Although the results indicate that automated progression prediction is technically feasible, the small sample size and the retrospective design represent significant limitations.

Table 5 provides an overview of the key open problems identified in the reviewed literature, together with recommendations for future research.

Table 5. Open problems and recommendations for future research

Open problem	Comment and recommendation
Lack of public datasets	Existing public datasets are small for modern deep models. Large multicenter datasets with standardized annotations are needed.
Neglected 3D nature of scoliosis	Most studies use 2D PA X-ray images, while 3D analysis and EOS systems are far less explored.
Rare external validation	Few studies validate models on independent datasets, so generalizability often remains unknown.
Prediction of disease progression	A clinically very important area that is still poorly explored and mostly based on retrospective studies.
Interpretability and uncertainty	Clinical models should enable interpretation of decisions and estimation of prediction uncertainty.
Regulatory framework	FDA, CE, and EU AI Act requirements are rarely considered despite their importance for clinical application.

CONCLUSION

This paper has presented an overview of the key approaches to applying deep neural networks in the detection and analysis of scoliosis, based on the relevant literature published between 2015 and 2024. The literature analysis covered several dominant architecture classes, with hybrid CNN–Transformer models representing the most advanced approaches for Cobb angle regression, with indicative Cobb angle MAE values of approximately 2–3 degrees (Table 4), which approaches the limit of inter-observer variability among experienced clinicians.

Despite significant technical progress, the field is still characterized by pronounced methodological limitations, including the dominance of single-institution private datasets, rare external validation, and a lack of prospective clinical studies. Further development requires the standardization of evaluation protocols, the development of larger publicly available datasets, multicenter validation of models, and closer collaboration between researchers, clinicians, and regulatory bodies.

Although the results to date indicate significant potential for deep learning in the automated analysis of scoliosis, the transition from experimental systems to routine clinical practice requires additional validation of the robustness, generalizability, and clinical reliability of the proposed approaches.

ACKNOWLEDGEMENTS

This study was supported by the Ministry *of Science, Technological Development and Innovation of the Republic of Serbia and these results are parts of the Grant No. 451-03-33/2026-03/ 200132 with University of Kragujevac – Faculty of Technical Sciences Čačak.*

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest regarding the research presented in this paper.

LITERATURE

1. Weinstein, S. L., Dolan, L. A., Cheng, J. C. Y., Danielsson, A., & Morcuende, J. A. (2008). Adolescent idiopathic scoliosis. The Lancet, 371(9623), 1527–1537. https://doi.org/10.1016/S0140-6736(08)60658-3
2. Morrissy, R. T., Goldsmith, G. S., Hall, E. C., Kehl, D., & Cowie, G. H. (1990). Measurement of the Cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error. JBJS, 72(3), 320-327. https://doi.org/10.2106/00004623-199072030-00002
3. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sanchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005
4. Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216
5. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118. https://doi.org/10.1038/nature21056
6. Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., Tse, D., Etemadi, M., Ye, W., Corrado, G., & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature Medicine*, 25(6), 954–961. https://doi.org/10.1038/s41591-019-0447-x
7. Negrini, S., Donzelli, S., Aulisa, A. G., Czaprowski, D., Schreiber, S., de Mauroy, J. C., Diers, H., Grivas, T. B., Knott, P., Kotwicki, T., Lebel, A., Marti, C., Maruyama, T., O'Brien, J., Price, N., Parent, E., Rigo, M., Romano, M., Stikeleather, L., ... Zaina, F. (2018). 2016 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. Scoliosis and Spinal Disorders, 13(1), 3. https://doi.org/10.1186/s13013-017-0145-8
8. Weinstein, S. L., Dolan, L. A., Wright, J. G., & Dobbs, M. B. (2013). Effects of bracing in adolescents with idiopathic scoliosis. New England Journal of Medicine, 369(16), 1512–1521. https://doi.org/10.1056/NEJMoa1307337

9. Konieczny, M. R., Senyurt, H., & Krauspe, R. (2013). Epidemiology of adolescent idiopathic scoliosis. *Journal of Children's Orthopaedics*, 7(1), 3–9. https://doi.org/10.1007/s11832-012-0457-4
10. Lonstein, J. E., & Carlson, J. M. (1984). The prediction of curve progression in untreated idiopathic scoliosis during growth. Journal of Bone and Joint Surgery, 66(7), 1061–1071. https://doi.org/10.2106/00004623-198466070-00013
11. Li, H., Qian, C., Yan, W., Fu, D., Zheng, Y., Zhang, Z., Meng, J., & Wang, D. (2024). Use of artificial intelligence in Cobb angle measurement for scoliosis: Retrospective reliability and accuracy study of a mobile app. Journal of Medical Internet Research, 26, e50631. https://doi.org/10.2196/50631
12. Russo, A. P., Pastorello, Y., Dénes, L., Caian, C., Margarit, A., Cosma, D., Pop, T. S., & Bodog, A. (2024). Automated Cobb angle measurements for scoliosis diagnosis and assessment: AI applications and accuracy enhancement through image processing techniques. Cureus, 16(8), e66736. https://doi.org/10.7759/cureus.66736
13. Wang, M. X., Kim, J. K., Choi, J.-W., Park, D., & Chang, M. C. (2024). Deep learning algorithm for automatically measuring Cobb angle in patients with idiopathic scoliosis. European Spine Journal, 33(11), 4155–4163. https://doi.org/10.1007/s00586-023-08024-5
14. Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4700–4708. https://doi.org/10.1109/CVPR.2017.243
15. Vergari, C., Skalli, W., & Gajny, L. (2020). A convolutional neural network to detect scoliosis treatment in radiographs. International Journal of Computer Assisted Radiology and Surgery, 15(6), 1069–1074. https://doi.org/10.1007/s11548-020-02173-4
16. Wang, Z., Lv, J., Yang, Y., Lin, Y., Li, Q., Li, X., & Yang, X. (2024). Accurate scoliosis vertebral landmark localization on X-ray images via shape-constrained multi-stage cascaded CNNs. Fundamental Research, 4(6), 1657-1665. https://doi.org/10.1016/j.fmre.2022.10.014
17. Zhang, T., Zhu, C., Zhao, Y., Zhao, M., Wang, Z., Song, R., Meng, N., Sial, A., Diwan, A., Liu, J., & Cheung, J. P. Y. (2023). Deep learning model to classify and monitor idiopathic scoliosis in adolescents using a single smartphone photograph. JAMA Network Open, 6(8), e2330617. https://doi.org/10.1001/jamanetworkopen.2023.30617
18. Sekuboyina, A., Husseini, M. E., Bayat, A., Loeffler, M., Liebl, H., Li, H., Tetteh, G., Kukacka, J., Payer, C., Stern, D., Urschler, M., Yang, M., Aubert, B., Bhatt, P., Wang, F., Chen, C., Bhatt, D. L., Armbrecht, G., Dosgani, L., ... Menze, B. H. (2021). VerSe: A Vertebrae labelling and segmentation benchmark for multi-detector CT images. Medical Image Analysis, 73, 102166. https://doi.org/10.1016/j.media.2021.102166
19. Duan, X., Ma, X., Zhu, M., Wang, L., You, D., Deng, L., & Niu, N. (2025). Deep learning-assisted screening and diagnosis of scoliosis: Segmentation of bare-back images via an attention-enhanced convolutional neural network. Journal of Orthopaedic Surgery and Research, 20, 161. https://doi.org/10.1186/s13018-025-05564-y
20. Zhao, M., Meng, N., Cheung, J. P. Y., Yu, C., Lu, P., & Zhang, T. (2023). SpineHRformer: A Transformer-Based Deep Learning Model for Automatic Spine Deformity Assessment with Prospective Validation. Bioengineering, 10(11), 1333. https://doi.org/10.3390/bioengineering10111333
21. Jin, C., Wang, S., Yang, G., Li, E., & Liang, Z. (2022). A review of the methods on cobb angle measurements for spinal curvature. Sensors, 22(9), 3258. https://doi.org/10.3390/s22093258
22. Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L., & Zhang, L. (2021). CvT: Introducing Convolutions to Vision Transformers. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 22–31. https://doi.org/10.1109/ICCV48922.2021.00009
23. Lee, J. S., Shin, K., Ryu, S. M., Jegal, S. G., Lee, W., Yoon, M. A., Hong, G.-S., Paik, S., & Kim, N. (2023). Screening of adolescent idiopathic scoliosis using generative adversarial

network (GAN) inversion method in chest radiographs. PLOS ONE, 18(5), e0285489. https://doi.org/10.1371/journal.pone.0285489
24. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z
25. Wu, H., Bailey, C., Rasoulinejad, P., & Li, S. (2017). Automatic landmark estimation for adolescent idiopathic scoliosis assessment using BoostNet. In: MICCAI 2017, Lecture Notes in Computer Science (Vol. 10433, pp. 127–135). Springer. https://doi.org/10.1007/978-3-319-66182-7_15
26. Yang, J., Zhang, K., Fan, H., Huang, Z., Xiang, Y., Yang, J., ... & Lin, H. (2019). Development and validation of deep learning algorithms for scoliosis screening using back images. Communications biology, 2(1), 390. https://doi.org/10.1038/s42003-019-0635-8
27. Tao, R., Liu, W., & Zheng, G. (2021). Spine-transformers: Vertebra labeling and segmentation in arbitrary field-of-view spine CTs via 3D transformers. Medical Image Analysis, 75, 102258. https://doi.org/10.1016/j.media.2021.10225828. Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine, 17*(1), 195. https://doi.org/10.1186/s12916-019-1426-2
29. Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Aviles-Rivero, A. I., Etmann, C., McCague, C., Beer, L., Weir-McCall, J. R., Teng, Z., Gkrania-Klotsas, E., Rudd, J. H. F., Sala, E., Schönlieb, C.-B., & Rueckert, D. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nature Machine Intelligence, 3(3), 199–217. https://doi.org/10.1038/s42256-021-00307-0
30. Topol, E. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7
31. Yahara, Y., Tamura, M., Seki, S., Kondo, Y., Makino, H., Watanabe, K., Kamei, K., Futakawa, H., & Kawaguchi, Y. (2022). A deep convolutional neural network to predict the curve progression of adolescent idiopathic scoliosis: A pilot study. BMC Musculoskeletal Disorders, 23, 610. https://doi.org/10.1186/s12891-022-05565-6
32. Aggarwal, R., Sounderajah, V., Martin, G., Ting, D. S. W., Karthikesalingam, A., King, D., Ashrafian, H., & Darzi, A. (2021). Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis. *npj Digital Medicine, 4*, 65. https://doi.org/10.1038/s41746-021-00438-z
33. Yu, A. C., Mohajer, B., & Eng, J. (2022). External validation of deep learning algorithms for radiologic diagnosis: a systematic review. Radiology: Artificial Intelligence, 4(3), e210064. https://doi.org/10.1148/ryai.210064