

RESEARCH ARTICLE

Automatic Lenke classification of adolescent idiopathic scoliosis with deep learning

Baolin Zhang^{1,2} | Kanghao Chen³ | Haodong Yuan^{1,2,4} | Zhiheng Liao^{1,2} |
Taifeng Zhou^{1,2} | Weiming Guo^{1,2} | Shen Zhao⁵ | Ruixuan Wang³ |
Peiqiang Su^{1,2} 

¹Department of Orthopaedic Surgery, First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China

²Guangdong Provincial Key Laboratory of Orthopedics and Traumatology, First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China

³School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, Guangdong, China

⁴Graduate School of Biomedical Engineering, UNSW Sydney, Sydney, New South Wales, Australia

⁵School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, Guangdong, China

Correspondence

Ruixuan Wang, School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, Guangdong, China.
Email: wangruix5@mail.sysu.edu.cn

Peiqiang Su, Department of Orthopaedic Surgery, First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong, China.
Email: supq@mail.sysu.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 81772293; “Three Major” Construction Scientific Research and Cultivation Special Projects of Sun Yat-sen University, Grant/Award Number: 80000-18843410

Abstract

Purpose: The Lenke classification system is widely utilized as the preoperative evaluation protocol for adolescent idiopathic scoliosis (AIS). However, manual measurement is susceptible to observer-induced variability, which consequently impacts the evaluation of progression. The goal of this investigation was to develop an automated Lenke classification system utilizing innovative deep learning algorithms.

Methods: Using the database from the First Affiliated Hospital of Sun Yat-sen University, the whole spinal x-rays images were retrospectively collected. Specifically, images collection was divided into AIS and control group. The control group consisted of individuals who underwent routine health checks and did not have scoliosis. Afterwards, relative features of all images were annotated. Deep learning was implemented through the utilization of the key-point based detection method to realize the vertebral detection, and Cobb angle measurement and scoliosis classification were performed based on relevant standards. Besides, the segmentation method was employed to achieve the recognition of lumbar vertebral pedicle to determine the type of lumbar spine modifier. Finally, the model performance was further quantitatively analyzed.

Results: In the study, a total of 2082 spinal x-ray images were collected from 407 AIS patients and 227 individuals in the control group. The model for vertebral detection achieved an F1-score of 0.809 for curve type evaluation and an F1-score of 0.901 for thoracic sagittal profile. The intraclass correlation efficient (ICC) of the Cobb angle measurement was 0.925. In the analysis of performance for vertebra pedicle

Abbreviations: AI, artificial intelligence; AIS, adolescent idiopathic scoliosis; ASPP, atrous spatial pyramid pooling; CSVL, central sacral vertical line; DL, deep learning; DNN, deep neural network; IOU, intersection over union; SMAPE, symmetric mean absolute percentage error.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Authors. *JOR Spine* published by Wiley Periodicals LLC on behalf of Orthopaedic Research Society.

segmentation model, the F1-score of lumbar modification profile was 0.942, the intersection over union (IOU) of the target pixels was 0.827, and the Hausdorff distance (HD) was 6.565 ± 2.583 mm. Specifically, the F1-score for ultimate Lenke type classifier was 0.885.

Conclusions: This study has constructed an automated Lenke classification system by employing the deep learning networks to achieve the recognition pattern and feature extraction. Our models require further validation in additional cases in the future.

KEYWORDS

AIS, deep learning, Lenke classification

1 | INTRODUCTION

Adolescent idiopathic scoliosis (AIS) is a complex three-dimensional deformity of the spine, including vertebral misalignment and axial rotation that typically arise in children at or around puberty.¹ The associated epidemiological researches indicate that the 1%–3% of children aged 10–16 years will have some degrees of spinal curvature, although most curves will not require surgical intervention.² However, for serious spinal misalignment in AIS with Cobb's angle $>45^\circ$ or rapid aggravation of scoliosis during the observation stage, corrective surgery becomes an indispensable strategy to alleviate disease progression. Previous literature has reported distinct clinical evaluative approaches to aid in the implementation of surgery, such as King classification standard.³ Among these multiple methodologies, the Lenke classification system has been accepted as a putative diagnostic method that can instruct the schematization of surgical plans, including determining the appropriate vertebral levels to be considered in an arthrodesis.^{3,4}

The Lenke classification is a triad classification system comprising: curve typology (1–6), lumbar spine modifier (a, b, c), and sagittal thoracic modifier (–, N, +).³ As shown in Figure 1, three Cobb angles typically are determined, with the largest angle being considered as the main Cobb angle. Based on the bending x-rays images, three curves can be classified as structural or non-structural. A curve is typically identified as structural if it maintains a measurement of over 25° on the bending x-ray. Conversely, if the curve measurement is less than 25° , it is considered a non-structural curve. Particularly those distinct subtypes are based on the different characteristics observed in spinal x-ray images, and the manual measurement method necessitates considerable technical expertise and is susceptible to both intra- and inter-observer errors of 2° – 8° .^{3,4} Moreover, due to complexity of classification, manually evaluating scoliosis in a large number of cases can be time-consuming and laborious.^{3,4} Thus, automating the process of Cobb angle measurement or classification is crucial to eliminate the subjectivity inherent in manual assessment and significantly enhance the efficiency.

Deep learning (DL), an essential branch of artificial intelligence (AI), had recently been rapidly developed and widely utilized in the

medical field, tremendously influencing the diagnosis and treatment of diseases.^{5,6} For instance, the installation of a DL radiologic-detection-system had assisted clinicians in detecting suspicious patients in the COVID-19 pandemic,⁷ and had helped discover signals of anemia through retinal fundus images in ophthalmology. Recently, several studies have investigated the application of deep learning in measuring Cobb angle. Totally, it can be divided into two categories: segmentation-based estimation methods^{8–10} and landmark-based estimation methods.^{11–16} The segmentation-based estimation methods primarily involve the initial segmentation of either the entire spinal region or each individual vertebra.⁹ Specifically, the procedure involved extracting the region of interest (ROI) of the spine, outlining the spine's edges, and subsequently applying U-Net for spine segmentation to calculate its curvature. A cascaded convolutional neural network is employed to segment both the spine and its centerline.⁸ Although segmentation methods may achieve higher accuracy, they often necessitate high-quality images and exhibit poor generalizability. Additionally, creating precise segmentation labels can be a time-consuming process.¹² For the landmark-based methods, the four corners of each vertebra are first identified to determine landmarks and the Cobb angles are subsequently calculated based on these landmarks.¹⁶ Therefore, utilizing landmark-based techniques provides better simulation of real-world clinical applications, while also reducing computational time.^{12,15} However, previous studies have mainly focused on the measurement of Cobb angle and there is a lack of intelligent systems for the classification of spinal curvature.

In this study, we propose to exploit the image data of AIS patients to construct an automated Lenke classification system. Upon uploading the relevant primitive image data, the system automatically establishes the vertebral detection and relevant Cobb angle measurement for each image, as well as generating a classification report.

2 | MATERIALS AND METHODS

Design pipeline: The procedures in systematic construction were illustrated in Figure 2, comprising the following major steps: (1) Establishment of image datasets with pre-processing and feature annotation.

(A)

Type of curve				
Type	Proximal thoracic	Main thoracic	Thoracolumbar/lumbar	Type of curve
1	Non-structural	Structural (main)	Non-structural	Main thoracic (MT)
2	Structural	Structural (main)	Non-structural	Double thoracic (DT)
3	Non-structural	Structural (main)	Structural	Double main (DM)
4	Structural	Structural (main)	Structural	Triple main (TM)
5	Non-structural	Non-structural	Structural (main)	Thoracolumbar/Lumbar (TL/L)
6	Non-structural	Structural	Structural (main)	Thoracolumbar/Lumbar-MT(TL/L-MT) (Lumbar curve>thoracic in $\geq 10^\circ$)

Structural criteria

Proximal Thoracic - Side Bending Cobb $\geq 25^\circ$
 - Kyphosis T2-T5 $\geq +20^\circ$

Main Thoracic - Side Bending Cobb $\geq 25^\circ$
 - Kyphosis T10-L2 $\geq +20^\circ$

Thoracolumbar / Lumbar - Side Bending Cobb $\geq 25^\circ$
 - Kyphosis T10-L2 $\geq +20^\circ$

Main : Largest Cobb measurement always structural

LOCATION OF APEX

CURVE	APEX
Thoracic	T2-T11/12 Disc
Thoracolumbar	T12-L1
Thoracolumbar/Lumbar	L1/2 Disc-L4

(B)

Lumbar Spine Modifier	CSVL up to Lumbar Apex
a	CSVL between pedicles
b	CSVL touches the apical body (or bodies)
c	CSVL completely medial

Sagittal Thoracic Profile T5 –T12	
- (Hypo)	$< 10^\circ$
N (Normal)	$10^\circ - 40^\circ$
+ (Hyper)	$> 40^\circ$

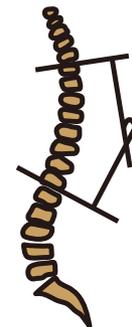
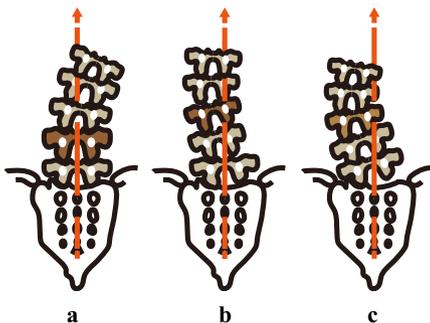


FIGURE 1 Lenke classification criteria. (A) The section displays the major criteria for determining the type of scoliosis. There are a total of six types that are identified based on the distribution of structural curves and the location of the main curve. (B) It shows the criteria for determining the modifier of lumbar spine and sagittal thoracic profile. The lumbar spine modifier can be divided into three categories (a-c) based on the relationship between CSVL and the lumbar pedicle. The sagittal thoracic profile is classified into three categories (hypo, normal, and hyper) based on the size of the Cobb angle at T5–T12 on lateral x-rays films.

This section is to provide elemental resources for the models; (2) Construction of deep learning frameworks: Based on the *Transformer* network module, vertebral landmark detection is performed for Cobb angle measurement and assessment of scoliosis structure. And the segmentation module is utilized to detect lumbar vertebral pedicle. By combining the results from both modules, Lenke classification report is obtained.

2.1 | Establishment of image dataset

2.1.1 | Data collection

The dataset for this study consisted of whole spinal x-rays images between 2015 and 2021, retrospectively gathered from the First Affiliated Hospital of Sun Yat-sen University. In particular, images

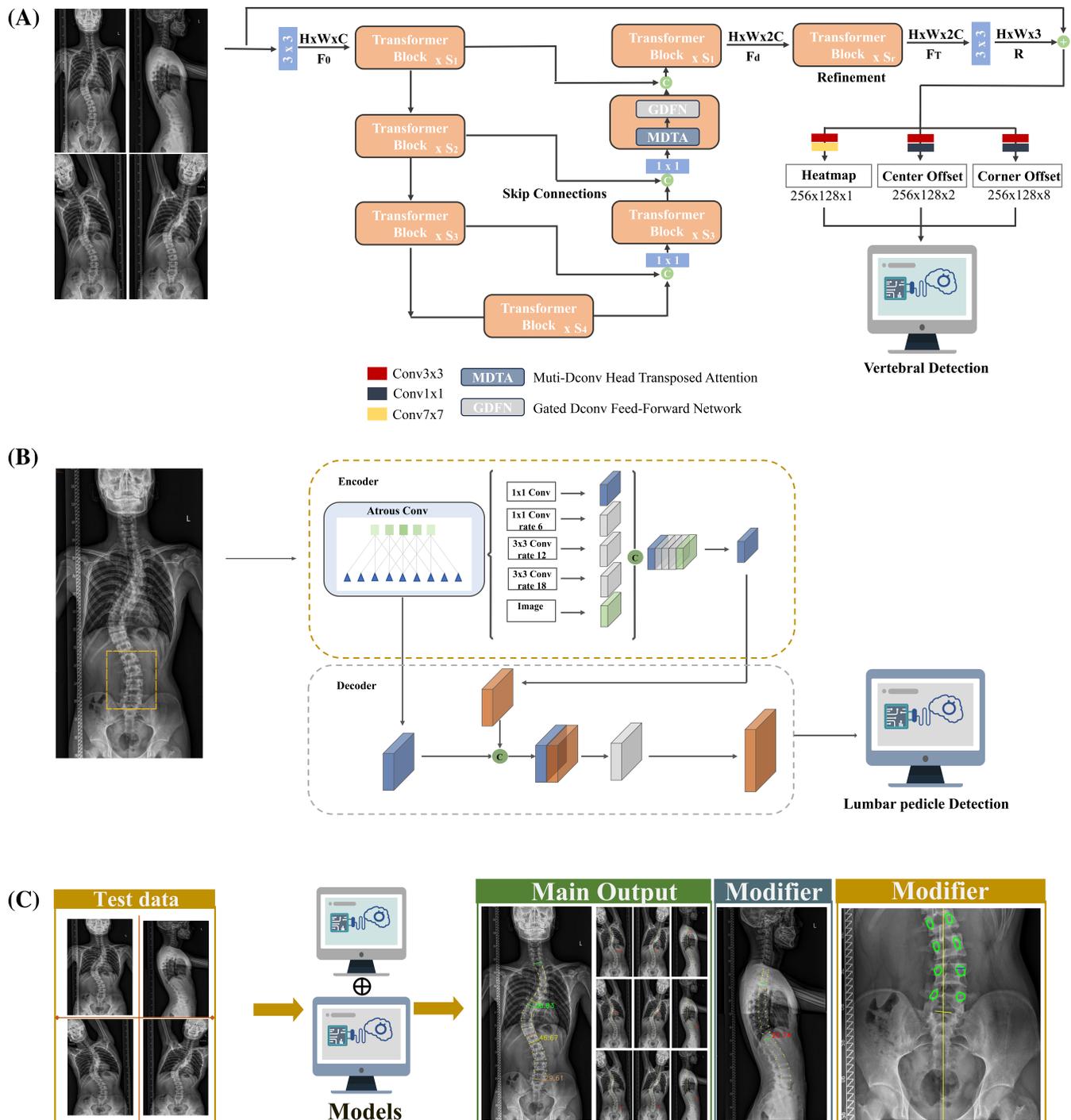


FIGURE 2 Fundamental architecture of construction for automated Lenke classification system. (A) Overview of development of the vertebral detection modules. The input to the model is the multi-view full-length spinal images. The *Restormer* framework is utilized for feature extraction and analysis and subsequently various modules such as Heatmap, center offset, and corner offset are employed to obtain landmarks of the vertebral bodies. These landmarks are crucial for further analysis and assessment of the spinal image. (B) The construction of lumbar pedicle detection modules. The overall approach utilizes the DeepLabV3+ algorithm framework to segment the pedicle of the lumbar vertebrae. Specifically, the encoder–decoder architecture helps capture both local and global contextual information to achieve accurate and precise segmentation results. (C) Test for the automated typing system. The relevant Cobb angle measurement in distinct images and the identification of lumbar pedicle and CSVL are displayed.

collection was divided into AIS and control group. As for AIS group, to preserve the credibility of the systems, enrolling criteria pertaining to basic characteristics were stipulated as follows: (1) Diagnosed as

AIS with age 10–18 years old; (2) No gender restriction; (3) Congenital bony deformity or degenerative scoliosis were excluded. Furthermore, data collection must possess 4 sorts of images, including anterior-

posterior (AP), lateral films, right and left bending films. Images that did not occupy intact clinical information were excluded prior to the subsequent image labeling tasks. Finally, the samples were randomly assigned into training datasets or test datasets at the ratio of 3:1. The control group comprised individuals who underwent regular health examinations and did not exhibit scoliosis, and images in the control group included both AP and lateral films. Specifically, the age and gender composition ratios of the control group remained essentially the same as those of the AIS group.

2.1.2 | Data preparation

Owing to the necessity to intake local information from interested regions according to Lenke classification standards, the feature labeling procedure did not include the cervical vertebrae. Therefore, for each x-ray image, the annotation regions included 17 vertebrae from the thoracic and the lumbar spine, as well as the upper boundary of sacrum. And each vertebra had 4 corner landmarks (Figure 3A), along with 2 landmarks from superior extremity of the sacrum. Those key points abovementioned were labeled by applying the *Surgimap* (detailed in *Software application*), and eventually each image occupied 70 landmarks. Besides, the silhouettes of lumbar pedicle (L1–L5) were portrayed by *labelme* (detailed in *Software application*) bilaterally with different tags, which was an essential portion for the decision of lumbar spine modifier (Figure 3A). All images were initially annotated by residents and then over-read or furtherly validated by professional spine surgeons with corrections made to the diagnostic labels as needed. The different types of AIS according to the Lenke classification in the training dataset and test dataset would be maintained in a similar pattern.

2.2 | Construction of the deep learning architecture

On the ground that the deep learning method possessed a capability to discriminatory and hierarchical representations from raw data, we utilized deep neural networks (DNN) as an elementary learning block. With the processed data and annotation, we trained the deep neural networks to make automated diagnosis for patients. In details, we divided the tasks into two parts: (1) initial vertebral landmark detection for AP, bending and lateral x-rays films; (2) lumbar pedicle detection for the judgment of lumbar spine modifier.

2.2.1 | Vertebral landmark detection

Considering that landmark detection can better simulate the process of clinical applications, we have continued to follow this approach and made improvements on the feature extraction module. Specifically, we firstly introduce the attention block with self-attention to model the long-range pixel dependencies in the x-ray images, since the

position of the vertebra is closely connected to each other. CNN-based models have limitations in modeling long-range pixel-level correlations due to their limited perceptual field, which can result in the insufficient capability to capture global structural information. Instead, attention-based models are not constrained by the perceptual field and can perform feature extraction and aggregation at the global dimension, leading to improved performance in various vision tasks. Therefore, we adopt the attention blocks from Restormer¹⁷ including multi-dconv head transposed attention (MDTA) and gated dconv feed-forward network (GDFN). Secondly, we also construct the feature extractor as a U-Net formation to further extract multi-scale feature of the image. Given a single image input, encoder captures the corresponding feature maps and up-convolution is conducted in decoder. Finally, inspired by,¹⁶ we define the regression head with three branches, which are defined as multi-layer convolutions in the model. the three branches¹⁶ captured three different outputs including of the center heatmap, center offset and corner offset. The center heatmap branch is assigned to localize the vertebrae by detecting their center points. Furthermore, instead of regressing the center points directly, we apply keypoint heatmap to define the center, which is generally used in pose joint localization and object detection. The center offset branch output center offset resulting from the down-sampling operation. For the purpose of recognizing the endplate, corner offset branch is designed to predict the vectors that start from the center and point to the vertebra corners. With the output of the center heatmap branch and center offset branch, we obtain the correct center points of the vertebra. Afterwards, the vectors captured by corner offset branch are added to the center points and the corner points are obtained. With the corner points of each vertebra, the inclining angles are obtained in order to conduct Lenke classification (Figure 2A).

2.2.2 | Cobb angle measurement and curvature categorization

The specific definition of the Cobb angle involves measuring the angle between the upper endplate of the highest vertebra and the lower endplate of the lowest vertebra (Figure 3A). For each individual vertebra, the upper endplate is defined by connecting the left-superior corner with the right-superior corner on the AP films and bending films, whereas on the lateral films, it is determined by connecting the posterior-superior corner with the anterior-superior corner. And the lower endplate is determined by connecting the left-inferior corner with the right-inferior corner on the AP films and bending films, while it is determined by connecting the posterior-inferior corner with the anterior-inferior corner on the lateral films (Figure 3A).

For AP x-ray films, each angle formed by different vertebrae is recursively computed by sequentially adjusting the upper and lower end vertebrae from the top to the bottom. Once the Cobb angle of the main curve is determined, similar methods are used to measure the Cobb angles in the upper and lower curves (Figure 3B). And after identifying each segment of the scoliosis, further measurements are

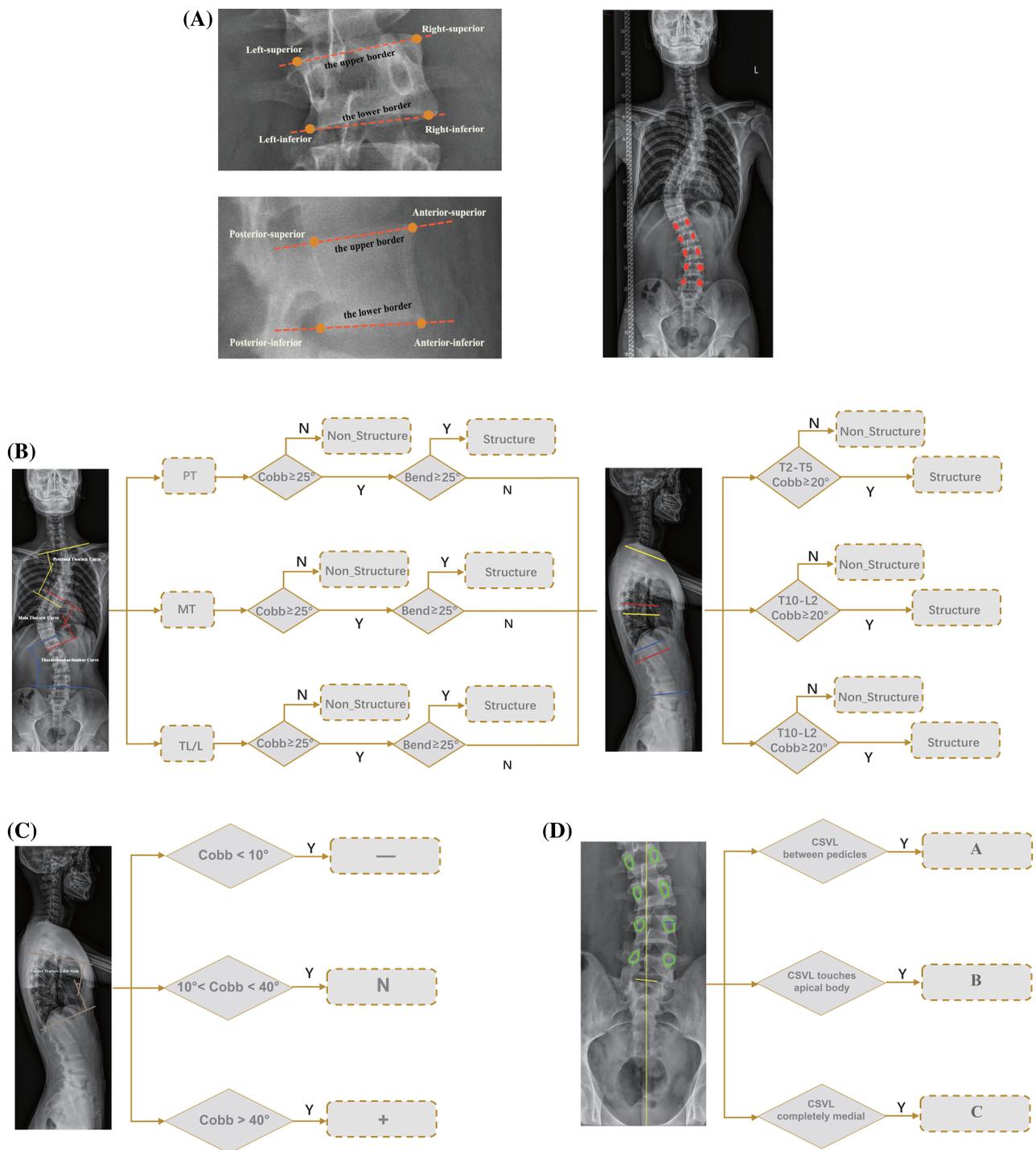


FIGURE 3 The procedure for determining the Lenke classification. (A) Annotation of vertebral landmarks and pedicle labeling on anteroposterior and lateral x-ray images. Each vertebral landmark on the anteroposterior image includes left-superior, right-superior, left-inferior, and right-inferior, while each vertebral landmark on the lateral image includes posterior-superior, anterior-superior, posterior-inferior, and anterior-inferior. Additionally, the upper border and lower border are formed by connecting the superior landmarks and inferior landmarks respectively on both anteroposterior and lateral images. (B–D) The criteria for determining the structural section of scoliosis, lumbar spine modifier and sagittal thoracic profile. Structure = structural curvature. Non_Structure = non-structural curvature. Y = the result of judgment is true. N = the result of judgment is false.

taken. This includes measuring the Cobb angle of the corresponding segments in the left and right bending films. Besides, the Cobb angles are measured between the upper endplate of T2 and the lower

endplate of T5 on lateral films, as well as between the upper endplate of T10 and the lower endplate of L2. And then the structural curve is identified according to the degree of side bending Cobb angles in

bending film and the degree of kyphosis in the corresponding segment of lateral films (Figure 3B). In addition, the sagittal thoracic modifier is classified by measuring the Cobb angle between the upper endplate of T5 and the lower endplate of T12 on lateral films (Figure 3C). Ultimately, by utilizing the aforementioned measurements of Cobb angles in each segment and the determination of structural scoliosis, a primary evaluation of the scoliotic type can be achieved.

2.2.3 | Lumbar pedicle detection

For lumbar pedicle detection, we employed the commonly used Deeplabv3+,^{18,19} a distinguished image segmentation learning framework, to produce semantically accurate predictions and detailed segmentation maps along interested object boundaries. Briefly, the overall architecture of this model still inherited the encoder-decoder frameworks. Among the encoder section, the elemental module consisted of DCNN (dilated convolution neural network) to implement feature extraction and ASPP (atrous spatial pyramid pooling) to complete the multiscale integration of contextual information.¹⁸ And in the decoder section, both low-level semantic feature and high-level one were coalesced to accomplish the multiple space representation integration and transform the feature vector to produce the segmentation maps. Specifically, for further refinement, open operation with circular structure was applied to filter the noise in the segmentation map. With combination of the segmentation map and the predicted CSVL, the spatial relationship between the pedicle of apical vertebra and CSVL was evaluated and the type of lumbar spine modifier was inferred according to Lenke classification standards (Figure 3D).

2.3 | Evaluation of the deep learning model

In the assessment of performance on the model, we used a new panel of radiologic test dataset with inclusion of the coronal, sagittal and bidirectional (left and right) bending films, which were input into model interface. And the consequences of output were compared with the professional clinical interpretation by senior spine surgeons and radiologists. First of all, we built the confusion matrix and evaluated the performance of different categorial tasks in terms of recall, specificity, precision and accuracy for the diagnosis. Particularly, we calculated the F1-score (the harmonic mean of sensitivity and specificity) to serve as overall evaluators in the performance of the models and implemented consistency test analysis of measurements in Cobb's angle of frontal scoliosis and corresponding sagittal kyphosis with statistics of intraclass correlation efficient (ICC) where the value ranged from 0 to 1, with 0 indicating no reliability among raters and 1 indicating perfect reliability among raters.²⁰ Moreover, we stratified the test dataset into different groups based on age and sex to perform the subgroup analysis. And in regard to evaluate the accuracy of object detection in lumbar pedicle segmentation model, we analyzed the intersection over union (IOU) of the target pixels and calculated dice coefficient (a spatial overlap index to assess the similarity between

predictive results and ground truth in segmentation model). Besides, the Hausdorff distance (HD) was used to evaluate the distance between the segmented surface and the ground truth surface. And a smaller HD indicates better segmentation accuracy. This metric was defined as:

$$HD = \max(h(S,GT),h(GT,S))$$

with

$$h(S,GT) = \max_{a \in S} \min_{b \in GT} \|a - b\|.$$

Among that, S represent the segmentation result while GT represent the ground truth.

2.4 | Software application

The images annotation tools in this study were included of *Surgimap* (*Surgimap*[®], a Nemaris Inc.[™] innovation) and *labelme* (open annotation tool, <http://labelme.csail.mit.edu>). The whole model construction procedures and training process were executed in a Linux server with an NVIDIA P100 GPU. The data preprocessing and model training and testing were constructed with Torch, version 1.9.0. The relevant statistical analyses including demographic characteristics' statistics, calculation of sensitivity, specificity, F1 scores and ICC coefficient were carried out in Python 3.9.5.

3 | RESULTS

3.1 | Basic characteristics of the datasets

Initially, for the construction of model dataset, we searched the radiographic datasets of the hospital and collected 600 patients of AIS totally. After exclusion of the ineligible objects, we obtained 407 cases in which the training dataset included 1470 images (from 300 patients and 135 normal objects) while the test dataset consisted of 612 images (from 107 patients and 92 normal objects). The relevant clinical characteristics are described in Table 1. The median age in the AIS group was 14 years (IQR 11–15) in the training dataset and 15 years (IQR 10–16) in the test dataset; 62% of patients in training datasets are women while 63% in test datasets. In addition, as illustrated in the table, the associated feature distribution of Lenke classification in both datasets which was approximately in accord with epidemiologic investigation in the adolescents.

3.2 | Performance of automated classification system for different cases

After a broad array of convolution transfer and subtle calibration of model parameters in our deep learning model, the results pertaining

TABLE 1 Baseline characteristics of datasets.

	Training dataset	Test dataset
Total number of images	1470	612
Total number of objects	435	199
AIS	300 (69%)	107 (54%)
Control group	135 (31%)	92 (46%)
Age, years		
AIS	14 (11–15)	15 (10–16)
Control group	15 (11–17)	13 (12–14)
Sex (in AIS)		
Male	115 (38%)	40 (37%)
Female	185 (62%)	67 (63%)
Lenke curve type in AIS		
Type 1 (MT)	162 (54%)	66 (62%)
Type 2 (DT)	60 (20%)	17 (16%)
Type 3 (DM)	30 (10%)	9 (8%)
Type 4 (TM)	9 (3%)	3 (3%)
Type 5 (TL/L)	33 (11%)	11 (9%)
Type 6 (TL/L-MT)	6 (2%)	4 (4%)

Abbreviations: DM, double major; DT, double thoracic; MT, main thoracic; TL/L, thoracolumbar/lumbar; TL/L-MT, thoracolumbar/lumbar-main thoracic; TM, triple major.

to vertebral-detection were displayed visually in Figure 4. Specifically, the ranges and Cobb angles of each segment for scoliosis on different x-ray images were shown. As established in Table 2, the classification evaluation indexes (F1-score) of different *Lenke* types in the vertebral detection model were 0.907, 0.653, 0.737, 0.857, 0.900 and 0.800 separately; the F1-score of the thoracic sagittal profile were 0.954, 0.948 and 0.800, respectively; and the results of the consistent test analysis showed the intraclass correlation efficient (ICC) of the Cobb angle measurement were 0.873–0.982, which indicated a high measurement consistency of the model (Table 3). Therefore, it was shown that as for the performance on multi-view detection such as AP and sagittal films, the model exhibited considerably great position-matching accuracy compared to ground truth.

The result for recognition of lumbar pedicle was shown in Figure 5. As expected, each vertebral pedicle in lumbar segment was discerned sequentially. In the analysis of performance for vertebra pedicle segmentation model, the intersection over union (IOU) of the target pixels was 0.827, the dice coefficient was 0.905 and Hausdorff distance (HD) was 6.565 ± 2.583 mm (Table 4). The F1-score in the lumbar modification judgment model was 0.942 on average, indicating considerable high accuracy in the performance. But in some conditions, the occurrence of abnormal spine rotation spatially engendered obvious disappearance of the pedicle shadow in the AP films, impeding the targeted area recognition. Furthermore, the

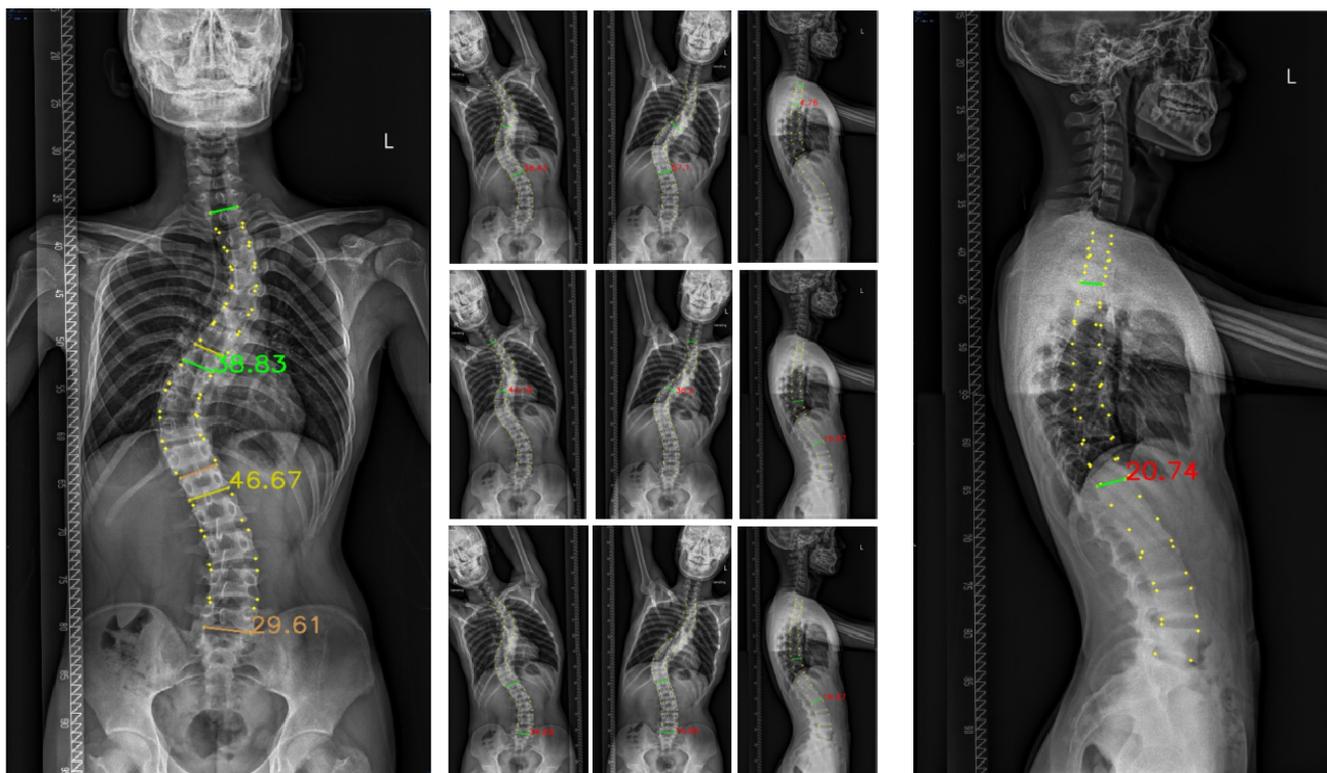


FIGURE 4 Representative examples for automated Lenke classification system in different radiologic images with exhibition of vertebral segmentation and automated measurement of Cobb's angle. As represented in the graphic, in the leftmost image, the corner landmarks of each vertebra are annotated with yellow point sequentially. And the spinal segment in scoliosis is partitioned spatially into three types: proximal thoracic (PT) curve, main thoracic (MT) curve and thoracolumbar/lumbar (TL/L) curve in which the Cobb's angles are depicted correspondingly (PT: green marker, MT: yellow marker, TL/L: brown marker). The measurement of Cobb's angles in bidirectional films and relevant sagittal sectional outputs are displayed in middle part of the plot. The rightmost image establishes the angle relative to sagittal modification profile.

TABLE 2 Performance for distinct categorization of automated Lenke classification system.

Type	Recall	Precision	Accuracy	F1-score
Lenke classification				
Lenke1	0.857	0.964	0.897	0.907
Lenke2	0.941	0.500	0.935	0.653
Lenke3	0.778	0.700	0.953	0.737
Lenke4	1	0.750	0.972	0.857
Lenke5	0.818	1	0.981	0.900
Lenke6	1	0.667	0.981	0.800
Overall performance	0.899	0.7635	0.953	0.809
Sagittal modification				
Hypo	0.933	0.977	0.963	0.954
Normal	0.948	0.948	0.944	0.948
Hyper	1	0.667	0.981	0.800
Overall performance	0.960	0.864	0.963	0.901
Lumbar modification				
A	0.941	0.960	0.953	0.950
B	0.913	0.955	0.972	0.934
C	0.970	0.914	0.963	0.941
Overall performance	0.941	0.943	0.963	0.942
Ultimate performance	0.934	0.847	0.938	0.885

spatial relationship between the paired pedicle on the lumbar apical vertebra with central sacral vertical line (CSVVL) was evaluated via image detection technique (Figure 5). The subgroup analyses in performance of both training models were elucidated in Tables 5 and 6, in which there were no distinctive discrepancies among sex and age group.

4 | DISCUSSION

Building on recent advancements in deep learning, we had constructed an automated-AIS-classification diagnostic system, which was capable of aiding clinicians to evaluate the severity of spine curvature and guiding the associated formulation of corrective surgery.

Generally, the object detection can be divided into anchor-based and anchor-free methods.²¹ Compared to conventional anchor-based approaches, the landmark based object detection, as a branch of anchor-free methods, had simplified the model structure and eliminated the manual adjustment of parameters for anchor configuration.^{22–24} And relevant researches had demonstrated that this method had established great performance in vertebral landmarks detection for x-ray films in comparison with other models such as regression or segmentation based approaches.¹⁶ Given that recognition patterns were similar in vertebral morphologic analysis for lateral x-rays film, we firstly deployed this approach in sagittal films detection and achieved an excellent performance in target recognition.

Additionally, in previous deep learning models for measuring Cobb angles, CNN modules were commonly used for feature

TABLE 3 The consistent test analysis of Cobb's angle measurement.

Category		ICC coefficient (95% CI)
Proximal thoracic	AP	0.973 (0.961–0.982)
	Bend-L	0.965 (0.950–0.976)
	Bend-R	0.982 (0.974–0.988)
Main thoracic	AP	0.976 (0.964–0.983)
	Bend-L	0.978 (0.969–0.985)
	Bend-R	0.977 (0.966–0.984)
TL/L	AP	0.966 (0.951–0.977)
	Bend-L	0.982 (0.974–0.988)
	Bend-R	0.950 (0.927–0.966)
PT kyphosis (T2–T5)		0.903 (0.860–0.932)
MT/L kyphosis (T10–L2)		0.873 (0.819–0.911)

Abbreviation: AP, anterior-posterior.

extraction and fusion.^{9,11–16} However, in recent times, *Transformer* modules have shown remarkable performance in various object detection tasks.¹⁷ Specifically, CNN-based models have limitations in capturing long-range pixel-level correlations due to their restricted receptive field. This can lead to insufficient representation of global structural information. Instead, the *Transformer* model utilizes self-attention operations to achieve input adaptability, capture long-range dependencies, and facilitate high-order spatial interactions.^{12,17} By establishing interaction mechanisms between different modules of the network, such as the encoder, decoder, and self-attention fusion

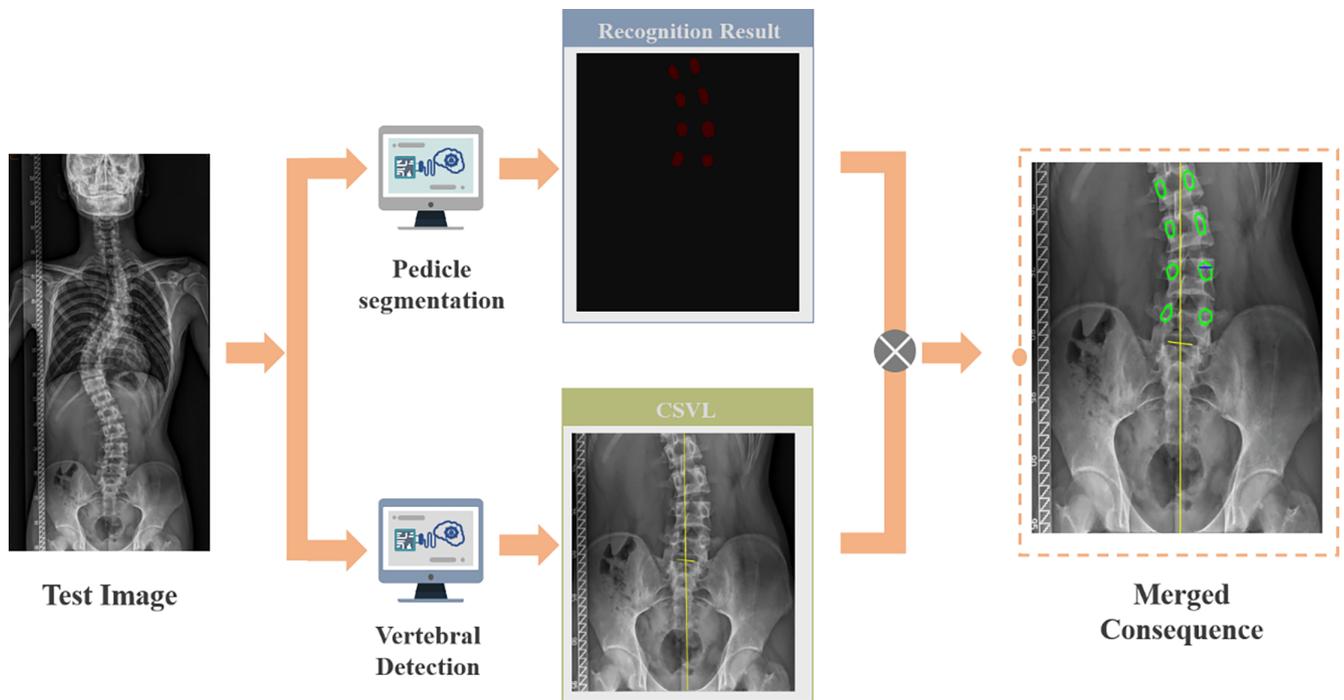


FIGURE 5 The identification for lumbar pedicle and CSVL. Each lumbar pedicle is labeled with green circle bilaterally while the upper sacral boundary and CSVL are drawn with yellow line. As established in image, the pedicle of apical vertebra does not touch the CSVL, therefore this situation belongs to A type in lumbar modification profile.

TABLE 4 Performance for vertebral pedicle segmentation model.

Model	Index	Recall	Accuracy	IOU	Dice coefficient	HD (mm)
Pedicle segmentation	Background	0.985	0.981	0.980	0.990	6.565 ± 2.583
	Target area	0.912	0.870	0.827	0.905	
	Mean	0.949	0.926	0.904	0.948	

Abbreviation: HD, Hausdorff distance.

TABLE 5 Subgroup analysis: Performance for distinct categorization of automated Lenke classification system by gender.

	Lenke classification model			Lumbar modification			Sagittal modification		
	Sensitivity	Specificity	F1	Sensitivity	Specificity	F1	Sensitivity	Specificity	F1
Female	0.92	0.95	0.93	0.95	0.97	0.96	0.81	0.97	0.88
Male	0.81	0.97	0.84	0.93	0.97	0.95	0.88	0.95	0.91

Note: F1 = an indicator for the harmonic mean of sensitivity and specificity.

TABLE 6 Subgroup analysis: Performance for distinct categorization of automated Lenke classification system by age.

	Lenke classification model			Lumbar modification			Sagittal modification		
	Sensitivity	Specificity	F1	Sensitivity	Specificity	F1	Sensitivity	Specificity	F1
[10, 15)	0.89	0.94	0.88	0.91	0.96	0.93	0.80	0.95	0.87
Age ≥15 years	0.88	0.97	0.81	0.95	0.98	0.96	0.90	0.97	0.93

Note: F1 = an indicator for the harmonic mean of sensitivity and specificity.

module, it enables the fusion of feature maps at different levels. This helps in capturing both intra-object and inter-object long-range contextual information.¹⁷ Therefore, in this study, we utilized a feature capturing structure based on the *Transformer* model, combined with landmark extraction, for vertebral detection.

As for semantic segmentation architecture, we used the DeepLabV3+ to accomplish the lumbar pedicle recognition. Previous literature had demonstrated several models for pedicle detection via traditional image descriptors including support vector regression (SVR) and histograms of oriented gradients (HOG)^{25,26}; However, those approaches required complex tuning of hyperparameters and computational costs, which limited the upgradation of segmentation efficiency. DeepLabV3+ was a novel visual segmentation architecture via the atrous convolution to extract multiple features and ASSP structure to capture information at different scales.²⁷⁻³⁰ And it had been used in medical image region and achieved a better performance in segmentation of radiologic images than other frameworks like FCN, SegNet, and U-Net.³¹⁻³⁴ Therefore, we tried this model in pedicle detection task and ultimate high-quality consequences for the first time revealed the potential of DeepLabV3+ to be useful for image segmentation in x-ray films.

In brief, this study shows the power of leveraging modern computing technology to potentially facilitate the diagnosis and treatments of AIS. Particularly, it is an integrated modality with object detection and semantic segmentation. When the raw images are input into the automated typing systems, it will automatically recognize the vertebral margin to realize the rapid spinal detection and calculate Cobb's angles in distinct spinal segments. Meanwhile, it demarcates lumbar pedicle to assess the extent of lumbar deviation and possesses high accuracy with a level of competence comparable to relevant experts.

The limitations in our works are demonstrated as follow. The scale of training samples should be expanded as far as possible, and larger datasets from diverse populations are able to improve the effectiveness and generalizability of learning models^{35,36} and achieve greater progress in abatement of fuzzy judgment for object identification. The amount in different types of Lenke classification do not possess the balanced distributions which predispose the excellent performance in some cases owning more samples.³⁷ And this study lacks validation on external datasets, so it is considered as preliminary research.

Future work could involve extensive augmentation in sample sizes for further optimization in performance stability and recognition accuracy³⁸ while multi-center datasets from different regions are indispensable sample sources for a deep learning model. In addition, the classification framework can be deployed in radiologic browser devices with enhancement in the convenience and feasibility for general application.

5 | CONCLUSION

Overall, we had built up an automated Lenke classification system by integrating the ability of images segmentation and feature

analysis of the deep neural networks. This approach can provide valuable guidance for the early diagnosis of scoliosis and assist in the formulation of surgical plans. However, the test set in our study had a small number of images, and therefore, the results of this experimental study need to be further validated in a larger sample size in the future.

AUTHOR CONTRIBUTIONS

B. Zhang, K. Cheng, R. Wang, and P. Su designed the research framework. B. Zhang, H. Yuan, T. Zhou, and W. Guo conducted the data collection and image annotations. K. Cheng constructed the deep learning networks. B. Zhang and K. Cheng performed the statistical analysis and wrote the manuscript. R. Wang and P. Su critically revised the manuscript. All the authors approved the final manuscript.

ACKNOWLEDGMENTS

We would like to show our sincere appreciation to the reviewers for their constructive comments on this article. Particularly, we appreciate the support of computing power from National Supercomputer Center (School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Due to the involvement of human participants, the accessibility of data is restricted in the manuscript. However, data are available to qualified researchers upon reasonable request from the corresponding authors and ethical approvals from the First Affiliated Hospital of Sun Yat-sen University. The relative codes of the overall frameworks of deep learning models are available upon reasonable request from the corresponding authors.

ORCID

Peiqiang Su  <https://orcid.org/0000-0001-5391-7558>

REFERENCES

- Weinstein SL, Dolan LA, Cheng JCY, Danielsson A, Morcuende JA. Adolescent idiopathic scoliosis. *Lancet*. 2008;371:1527-1537.
- Altaf F, Gibson A, Dannawi Z, Noordeen H. Adolescent idiopathic scoliosis. *BMJ (Clin Res Ed)*. 2013;346:f2508.
- Lenke LG. The Lenke classification system of operative adolescent idiopathic scoliosis. *Neurosurg Clin N Am*. 2007;18:199-206.
- Arlet V, Reddi V. Adolescent idiopathic scoliosis: Lenke type I-VI case studies. *Neurosurg Clin N Am*. 2007;18:e1-e24.
- Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018;172:1122-1131.e9.
- Olczak J, Fahlberg N, Maki A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthop*. 2017;88:581-586.
- Wang M, Xia C, Huang L, et al. Deep learning-based triage and analysis of lesion burden for COVID-19: a retrospective study with external validation. *Lancet Digital Health*. 2020;2:e506-e515.
- Pan Y, Chen Q, Chen T, et al. Evaluation of a computer-aided method for measuring the cobb angle on chest x-rays. *Eur Spine J*. 2019;28:3035-3043.

9. Horng M-H, Kuok C-P, Fu M-J, Lin C-J, Sun Y-N. Cobb angle measurement of spine from x-ray images using convolutional neural network. *Comput Math Methods Med.* 2019;2019:6357171.
10. Anitha H, Prabhu GK. Automatic quantification of spinal curvature in scoliotic radiograph using image processing. *J Med Syst.* 2012;36:1943-1951.
11. Wu H, Bailey C, Rasoulinejad P, Li S. Automatic landmark estimation for adolescent idiopathic scoliosis assessment using BoostNet. In: Descoteaux M, Maier-Hein L, Franz A, Jannin P, Collins DL, Duchesne S, eds. *Medical Image Computing and Computer Assisted Intervention—MICCAI 2017.* Springer International Publishing; 2017:127-135.
12. Yao Y, Yu W, Gao Y, et al. W-transformer: accurate cobb angles estimation by using a transformer-based hybrid structure. *Med Phys.* 2022;49:3246-3262.
13. Sun Y, Xing Y, Zhao Z, Meng X, Xu G, Hai Y. Comparison of manual versus automated measurement of Cobb angle in idiopathic scoliosis based on a deep learning keypoint detection technology. *Eur Spine J.* 2022;31:1969-1978.
14. Caesarendra W, Rahmaniar W, Mathew J, Thien A. Automated cobb angle measurement for adolescent idiopathic scoliosis using convolutional neural network. *Diagnostics (Basel, Switzerland).* 2022;12:12.
15. Sun H, Zhen X, Bailey C, Rasoulinejad P, Yin Y, Li S. Direct estimation of spinal cobb angles by structured multi-output regression. In: Niethammer M, Styner M, Aylward S, et al., eds. *Information Processing in Medical Imaging.* Springer International Publishing; 2017:529-540.
16. Yi J, Wu P, Huang Q, Qu H, Metaxas DN. Vertebra-focused landmark detection for scoliosis assessment. *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI).* IEEE; 2020:736-740.
17. Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang M-H. Restormer: Efficient Transformer for High-Resolution Image Restoration. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 2022: 5718-5729.
18. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell.* 2018;40:834-848.
19. Chen L-C, Papandreou G, Schroff F, Adam H. *Rethinking Atrous Convolution for Semantic Image Segmentation.* 2017. arXiv:abs/1706.05587.
20. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15:155-163.
21. Liu S, Zhou H, Li C, Wang S. Analysis of anchor-based and anchor-free object detection methods based on deep learning. *2020 IEEE International Conference on Mechatronics and Automation (ICMA).* ICMA; 2020:1058-1065.
22. Zhu C, Chen F, Shen Z, Savvides M. Soft anchor-point object detection. *Computer vision—ECCV 2020. ECCV 2020. Lect Notes Comput Sci.* 2020 :91-107.
23. Law H, Deng J. CornerNet: detecting objects as paired keypoints. *Proceedings of the European Conference on Computer Vision (ECCV).* ECCV; 2018:734-750.
24. Tychsen-Smith L, Petersson L. Denet: scalable real-time object detection with directed sparse sampling. *Proceedings of the IEEE International Conference on Computer Vision, ICV 2017:428-436.*
25. Ebrahimi S, Gajny L, Vergari C, Angelini ED, Skalli W. Vertebral rotation estimation from frontal x-rays using a quasi-automated pedicle detection method. *Eur Spine J.* 2019;28:3026-3034.
26. Cunha P, Moura DC, Barbosa JG. Pedicle detection in planar radiographs based on image descriptors. *ICIAI 2012: Image Analysis and Recognition.* Vol 7325. Springer Berlin Heidelberg; 2012:278-285.
27. Wang J, Liu X. Medical image recognition and segmentation of pathological slices of gastric cancer based on Deeplab v3+ neural network. *Comput Methods Programs Biomed.* 2021;207:106210.
28. Baheti B, Innani S, Gajre S, Talbar S. Semantic scene segmentation in unstructured environment with modified DeepLabV3+. *Pattern Recogn Lett.* 2020;138:223-229.
29. Polat H. A modified DeepLabV3+ based semantic segmentation of chest computed tomography images for COVID-19 lung infections. *Int J Imaging Syst Technol.* 2022;32:1481-1495. doi:10.1002/ima.22772
30. Ebigbo A, Mendel R, Scheppach MW, Probst A, Shahidi N, Prinz F, Fleischmann C, Römmele C, Goelder SK, Braun G, Rauber D, Rueckert T, de Souza LA, Papa J, Byrne M, Palm C, & Messmann H. Vessel and tissue recognition during third-space endoscopy using a deep learning algorithm. *Gut.* 2020; 71(12): 2388–2390.IF: NANANA
31. Khan Z, Yahya N, Alsaih K, Ali SSA, Meriaudeau F. Evaluation of deep neural networks for semantic segmentation of prostate in T2W MRI. *Sensors (Basel, Switzerland).* 2020;20:3183.
32. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder–decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell.* 2017;39(12):2481-2495.
33. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Lect Notes Comput Sci.* 2015;9351:234-241.
34. Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell.* 2017;39:640-651.
35. Fang Y, Wang J, Ou X, et al. The impact of training sample size on deep learning-based organ auto-segmentation for head-and-neck patients. *Phys Med Biol.* 2021;66:185012.
36. Rukundo O. *Effects of Image Size on Deep Learning.* 2021. arXiv:2101.11508v1.
37. Johnson JM, Khoshgoftaar TM. Deep learning and data sampling with imbalanced big data. *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI).* IEEE; 2019:175-183.
38. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition.* IEEE; 2009:248-255.

How to cite this article: Zhang B, Chen K, Yuan H, et al. Automatic Lenke classification of adolescent idiopathic scoliosis with deep learning. *JOR Spine.* 2024;7(2):e1327. doi:10.1002/jsp2.1327