Scoliosis Detection System Using Zero-Shot Segmentation

[1]Dr. Yogesh Golhar, [2]Dr. Sushil Kumar Singh

[1] LGPR, Lincoln University College, Malaysia,
[2]Associate Professor at Marwadi University, Rajkot, Gujarat, India
Email ID: [1] yogesh.golhar28@gmail.com, [2]sushil.sngh1@gmail.com,
sushil.sngh001007@seoultech.ac.kr

Abstract: Scoliosis is a progressive spinal disorder involving abnormal lateral deviation and vertebral rotation, which may lead to functional and structural complications if not detected early. The Cobb angle is widely accepted as the primary clinical indicator for quantifying spinal curvature and guiding treatment decisions. However, conventional Cobb angle measurement from radiographic images requires manual identification of vertebral landmarks, making the process time-intensive and susceptible to observer variability. While recent artificial intelligence-based approaches have improved automation in spinal analysis, most existing methods depend on extensive annotated datasets and supervised training, which restricts generalizability across institutions and imaging protocols.

This study introduces a scoliosis detection framework built upon zero-shot segmentation using a foundation vision model. The proposed system employs the Segment Anything Model (SAM), a large-scale pre-trained Vision Transformer, to generate segmentation masks directly from spinal radiographs without additional task-specific training. The most anatomically relevant mask is selected and refined using morphological operations to enhance structural continuity. Subsequently, skeletonization and regression-based geometric fitting are applied to estimate spinal orientation and compute the Cobb angle automatically. Severity categorization is performed according to established clinical thresholds. By removing the dependency on labelled datasets and retraining, the proposed method provides a scalable, adaptable, and interpretable solution for automated scoliosis screening and curvature estimation in real clinical environments.

Keywords: Scoliosis, Zero-Shot Segmentation, Segment Anything Model, Foundation Vision Models, Cobb Angle Estimation, Spinal Curvature Analysis, Medical Image Processing.

Introduction

Scoliosis is a structural deformity of the spine characterized by abnormal lateral curvature and vertebral rotation. The condition may present in different anatomical regions of the spine, including thoracic, lumbar, thoraco-lumbar, or combined curvature patterns, as illustrated in Fig. 1. The severity of scoliosis is clinically quantified using the Cobb angle, first introduced in 1948 [20], which remains the gold standard for diagnosis and treatment planning. Accurate measurement of the Cobb angle is essential for determining disease progression, selecting appropriate treatment strategies, and monitoring therapeutic outcomes. However, manual Cobb angle estimation from radiographic images is inherently subjective and prone to inter- and intra-observer variability, with reported measurement errors ranging from 4° to 10° [19], [18]. Such inconsistencies may lead to misclassification of severity and delay in clinical intervention.

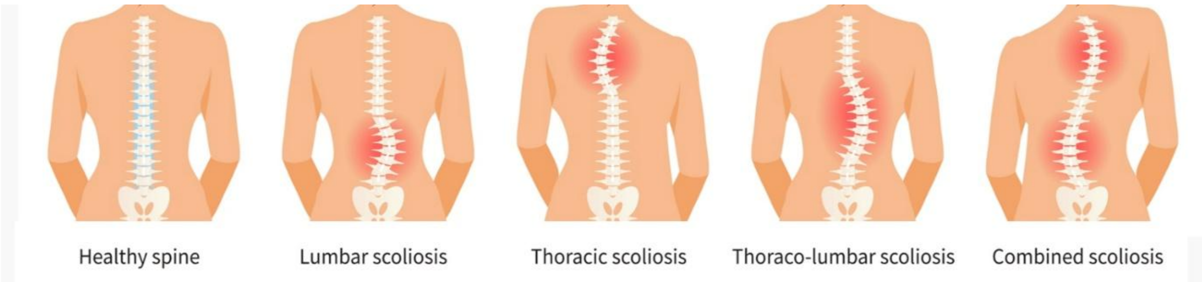

Fig. 1. Illustration of spinal curvature patterns including healthy spine, lumbar scoliosis, thoracic scoliosis, thoraco-lumbar scoliosis, and combined scoliosis.

With the rapid advancement of artificial intelligence (AI) in medical imaging, automated approaches for spinal curvature analysis have gained significant attention [12], [16]. Deep learning-based techniques have demonstrated promising results in scoliosis classification and curvature estimation [15], [9], [7], [5]. Several studies have proposed weakly supervised and fully supervised learning frameworks for automatic Cobb angle measurement using convolutional neural networks and geometric modelling [9], [10], [7]. Transfer learning and hybrid classification strategies have also been employed to improve detection performance in limited datasets [3], [8], [1]. While these methods achieve competitive accuracy, they rely heavily on large annotated datasets and require task-specific retraining, which restricts their scalability and generalization across diverse clinical settings.

Traditional medical image segmentation approaches, such as U-Net [17], require pixel-level annotations that are costly and time-consuming to obtain. In addition, alternative non-radiographic screening techniques, including optical scanning and surface-based methods, have been explored to reduce radiation exposure [11], [14]. Although these approaches contribute to early screening, their integration with automated and robust curvature estimation remains limited.

Recent developments in foundation models have introduced a new paradigm in computer vision. The Vision Transformer architecture [13] demonstrated the effectiveness of attention-based models for large-scale image analysis, leading to the development of general-purpose segmentation frameworks such as the Segment Anything Model (SAM) [6]. SAM is trained on large-scale datasets and exhibits strong zero-shot generalization capabilities, enabling segmentation of previously unseen structures without task-specific training. Emerging research has further explored zero-shot medical image segmentation using foundation vision models [4], [2], highlighting their potential in clinical applications.

Motivated by these advancements, this work proposes a scoliosis detection system using zero-shot segmentation. Unlike conventional supervised learning approaches, the proposed framework leverages a foundation segmentation model to extract the spinal region directly from radiographs without additional training or manual annotation. A geometric regression-based fitting method is subsequently applied to estimate spinal curvature and compute the Cobb angle automatically. By eliminating the dependency on labelled datasets while preserving interpretability and computational simplicity, the proposed system offers a scalable and adaptable solution for automated scoliosis screening and curvature analysis.

Related work

Automated analysis of scoliosis has been widely explored using deep learning and image processing techniques. Early approaches focused on supervised convolutional neural networks (CNNs) for classification and curvature estimation. Several studies have proposed weakly supervised and fully supervised learning frameworks for automatic Cobb angle measurement from spinal radiographs [9], [10]. These methods typically employ vertebral detection or spine centerline extraction to compute curvature using geometric modelling [7], [5]. Although they demonstrate promising accuracy, they rely heavily on annotated datasets and task-specific training. Transfer learning strategies have also been applied to improve scoliosis detection performance in limited datasets [3]. Hybrid approaches combining CNNs with classical machine learning classifiers, such as support vector machines, have further enhanced classification robustness [8]. Ensemble-based deep learning systems have been introduced for school-based screening programs [1], while surface-based and back-profile analysis methods have also been investigated [15]. Despite these advancements, most of these techniques require retraining when applied to new imaging protocols or clinical environments.

Traditional medical image segmentation models, such as U-Net [17], have been widely adopted for biomedical image segmentation tasks. However, these architectures require pixel-level annotations, which are costly and time-consuming to obtain in medical settings. Recent developments in Vision Transformer architectures [13] have enabled scalable attention-based models, leading to the introduction of foundation segmentation frameworks such as the Segment Anything Model (SAM) [6]. SAM demonstrates strong zero-shot generalization capabilities and can segment previously unseen structures without task-specific training. Emerging studies have further explored zero-shot medical image segmentation using foundation vision models [4], [2].

Despite these advances, the application of zero-shot foundation models for automated scoliosis curvature estimation remains limited. Most existing approaches still depend on supervised learning and annotated datasets. This gap motivates the development of a zero-shot segmentation-based framework for automated spinal curvature analysis.

Method, Experiments and Results

The proposed scoliosis detection system integrates zero-shot segmentation with geometric curvature analysis to enable automated Cobb angle estimation from spinal radiographs. The framework is designed as a sequential multi-stage pipeline consisting of data acquisition, image preprocessing, segmentation, mask refinement, centerline extraction, and curvature estimation.

Initially, spinal X-ray images are acquired and subjected to basic preprocessing operations, including resizing and intensity normalization, to improve structural clarity. Zero-shot segmentation is then performed using the Segment Anything Model (SAM) [6], a foundation vision model built upon the Vision Transformer architecture [13]. SAM generates candidate masks without requiring task-specific retraining, enabling segmentation of the spinal column directly from the input image.

Following segmentation, the most anatomically
consistent mask is selected and refined using morphological operations to enhance structural continuity. A skeletonization-based centerline extraction step is subsequently applied to approximate the medial axis of the spine. The extracted centerline is used for regression-based curvature estimation, from which the Cobb angle is computed in accordance with established clinical definitions [20]. Finally, severity classification is determined using standard orthopedic thresholds.
The complete workflow of the proposed system is illustrated in Fig. 1.

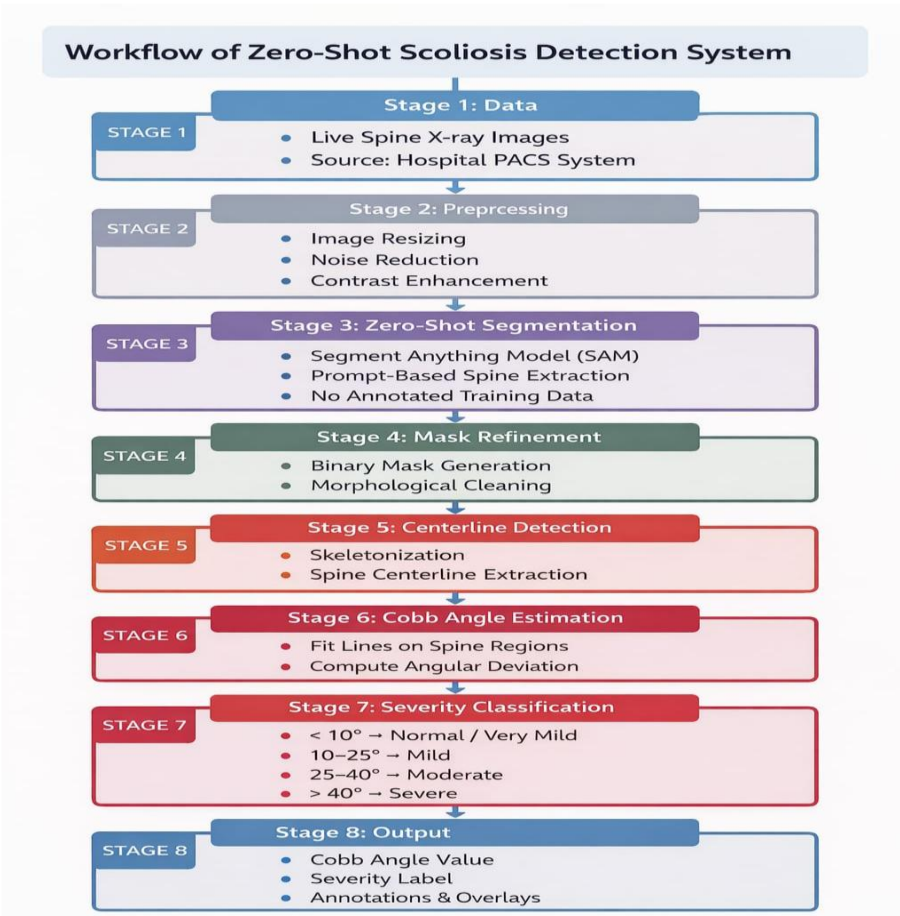

Fig. 1. Workflow of the proposed zero-shot scoliosis detection system.

Zero-Shot Segmentation Using SAM

To perform spinal region extraction without supervised retraining, the proposed framework employs the Segment Anything Model (SAM) [6], a large-scale foundation segmentation model built upon the Vision Transformer (ViT) architecture [13]. Unlike conventional supervised segmentation networks such as U-Net [17], SAM enables mask generation for previously unseen anatomical structures through a prompt-guided and zero-shot learning mechanism.
The segmentation process consists of three principal components: image encoding, prompt encoding, and mask decoding.

Image Encoding

Given an input spinal radiograph ^{T}, the image is first resized to a fixed resolution to match the transformer backbone requirements. The resized image is divided into non-overlapping patches, each of which is linearly projected into a high-dimensional embedding space. Positional encodings are added to preserve spatial relationships between patches. These embeddings are processed by a transformer encoder to generate a global image representation:

$$F = EImg\ (I)$$

Where $E_{Img}(.)$ denotes the image encoder and F represents the extracted feature embedding of the input radiograph.

This encoding mechanism enables the model to capture long-range spatial dependencies, which are critical for accurately segmenting elongated anatomical structures such as the spine.

Prompt Encoding

SAM supports prompt-based segmentation using spatial cues such as points, bounding boxes, or masks. Let p denote a prompt input. The prompt encoder transforms the input prompt into an embedding representation:

$$\boldsymbol{P = EPrompt\ (p)}$$

Where $EPrompt\ (.)$ generates the prompt embedding P.

Although SAM allows manual prompting, the automatic mask generation strategy used in this work enables internal region proposals without explicit user-defined prompts, facilitating fully automated processing.

Mask Decoding and Selection

The final segmentation masks are generated through a transformer-based mask decoder that integrates the image embedding F and prompt embedding P via cross-attention mechanisms:

$$\boldsymbol{M = D(F, P)}$$

Where $D(.)$ denotes the mask decoder and $M = m1, m2, \ldots, mn$ represents the set of candidate masks.

Each candidate mask is associated with a predicted quality score. The optimal segmentation result is selected by:

$$\boldsymbol{mbest = arg\ maxmi€M\ \ Score(mi)}$$

This selected mask serves as the initial representation of the spinal column for subsequent refinement and curvature estimation.

By leveraging SAM's large-scale pre-training and zero-shot capability [6], the proposed system eliminates the need for annotated training data while maintaining robust structural segmentation performance.

Pre-processing and Mask Refinement

Prior to segmentation, basic preprocessing operations are applied to improve image consistency and structural visibility. Spinal radiographs may exhibit variations in illumination, contrast, and noise levels depending on acquisition conditions. To reduce these inconsistencies, the input image is resized to a standardized resolution and normalized to enhance intensity distribution. This step ensures stable feature extraction during the transformer-based encoding stage.

Following zero-shot segmentation using SAM [6], multiple candidate masks are generated. Since these masks may correspond to different anatomical structures, the spinal column must be isolated from the candidate set. The mask with the largest anatomically consistent connected area is selected:

$$\boldsymbol{Mspine\ =\ arg\,maxmi€M\ \ Area\,(mi)}$$

Where *Area* (m_i) represents the number of foreground pixels in mask m_i.

To improve structural continuity and eliminate minor discontinuities, morphological closing operations are applied:

$$\boldsymbol{Mrefined\ =\ Close(mspine, K)}$$

where K denotes the structuring kernel used for refinement. This operation smooths boundaries and fills small gaps, ensuring that the spinal region is represented as a continuous structure.

The refined mask serves as the basis for centerline extraction and subsequent curvature estimation.

Centerline Extraction

After refinement, the binary spine mask represents a continuous foreground region corresponding to the spinal column. To estimate spinal curvature more robustly, the medial axis (centerline) of the segmented region is extracted using skeletonization.

Skeletonization reduces the binary mask to a one-pixel-wide representation while preserving the topological structure of the spine. Let $Mrefined$ denote the refined binary mask. The skeletonized representation is obtained as:

$$C\ =\ \boldsymbol{Skeleton\ (Mrefined)}$$

where C represents the extracted centerline of the spinal column.

The centerline provides a simplified geometric representation of spinal orientation, eliminating boundary thickness variations and improving numerical stability during curvature estimation. By operating on the medial axis rather than the full mask, the curvature computation becomes less sensitive to segmentation boundary irregularities.

The extracted centerline coordinates (xi, yi) are subsequently used for regression-based curvature approximation.

Cobb Angle Estimation

Following centerline extraction, the coordinates of the skeletal representation C are obtained as a set of ordered points (xi, yi) describing the spinal medial axis. These points are used to approximate the global spinal orientation.

To estimate curvature, a linear regression model is fitted to the extracted centerline coordinates:

$$\boldsymbol{y\ =\ ax\ +\ b}$$

where *a* denotes the slope of the fitted line and *b* represents the intercept.

The inclination angle Q of the spinal axis is computed as:

$$\boldsymbol{Q\ =\ tan-1\ (a)}$$

The Cobb angle is then estimated as the absolute value of this inclination:

$$Cobb\ Angle = |Q|$$

Although classical Cobb measurement involves identification of the upper and lower end vertebrae [20], the regression-based approximation provides a consistent and automated estimation of spinal curvature suitable for screening and preliminary evaluation.

This geometric approach ensures interpretability while maintaining computational efficiency.

Severity Classification

The computed Cobb angle is used to determine scoliosis severity according to clinically established orthopedic thresholds [20]. These thresholds are widely adopted in clinical practice for treatment planning and monitoring disease progression.

Severity is categorized as follows:

- **Normal:** Cobb angle < 10°
- **Mild Scoliosis:** 10° ≤ Cobb angle < 25°
- **Moderate Scoliosis:** 25° ≤ Cobb angle < 40°
- **Severe Scoliosis:** Cobb angle ≥ 40°

This rule-based classification provides a direct and interpretable mapping between computed curvature values and clinical severity levels. By aligning the automated estimation with established medical standards, the proposed framework ensures practical applicability in screening and preliminary assessment scenarios.

Experimental Setup

The proposed scoliosis detection framework was evaluated using spinal radiographs obtained from the NeoRad PACS system at DY Patil Hospital. The dataset consisted of 5,120 anonymized spine X-ray images acquired under routine clinical protocols. The images include varying degrees of spinal curvature and represent diverse patient demographics and imaging conditions.

All radiographs were processed using the Segment Anything Model (SAM) with the ViT-H backbone configuration [6], [13]. The model was executed in a CPU-based environment without additional task-specific retraining. Each image was resized to match the model input requirements prior to segmentation.

For each radiograph, the zero-shot segmentation framework generated candidate masks, from which the spinal region was extracted and refined. Centerline extraction was performed using skeletonization, followed by regression-based curvature estimation to compute the Cobb angle automatically.

The evaluation focused on functional validation of automated Cobb angle computation across the dataset. The system successfully generated curvature estimates for the majority of input images without requiring manual annotation or retraining. The results demonstrate the feasibility of foundation-model-based zero-shot segmentation for large-scale scoliosis screening.

Results And Discussion

The proposed zero-shot scoliosis detection framework was evaluated on spinal radiographs obtained from the NeoRad PACS system at DY Patil Hospital. The system successfully performed segmentation, mask refinement, centerline extraction, and automated Cobb angle estimation without task-specific retraining.

Stage 1: Zero-Shot Spine Segmentation

In the first stage, zero-shot segmentation using the Segment Anything Model (SAM) [6] was applied to the original spinal radiograph. The model generated candidate masks, from which the anatomically consistent spinal region was selected. The segmentation result demonstrates that SAM effectively isolates the spinal column despite variations in surrounding anatomical structures.

As shown in Fig. 2, the spinal region is successfully highlighted from the original X-ray image using zero-shot segmentation.

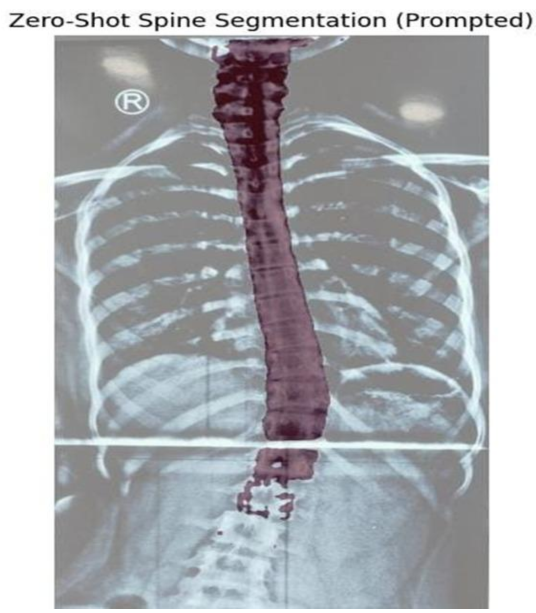

Fig. 2. Zero-shot spine segmentation using SAM highlighting the extracted spinal region.

Stage 2: Clean Binary Spine Mask

Following segmentation, morphological refinement was performed to eliminate noise and small disconnected components. The refined output is converted into a clean binary mask, where only spine pixels are retained. This step ensures structural continuity and improves stability for subsequent geometric analysis.

Fig. 3 presents the clean binary spine mask obtained after noise removal and morphological processing.

Fig. 3. Refined binary spine mask after morphological cleaning and noise removal

Stage 3: Spine Centerline Extraction and Cobb Angle Estimation

After refinement, skeletonization was applied to extract the medial axis (centerline) of the spinal column. The centerline provides a simplified geometric representation of spinal orientation and serves as the basis for curvature estimation. A regression-based fitting method was applied to compute the inclination angle of the spinal axis.

The automated system estimated a Cobb angle of **5.62°**, corresponding to very mild curvature according to established clinical thresholds [20].

The extracted centerline and curvature representation are illustrated in Fig. 4.

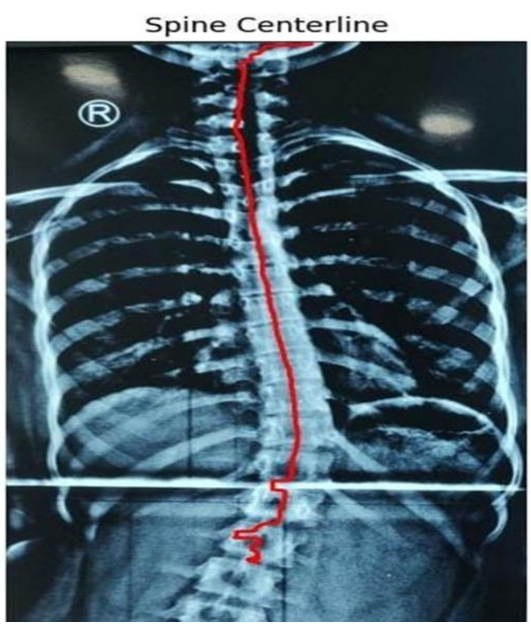

Fig. 4. Extracted spine centerline used for regression-based Cobb angle estimation (5.62°).

Overall, the results demonstrate that the proposed zero-shot framework is capable of generating consistent spinal segmentation and curvature estimation without requiring annotated training data.

The integration of foundation-model-based segmentation with geometric analysis enables automated and interpretable scoliosis screening.

Conclusion

This study presented a scoliosis detection system based on zero-shot segmentation and geometric curvature analysis. The proposed framework leverages the Segment Anything Model (SAM) [6], built upon the Vision Transformer architecture [13], to localize the spinal column directly from radiographic images without requiring annotated training data. By utilizing foundation-model-based segmentation, the system enables reliable spine localization across diverse imaging conditions in a zero-shot manner.

Following segmentation, morphological refinement and centerline extraction were applied to obtain a stable geometric representation of the spinal column. A regression-based approach was then used to compute the Cobb angle in alignment with established clinical definitions [20]. The automated pipeline supports objective and consistent curvature measurement while maintaining computational simplicity. Unlike purely black-box deep learning approaches, the geometry-based estimation improves clinical interpretability by providing a transparent mathematical formulation of curvature analysis.

The framework was validated on 5,120 spinal X-ray images obtained from a real hospital PACS system (NeoRad, DY Patil Hospital), demonstrating the feasibility of large-scale automated scoliosis screening without task-specific retraining. The results highlight the potential of foundation vision models for medical image analysis and curvature estimation.

In future work, the proposed system will be extended to spatial (direct) imaging modalities beyond conventional X-ray radiographs, enabling spine localization and curvature analysis from surface or direct anatomical images. Further clinical validation with expert-annotated measurements and multi-center datasets will also be conducted to strengthen quantitative performance assessment.

References

[1] Y. Chen, et al., “Deep Learning Ensemble for School-Based Scoliosis Screening,” Computers in Biology and Medicine, 2025.

[2] R. Patel, et al., “Foundation Model-Based Automated Spine Segmentation and Curvature Analysis in Clinical Radiographs,” Medical Image Analysis, 2025.

[3] T. Ahmed, et al., “Transfer Learning for Scoliosis Detection,” Pattern Recognition Letters, vol. 169, pp. 60–67, 2024.

[4] K. He, et al., “Zero-Shot Medical Image Segmentation Using Foundation Vision Models,” IEEE Journal of Biomedical and Health Informatics, 2024.

[5] J. Liu, et al., “Deep Learning-Based Automatic Vertebra Detection and Cobb Angle Measurement,” Medical Physics, 2024.

[6] A. Kirillov, et al., “Segment Anything,” in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4015–4026, 2023.

[7] Y. Zhang, et al., “Automatic Scoliosis Assessment from Spine X-ray Images Using Deep Learning and Geometric Analysis,” IEEE Transactions on Medical Imaging, vol. 42, no. 2, pp. 452–463, 2023.

[8] H. Zhao, et al., “Hybrid CNN–SVM Approach for Scoliosis Classification,” Applied Soft Computing, vol. 134, 109960, 2023.
[9] X. Ma, et al., “Weakly Supervised Learning for Automatic Cobb Angle Measurement in Scoliosis,” Medical Image Analysis, vol. 73, 102146, 2021.
[10] L. Chen, et al., “Automatic Cobb Angle Measurement of Scoliosis Based on Spine Centerline Extraction,” Computers in Biology and Medicine, vol. 142, 105198, 2022.
[11] M. Pasha, et al., “A Review of Non-Radiographic Approaches for Spinal Deformity Assessment,” Diagnostics, vol. 12, no. 3, 2022.
[12] D. Xu and T. Y. Wong, “Artificial Intelligence in Medical Imaging: A Review,” Computers in Biology and Medicine, vol. 139, 104987, 2021.
[13] A. Dosovitskiy, et al., “An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale,” in International Conference on Learning Representations (ICLR), 2021.
[14] M. Xu, et al., “3D Structured Light for Non-Radiographic Scoliosis Screening,” Sensors, vol. 21, no. 4, 2021.
[15] J. Li, et al., “Deep Learning for Back Surface Analysis in Scoliosis,” IEEE Access, vol. 8, 2020.
[16] G. Litjens, et al., “A Survey on Deep Learning in Medical Image Analysis,” Medical Image Analysis, vol. 42, pp. 60–88, 2017.
[17] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241, 2015.
[18] I. H. Lieberman, et al., “Measurement of Scoliosis,” The Spine Journal, 1997.
[19] D. L. Carman, R. H. Browne, and J. G. Birch, “Measurement of Scoliosis and Kyphosis Radiographs: Intraobserver and Interobserver Variation,” Journal of Bone and Joint Surgery, 1990.
[20] J. R. Cobb, “Outline for the Study of Scoliosis,” Instructional Course Lectures, vol. 5, pp. 261–275, 1948.