Conditional generative adversarial network-assisted system for radiation-free evaluation of scoliosis using a single smartphone photograph: a model development and validation study

Zhong He,[a] Neng Lu,[b] Yi Chen,[a] Elvis Chun-Sing Chui,[c] Zhen Liu,[a] Xiaodong Qin,[a] Jie Li,[a] Shengru Wang,[d] Junlin Yang,[e] Zhiwei Wang,[f] Yimu Wang,[g] Yong Qiu,[a] Wayne Yuk-Wai Lee,[c] Jack Chun-Yiu Cheng,[c] Kenneth Guangpu Yang,[c] Adam Yiu-Chung Lau,[c] Xiaoli Liu,[c] Xipu Chen,[a] Wu-Jun Li,[b,h,i,**] and Zezhang Zhu[a,*]

[a]Division of Spine Surgery, Department of Orthopedic Surgery, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, China
[b]National Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, Nanjing, China
[c]Department of Orthopaedics and Traumatology, The Chinese University of Hong Kong, Hong Kong, China
[d]Department of Orthopedics, Peking Union Medical College Hospital, Beijing, China
[e]Spine Center, Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai, China
[f]Department of Orthopaedic Surgery, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
[g]David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Canada
[h]Center of Medical Big Data, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, China
[i]National Institute of Healthcare Data Science at Nanjing University, Nanjing, China

eClinicalMedicine
2024;75: 102779
Published Online xxx
https://doi.org/10.1016/j.eclinm.2024.102779

Summary

Background Adolescent idiopathic scoliosis (AIS) is the most common spinal disorder in children, characterized by insidious onset and rapid progression, which can lead to severe consequences if not detected in a timely manner. Currently, the diagnosis of AIS primarily relies on X-ray imaging. However, due to limitations in healthcare access and concerns over radiation exposure, this diagnostic method cannot be widely adopted. Therefore, we have developed and validated a screening system using deep learning technology, capable of generating virtual X-ray images (VXI) from two-dimensional Red Green Blue (2D-RGB) images captured by a smartphone or camera to assist spine surgeons in the rapid, accurate, and non-invasive assessment of AIS.

Methods We included 2397 patients with AIS and 48 potential patients with AIS who visited four medical institutions in mainland China from June 11th 2014 to November 28th 2023. Participants data included standing full-spine X-ray images captured by radiology technicians and 2D-RGB images taken by spine surgeons using a camera. We developed a deep learning model based on conditional generative adversarial networks (cGAN) called Swin-pix2pix to generate VXI on retrospective training (n = 1842) and validation (n = 100) dataset, then validated the performance of VXI in quantifying the curve type and severity of AIS on retrospective internal (n = 100), external (n = 135), and prospective test datasets (n = 268). The prospective test dataset included 268 participants treated in Nanjing, China, from April 19th, 2023, to November 28th, 2023, comprising 220 patients with AIS and 48 potential patients with AIS. Their data underwent strict quality control to ensure optimal data quality and consistency.

Findings Our Swin-pix2pix model generated realistic VXI, with the mean absolute error (MAE) for predicting the main and secondary Cobb angles of AIS significantly lower than other baseline cGAN models, at 3.2° and 3.1° on prospective test dataset. The diagnostic accuracy for scoliosis severity grading exceeded that of two spine surgery experts, with accuracy of 0.93 (95% CI [0.91, 0.95]) in main curve and 0.89 (95% CI [0.87, 0.91]) in secondary curve. For main curve position and curve classification, the predictive accuracy of the Swin-pix2pix model also surpassed that of the baseline cGAN models, with accuracy of 0.93 (95% CI [0.90, 0.95]) for thoracic curve and 0.97 (95% CI [0.96, 0.98]), achieving satisfactory results on three external datasets as well.

*Corresponding author. Division of Spine Surgery, Department of Orthopedic Surgery, Nanjing Drum Tower Hospital, Affiliated Hospital of Medical School, Nanjing University, No.321 Zhongshan Road, Nanjing 210008, China.
**Corresponding author. National Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, No.163 Xianlin Road, Nanjing 210023, China.
E-mail addresses: zhuzezhang@nju.edu.cn (Z. Zhu), liwujun@nju.edu.cn (W.-J. Li).

Translated abstract For the Chinese translation of the abstract, see the Supplementary Materials section.

Interpretation Our developed Swin-pix2pix model holds promise for using a single photo taken with a smartphone or camera to rapidly assess AIS curve type and severity without radiation, enabling large-scale screening. However, limited data quality and quantity, a homogeneous participant population, and rotational errors during imaging may affect the applicability and accuracy of the system, requiring further improvement in the future.

Funding National Key R&D Program of China, Natural Science Foundation of Jiangsu Province, China Postdoctoral Science Foundation, Nanjing Medical Science and Technology Development Foundation, Jiangsu Provincial Key Research and Development Program, and Jiangsu Provincial Medical Innovation Centre of Orthopedic Surgery.

Keywords: Adolescent idiopathic scoliosis; Radiation-free; Virtual X-ray images; Conditional generative adversarial networks

Research in context

Evidence before this study

PubMed was searched on January 28, 2024 for all research articles containing the keywords "artificial intelligence" OR "deep learning" AND "radiation-free" OR "no-radiation" AND "scoliosis", without any date or language restrictions. Further searches were conducted for references cited in the articles. Nine relevant studies were identified, which used techniques including RGBD images, 2D-RGB images, raster stereographic back images, ultrasound images, and point clouds of the back for radiation-free identification of scoliosis. These studies mostly had limitations due to (1) limited data volume; (2) the need for specialized equipment for collection and analysis; (3) accuracy or visualization levels not meeting clinical needs; and (4) insufficient multi-centre external data validation. 2D-RGB images are the easiest type of data for patients to obtain in daily life, and currently, only a few studies have attempted to use deep learning techniques to classify and predict using 2D-RGB images, and no teams have successfully used 2D-RGB images to generate medical images.

Added value of this study

To our knowledge, this study represents the first development of an innovative cGAN system capable of generating realistic spin X-ray images from 2D-RGB images captured by common 2D cameras or smartphones, achieving results comparable to those generated from RGBD images in previous research. It holds promise for future screening in medical centres, schools, communities, and homes, significantly enhancing the accessibility of scoliosis screening.

Implications of all the available evidence

The radiation-free scoliosis detection platform developed in this study, based on 2D photographs, can efficiently, accurately, and conveniently assess the curve type and severity of scoliosis. However, firstly, there is room for improvement in both the quantity and quality of the dataset used in this study, which may affect the accuracy and generalizability of the model. Secondly, this study only includes the Han Chinese population from different regions, lacking application data from other ethnic groups, hence it is uncertain whether the procedure can be extended to other ethnicities. Finally, due to the inability to completely eliminate rotational errors during imaging, this may be a contributing factor to the remaining differences between fake X-rays and real X-rays.

Introduction

Adolescent Idiopathic Scoliosis (AIS) is a three-dimensional spinal deformity that occurs in individuals aged 10–18 years and is characterized by a spinal curvature of 10° or more. AIS is the most common spinal disorder among adolescents, with a prevalence rate of 1%–4% in this age group, and female adolescents are more susceptible. The exact etiology of AIS is not fully understood.[1–4] Mild scoliosis often presents no noticeable symptoms. However, moderate to severe scoliosis progression can lead to back deformities, respiratory dysfunction, paralysis, and, in extreme cases, can be life-threatening.[5–7] Therefore, early screening and follow-up treatment are crucial for patients with AIS, which allow spine surgeons to intervene promptly and control the rapid progression of the curvature.[6,8]

Currently, the clinical diagnosis of AIS primarily relies on full-spine standing X-ray imaging. However,[9] this diagnostic method has several drawbacks: ① it may cause radiation exposure damage to adolescents; ② in most primary healthcare facilities, only chest radiographs can be taken, potentially leading to missed diagnoses, particularly of lumbar curvature; ③ patients may be unable to undergo screening or regular follow-up due to factors such as economic, transportation, healthcare resources, and epidemics. To solve these problems, radiation-free scoliosis assessment methods have been extensively researched and developed, which include traditional methods like appearance inspection, forward

bending tests, scoliometer measurements, and Moiré topography.[10–13] However, these traditional assessment methods not only require significant human resources but also lead to large measurement errors and high rates of missed diagnoses.

In recent years, with the rapid development of deep learning technologies, this challenge is likely to be overcome. Conditional Generative Adversarial Networks (cGAN)[14] have demonstrated remarkable performance in the medical imaging field,[15–17] with one of the key advantages being their ability to learn data distributions and mapping relationships, facilitating powerful image generation capabilities beneficial for tasks requiring data transformation in the absence of data. To our knowledge, the application of cGAN in the field of spine deformity is very limited. Zhang et al.[18] was the first to propose using deep cameras to collect the RGB-Depth (RGBD) images of the backs of patients with AIS and generate radiograph-comparable images (RCI) through anatomical point identification registration and style transfer methods for precise scoliosis assessment, significantly improving screening efficiency and accuracy. However, this method still requires professionals to use specialized equipment for screening, which is not available in most underdeveloped areas. In the extant models, paired datasets were not employed, meaning there was no direct correlation between the patients' X-ray images and the two-dimensional Red Green Blue (2D-RGB) back images. Nonetheless, we maintain that this foundation is amenable to facilitating a significant technological innovation. Could we further break through on this basis, use correlated information and develop a system that allows patients or their guardians without professional knowledge to assess themselves without the aid of other special equipment?

In this study, we aim to develop the first AIS screening system for use by general population or professionals in more schools, communities and hospitals, using commonly available 2D cameras to capture the 2D-RGB images of the backs of patients with AIS, and generating Virtual X-ray Images (VXI) to assist spine surgeons in rapidly, accurately, and non-invasively assessing the severity and type of scoliosis in patients with AIS, along with providing treatment recommendations. We built a paired dataset and for the first time developed an innovative Swin-pix2pix network structure. Experiments conducted on retrospective and prospective paired datasets of 2D-RGB back appearance images and X-ray images (ground truth X-ray image, GT) from four hospitals in Mainland China have shown that our method can produce high-quality VXI, featuring optimal assessment accuracy and robustness.

Methods

Study design and participants

This manuscript adhered to STROBE guidelines. We collected retrospective multi-centre data from four medical institutions in China from June 11th 2014 to March 17th 2022: Nanjing Drum Tower Hospital (NJDTH), Peking Union Medical College Hospital (PUMCH), Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine (XHASJU), and the Second Affiliated Hospital Zhejiang University School of Medicine (SAHZU). The inclusion criteria were as follows: (1) diagnosed with AIS; (2) underwent standing full-spine posterior-anterior, lateral, and bending X-ray; (3) had bare back 2D-RGB images taken. The exclusion criteria included: (1) age under 11 years or over 18 years; (2) having any disease that could affect standing posture; (3) history of spine surgery; (4) obstructions in the back area in the 2D-RGB images or any appearance abnormalities other than AIS; (5) unequal leg lengths; (6) incomplete clinical or imaging data. Some patients, although their main curve Cobb angle was less than 10° after conservative treatment, were diagnosed as AIS and included in this study to supplement the insufficient number of samples of mild scoliosis in the training dataset.

Additionally, from April 19th 2023 to November 28th 2023, we prospectively recruited consecutive participants from NJDTH. The inclusion criteria were as follows: (1) diagnosed with AIS or potential AIS ("Potential AIS" is defined as adolescents with a Cobb angle of less than 10° who do not meet the diagnostic criteria for AIS. These patients are usually identified during school physical examinations when doctors suspect AIS and are then referred for X-ray examinations at the request of their guardians); (2) underwent standing full-spine posterior-anterior and lateral X-ray, and additional bending X-rays for patients with AIS. (3) willing to have bare back 2D-RGB images taken. The exclusion criteria included: (1) age under 11 years or over 18 years; (2) having any disease that could affect standing posture; (3) history of spine surgery; (4) back area abnormalities other than AIS; (5) unequal leg lengths; (6) patients with AIS who could not undergo Lenke classification due to the absence of spine bending X-rays. The specific collection requirements for prospective and retrospective data are detailed in the eAppendix 1.

After careful review and screening, a total of 2177 retrospective patients with AIS and 268 prospective participants were included. Among the retrospective patients with AIS, 2042 were from NJDTH, 50 from PUMCH, 20 from XHASJU, and 65 from SAHZU (Fig. 1). The baseline demographic characteristics of the included patients are shown in Table 1.

Nanjing University Medical School Affiliated Drum Tower Hospital has obtained approval from the authoritative institutional medical ethics committee (approval number 2022-609-01), following the unified protocol of the ethics committee. It is also registered in the National Medical Research Registration Information System (registration number MR-32-23-049234). All

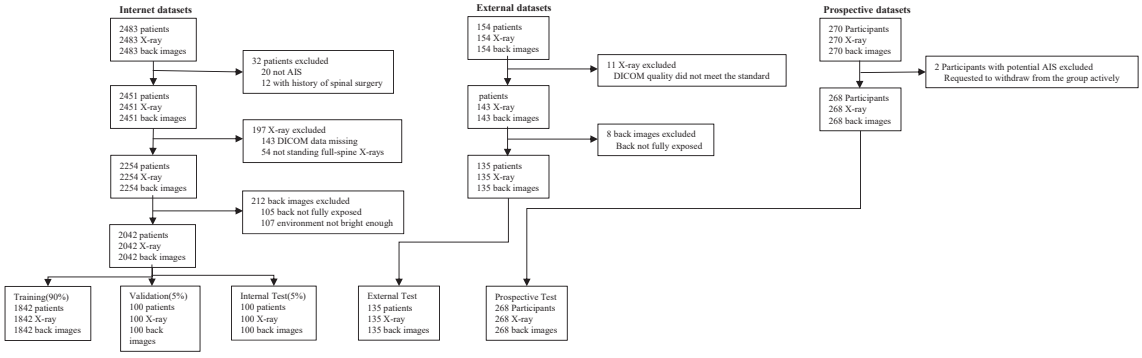

***Fig. 1:* Study flow diagram of participant enrolment.** The final cohort included 2445 participants. For training, validation, and internal testing, 1842, 100, and 100 participants were included, respectively. The external testing and prospective testing included 135 and 268 participants, respectively.

recruited participants and their guardians provided informed consent for this trial.

Datasets distribution

All included patients with AIS from the four medical institutions underwent standing full-spine X-ray examinations by radiology technicians and were photographed by spine surgeons with a 2D camera, obtaining the coronal X-ray images and back appearance 2D-RGB images. All X-ray images were in the Digital Imaging and Communications in Medicine (DICOM) format. The two sets of image data were paired and named using the same ID number. After data annotation, the retrospective dataset from NJDTH was randomly divided into training, validation, and internal testing datasets. The retrospective datasets from PUMCH, XHASJU, and SAHZU were all used as external testing datasets. The prospective dataset from NJDTH was used as testing dataset too. The specific dataset distribution is as follows:

(1) Training dataset: Uses X-ray images and back 2D-RGB images of 1842 patients with AIS from NJDTH to train the model.
(2) Validation dataset: Uses X-ray images and back 2D-RGB images of 100 patients with AIS from NJDTH to validate the model and select the best hyperparameters.
(3) Internal testing dataset: Uses X-ray images and back 2D-RGB images of 100 patients with AIS from NJDTH for internal testing of the model.
(4) External testing dataset 1: Uses X-ray images and back 2D-RGB images of 50 patients with AIS from PUMCH for external testing of the model.
(5) External testing dataset 2: Uses X-ray images and back 2D-RGB images of 20 patients with AIS from XHASJU for external testing of the model.
(6) External testing dataset 3: Uses X-ray images and back 2D-RGB images of 65 patients with AIS from SAHZU for external testing of the model.
(7) Prospective testing dataset: Uses X-ray images and back 2D-RGB images of 220 patients with AIS and 48 potential patients with AIS from NJDTH for prospective testing of the model.

Data preprocessing

To enhance the Signal-to-Noise Ratio (SNR), computational efficiency, interpretability, and accuracy of the input images (back 2D-RGB and X-ray images), we established an elaborate automated data preprocessing workflow (Fig. 2, eAppendix 3), with manual adjustments made for data that failed automatic processing. The specific steps are as follows:

(1) Collection of back 2D-RGB images: The quality of X-ray generation is susceptible to variables such as the patient's posture, rotational deviations, and environmental conditions. Although the retrospective datasets from the four centres did not adhere to strict shooting protocols, they were collected by designated personnel following standardized procedures to ensure a certain degree of data consistency. In contrast, the prospective datasets had stringent requirements for shooting equipment, distance, height, lighting conditions, and trunk rotation (eAppendix 1), which were aimed at reducing the impact of environmental differences on shooting errors. Additionally, during image processing, precise cropping around key anatomical landmarks (such as the acromion and pelvic regions) on the 2D-RGB photos and X-rays was performed to standardize scales and align the datasets. Including images obtained under various environmental conditions also increased the diversity of the training dataset, thereby enhancing the model's feature recognition capabilities and robustness across different settings.
(2) Object detection in X-ray images: We used a portion of the X-ray image data to train a

Datasets (number)	Training (1842)	Validation (100)	Internal test (100)	External test			Prospective test (268)	P
				PUMCH(50)	XHASJU (20)	SAHZU (65)		
Gender (male/female)	411/1431	29/71	23/77	12/38	4/16	12/53	47/221	0.34
Age (years)	14.4 ± 2.3	14.7 ± 1.9	14.8 ± 2.1	14.3 ± 2.2	14.8 ± 1.6	14.3 ± 1.7	14.5 ± 2.4	0.57
Risser	3.3 ± 1.7	3.0 ± 1.9	3.4 ± 1.7	3.5 ± 1.7	3.3 ± 2.2	2.8 ± 1.6	3.1 ± 1.5	0.10
BMI (kg/m2)	18.5 ± 2.4	18.8 ± 2.2	18.6 ± 2.1	18.4 ± 1.4	18.6 ± 1.9	18.9 ± 1.6	18.6 ± 1.4	0.15
Cobb angle (degree)								
Main curve	50.1 ± 12.9	50.5 ± 13.3	49.1 ± 13.1	38.5 ± 20.7	31.3 ± 16.9	49.9 ± 9.1	40.6 ± 20.3	<0.01
Secondary curve	30.3 ± 10.8	30.8 ± 10.1	29.8 ± 11.4	23.2 ± 15.6	21.3 ± 14.5	30.7 ± 9.2	25.4 ± 10.4	<0.01
Scoliosis severity grading								
Main curve								
1	42	3	2	11	7	0	58	<0.01
2	131	5	7	18	5	4	68	
3	1346	72	75	13	6	50	89	
4	248	16	13	5	2	11	38	
5	75	4	3	3	0	0	15	
Secondary curve								
1	356	21	19	18	9	3	118	<0.01
2	1141	59	63	22	9	53	105	
3	303	16	16	4	2	8	35	
4	32	2	2	3	0	1	7	
5	10	2	0	0	0	0	3	
Main curve position								
T	1253	69	67	25	11	43	182	0.18
TL/L	589	31	33	25	9	22	86	
Curve classification								
1	927	55	51	16	9	35	127	0.72
2	120	5	6	4	0	1	15	
3	125	8	7	2	2	4	27	
4	61	4	3	3	0	2	13	
5	560	26	30	24	8	22	77	
6	49	2	3	1	1	1	9	

Continuous value was presented with mean ± standard deviation.

Table 1: **Demographic information of the datasets.**

YOLOv8 model for object detection on X-ray images. This process extracted bounding box coordinates for the entire spine (T1-L5) to the pelvic region from the X-ray images, cropping the images based on these coordinates to ensure that the primary information (i.e., spinal curvature in the X-ray images) was highlighted while reducing or eliminating other unnecessary information and noise (Supplementary eFig. S2 and eTable S2).

(3) Instance segmentation of back 2D-RGB images: An enhanced Swin-YOLOv8 model (eAppendix 3), trained on a subset of back 2D-RGB image data, was employed for instance segmentation of back 2D-RGB images. This identified and segmented the back and pelvic regions from the back 2D-RGB images, cropping these images to match the spatial occupancy ratio of the human body in X-ray images. This ensured that the main information (i.e., the spine in the back 2D-RGB images) was highlighted, while other unnecessary information and noise were reduced or removed. These initial two steps significantly improved the SNR of the data images and enhanced performance upon application. The cropped back 2D-RGB images and X-ray images were then aligned to ensure spatial consistency, minimizing differences between the two image types and enhancing the quality of the transformation (Supplementary eFig. S3 and eTable S2).

(4) Image resizing and stitching: To meet the model input data requirements and ensure uniform input image scale, all back 2D-RGB images and X-ray images underwent standardization, being uniformly resized to 512 × 1280 pixels. Subsequently, the processed back RGB images and corresponding X-ray images were horizontally stitched along the width direction.

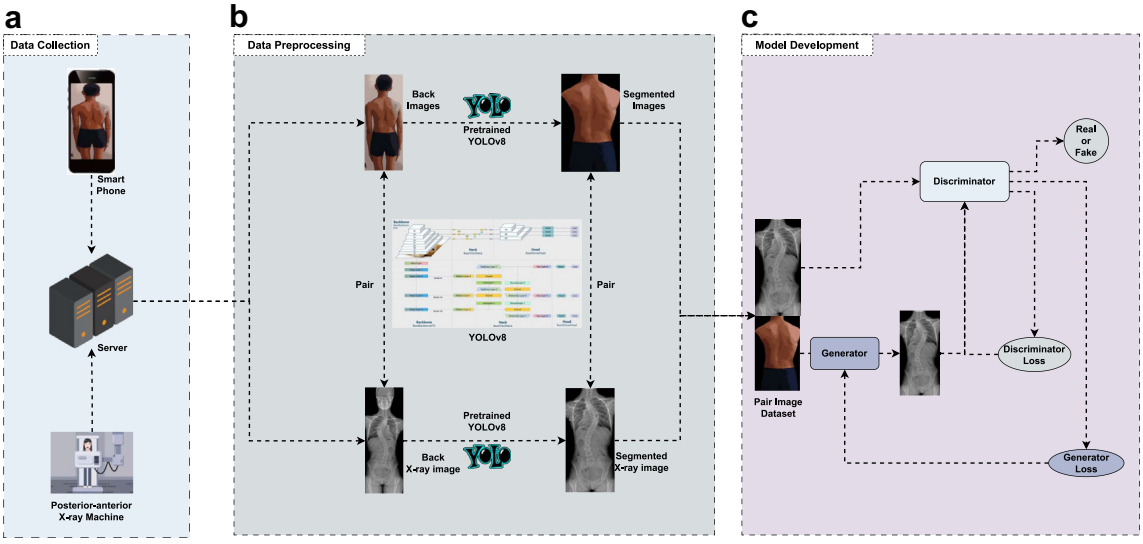

Fig. 2: **The pipeline of system for radiation-free evaluation of scoliosis based on back 2D-RGB images.** This system comprises several components: a) Data Collection Module: Back image data of patients with AIS are collected using smartphones and a posterior-anterior X-ray machine and uploaded to the server. b) Preprocessing Module: A portion of the collected data is annotated, and the YOLOv8 model is trained to complete segmentation and detection tasks, cropping and standardizing 2D-RGB and X-ray images. c) Model Development Module: The preprocessed image data are input into the cGAN model for training. The assessment performance of the model is improved by discriminating between the generated VXI and real X-ray images. Abbreviations: YOLOv8: You Only Look Once Version 8; VXI: Virtual X-ray images are defined as synthetic X-rays generated by a model from back photographs. Clinically meaningful VXI includes key anatomical landmarks such as vertebrae, shoulders, and pelvis. These X-rays allow doctors to identify and assess AIS by measuring these landmarks.

Model development

In this study, we developed a cGAN named Swin-pix2pix, which can convert the back 2D-RGB images into standing full-spine posterior-anterior X-ray images and accurately predict the curve type and severity of patients with AIS. In our proposed Swin-pix2pix model, we innovatively replaced the downsampling convolutional layers of the generator UNet-256 structure in the original pix2pix[19] with Swin modules (Fig. 3), which are implemented by the Swin Transformer in the mmseg package.[20–22] Due to the limited scale of our dataset, we did not employ the full Swin Transformer model. Instead, we meticulously selected the most critical component—the Swin block—to construct a network architecture tailored to our specific needs. This innovative network structure enhances the model's performance on small datasets by reducing the number of parameters and incorporating local window attention mechanisms and hierarchical feature fusion. Additionally, Liu et al., the authors of Swin Transformer, have also showcased its performance across multiple datasets,[20] including smaller datasets like CIFAR-10 and CIFAR-100. They reported that Swin Transformer achieves highly competitive results on these datasets, further confirming the effectiveness of the Swin Transformer architecture for small datasets.

These Swin modules can capture long-distance dependencies by using self-attention mechanisms, thus enhancing the global receptive field of the generator and learning more rich feature information on the back 2D-RGB images. The datasets that performed well on the original pix2pix model have different surface appearances for the input and output, but they share the same basic structure. For our study, the back 2D-RGB images and X-ray images have a huge structural span, so theoretically, improving the global receptive field can help capture more back appearance information (trunk contour, back texture, etc.) and strengthen the connection between the two datasets.

(1) Training:

In the training phase, we first preprocessed the input dataset and resized it to 512×1280 pixels. Given the input back image X and X-ray image Y of size $512 \times 1280 \times 3$, we denoted the generator as $G(\cdot)$ and the discriminator as $D(\cdot)$. The model learned the mapping $G : x \rightarrow y$ from the back image $x \in X$ to the X-ray image $y \in Y$. We fed the processed back 2D-RGB image and X-ray image pairs into the model for training. The generator transformed the back image into a VXI, and passed the VXI and the back image to the discriminator, which judged whether the image pair was real or

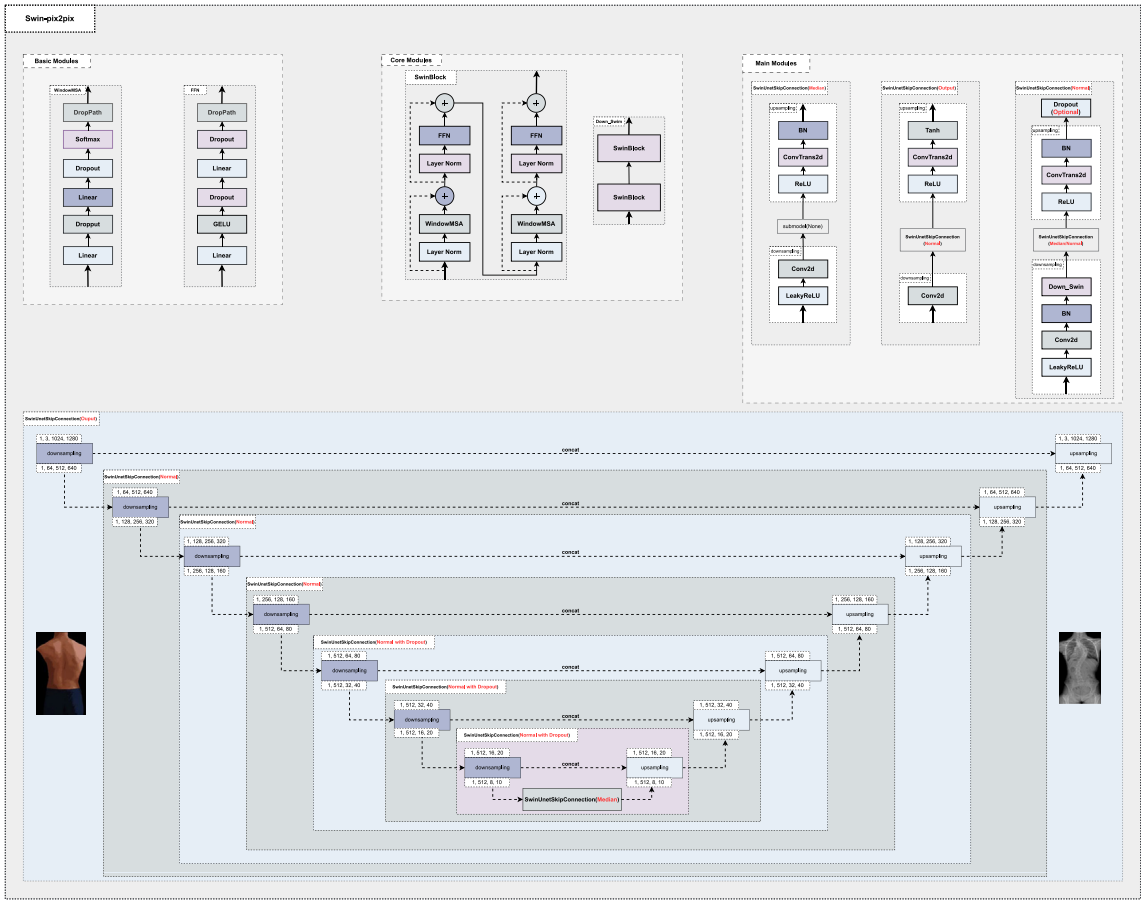

***Fig. 3:* The network architecture of the generator within the swin-pix2pix model.** Innovatively, this study replaces the down sampling convolutional layers in the original pix2pix generator's UNet-256 structure with Swin Transformer blocks. These Swin modules use the self-attention mechanism to capture long-range dependencies, thereby enhancing the generator's global receptive field. This allows for learning richer feature information on back 2D-RGB images, improving the model's ability to generate detailed and accurate VXI from the input 2D-RGB images. Abbreviations: VXI: Virtual X-ray Image.

not. To make the generator fool the discriminator and make the discriminator believe that the generated image pair was real, we used a loss function $\mathcal{L}_{\mathrm{G}}$ that consisted of two parts:

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda \times \mathcal{L}_{pix} = -\mathbb{E}_x\left[log\left(D\left(G(x), y\right)\right)\right] - \lambda \times \mathbb{E}_{x,y}\left[\|B - G(x)\|_1\right]$$

Here, $\mathcal{L}_{\mathrm{adv}}$ is a binary cross-entropy loss function that is opposite to the discriminator, giving a high probability to the generated image pair. $\mathcal{L}_{\mathrm{pix}}$ is an L1 norm loss function that measures the difference between the generated and real images. The generator's loss function $\mathcal{L}_{\mathrm{G}}$ is controlled by the weight λ.

The discriminator's task is to judge whether the output image is a real X-ray image or a virtual X-ray image generated by the generator, given the input image. To make the generator give a high probability to the real image pair and a low probability to the generated image pair, we use a binary cross-entropy loss function:

$$\mathcal{L}_D = \mathbb{E}_{x,y}\left[log\left(D(y, x)\right)\right] + \mathbb{E}_x\left[log(1 - D(G(x), x))\right]$$

Here, the discriminator's loss function $\mathcal{L}_{\mathrm{D}}$ is the sum of the losses for the real and generated image pairs.

The generator and the discriminator have a competitive and cooperative relationship. Based on the generator's loss function and the discriminator's loss function, the weights of the generator W_G and the discriminator W_D are updated until the generator can generate output images that match the target images.

(2) Inferencing

In the inference phase, the back 2D-RGB images are fed to the generator, which transforms it into a VXI. The VXI and the back 2D-RGB image are then fed to the

discriminator, which determines whether they are a matching pair. If the pair matches, the discriminator considers it as a real image. Otherwise, it considers it as a fake image. This process is repeated until the generator produces a VXI that matches the back image. After the inference phase is over, the output real back 2D-RGB image and VXI are resized to the pixel size before the resizing step in the preprocessing step (3), and the real X-ray image is resized to the original X-ray image pixel size.

(3) Experiment setup
We implemented all the experiments using the PyTorch toolkit. During the training process, we used the Adam optimizer with momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$, and an initial learning rate of 0.0002. We divided the training process into 200 epochs, keeping the learning rate constant for the first 100 epochs and linearly decaying it for the next 100 epochs. We chose a batch size of 1 and a lambda of 10 for the experimental setting, because the experimental results on the validation dataset showed that the quality of VXI was the highest when when choosing this set of hyperparameters.

We employed a grid search approach to precisely select the hyperparameters, optimizing them based on computational metrics and the Cobb angle of images. Furthermore, the number of training epochs was determined by observing the trend of the loss curve (eAppendix 5 and Supplementary eFig. S6).

Definition of scoliosis severity grading, main curve position and curve classification

To assess the clinical significance of this project, we defined three classification standards based on the clinical treatment strategies for AIS: scoliosis severity grading, main curve position, and curve classification (eAppendix 2).

(1) Scoliosis Severity Grading: Participants are classified into five levels based on the Cobb angle: Level 1: 0–19°; Level 2: 20–39°; Level 3: 40–59°; Level 4: 60–79°; Level 5: ≥80°.
(2) Main Curve Position: Participants are classified into two categories based on the location of the main curve: T (thoracic curve) where the apex is located between T2 and T11/T12 intervertebral discs; TL/L (thoracolumbar/lumbar curve) where the apex is located at T12-L1 for thoracolumbar or between L1/L2 intervertebral disc and L4 for lumbar.
(3) Curve Classification: Since this system can only generate coronal VXI and cannot produce sagittal or bending position images, to assess the accuracy of the cGAN model in diagnosing curve types, we propose the curve classification as follows: If the Cobb angle of the curve (X) in the GT coronal X-ray image is ≥ 25°, and the Cobb angle of the same curve (Y) in the bending position image does not improve to <25°, we consider the curve in the GT as structural. Otherwise, it is considered non-structural. Similarly, if in the VXI, the same curve's Cobb angle (Z) is ≥ 25°, and Z-(X–Y) is ≥ 25°, then the curve in the VXI is also considered structural; otherwise, it is non-structural. Based on the position of structural and non-structural curves in the upper T, T, and TL/L, AIS can be categorized into six types (eAppendix 2, Supplementary eFig. S1 and eTable S1).

Statistical analysis

To quantitatively evaluate the linear correlation between the actual Cobb angle on GT and the predicted Cobb angle on VXI, and to analyse the independent risk factors affecting the mean absolute error (MAE), we used linear regression analysis. Linear regression analysis is one of the most commonly used statistical methods, which is used to establish a linear relationship model between one or more independent variables and a dependent variable. It estimates the model parameters by minimizing the sum of squared residuals, thus finding the best fitting line or hyperplane.

To further compare the consistency between the actual Cobb angle on GT and the predicted Cobb angle on VXI, we also used Bland–Altman analysis. Bland–Altman analysis (or difference-mean plot) is a graphical method for measuring the agreement between two sets of measurements. It compares the measurements of the two methods, identifies the systematic and random errors, and evaluates the consistency of the two methods.

For scoliosis severity grading, main curve position and curve classification, we calculated five diagnostic test parameters, including accuracy, sensitivity, PPV and NPV, which are defined and calculated as follows, where TP is True Positives, TN is True Negatives, FP is False Positives, and FN is False Negatives:

(1) Accuracy: The proportion of correctly classified samples (whether positive or negative) out of the total number of samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

(2) Sensitivity: The proportion of positive samples that are correctly identified as positive out of all the actual positive samples.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

(3) Specificity: The proportion of negative samples that are correctly identified as negative out of all the actual negative samples.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

(4) PPV: The proportion of actual positive samples out of all the samples that are identified as positive.

$$PPV = \frac{TP}{TP + FP}$$

(5) NPV: The proportion of actual negative samples out of all the samples that are identified as negative.

$$\text{NPV} = \frac{TN}{TN + FN}$$

All the statistical analyses were performed using Python (v3.8.16) and several Python packages, including Numpy (v1.23.5), SciPy (v1.10.1), Statsmodels (v0.14.0), OpenCV (v4.7.0.72), PyTorch (v1.12.1), Matplotlib (v3.7.1), and Pandas (v1.5.3).

Demographic data analysis was performed using SPSS 25.0 statistical software (SPSS Inc., Chicago, IL). The Chi-square test was utilized to compare categorical variables, including gender, Risser, scoliosis severity grading, main curve position and curve classification, across seven different datasets. Repeated measures ANOVA was employed to compare continuous variables. A P value less than 0.05 was considered statistically significant.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to the data in the study and had final responsibility for the decision to submit for publication.

Results

Datasets

This study included a total of 2445 participants, comprising 1842 in the training dataset, 100 in the internal test dataset, 135 in the external test datasets, and 268 in the prospective datasets (Fig. 1). We recorded patient demographics including gender, age, Risser sign, BMI, Cobb angles (for both the main and secondary curves), scoliosis severity grading (levels 1–5 for both the main and secondary curves), main curve position (T and TL/L), and curve classification (levels 1–6). Detailed demographic information is available in Table 1. The analysis of variance conducted on various datasets revealed statistically significant differences in Cobb angle and scoliosis severity grading. Most notably, the prospective test dataset exhibited a considerably wider distribution of Cobb angles compared to other groups. The significant variability in Cobb angles allows for a more rigorous assessment of the model's performance. By evaluating the model against a dataset that mirrors the diverse characteristics of the broader population, we can better understand its effectiveness in real-world scenarios. This is particularly important in clinical applications where the ability to accurately predict outcomes across diverse patient groups is paramount.

Performance of our Swin-pix2pix and three baseline models for VXI synthesis

We evaluated the performance of Swin-pix2pix against three baseline models—pix2pix,[19] pix2pixHD,[23] and cycleGAN[24] in synthesizing VXI on the internal test, external, and prospective dataset. Representative VXI generated by these models are shown in Fig. 4. The paired back 2D-RGB images and spinal X-ray images, after preprocessing with the You Only Look Once version 8 (YOLOv8) model, showed strong consistency and corresponding internal structures. The preprocessed back 2D-RGB images were input into Swin-pix2pix and the three baseline models. Except for CycleGAN, all models generated clinically meaningful VXI. Swin-pix2pix has demonstrated exceptional comprehensive performance (Supplementary eTable S3), notably securing the top position in retrospective datasets and sharing the lead in prospective datasets. In both datasets, it achieved the highest scores for peak signal-to-noise ratio (PSNR) and learned perceptual image patch similarity (LPIPS), indicating its superior accuracy in image reconstruction and its outstanding visual quality. The model also excelled in maintaining structural information, as evidenced by its top-tier ranking in structural similarity index (SSIM) for retrospective datasets and a strong second place in prospective datasets. Furthermore, its Fréchet inception distance (FID) scores were second-best in both datasets, significantly outperforming the third-ranked model, showcasing its robust capabilities in content and style consistency of images. Although its natural image quality evaluator (NIQE) scores were second in retrospective and third in prospective datasets, they were closely competitive with the top-ranked model, reflecting the model's high-level performance in image naturalness. Overall, these metrics collectively affirm the model's exceptional performance across multiple critical aspects (eAppendix 6).

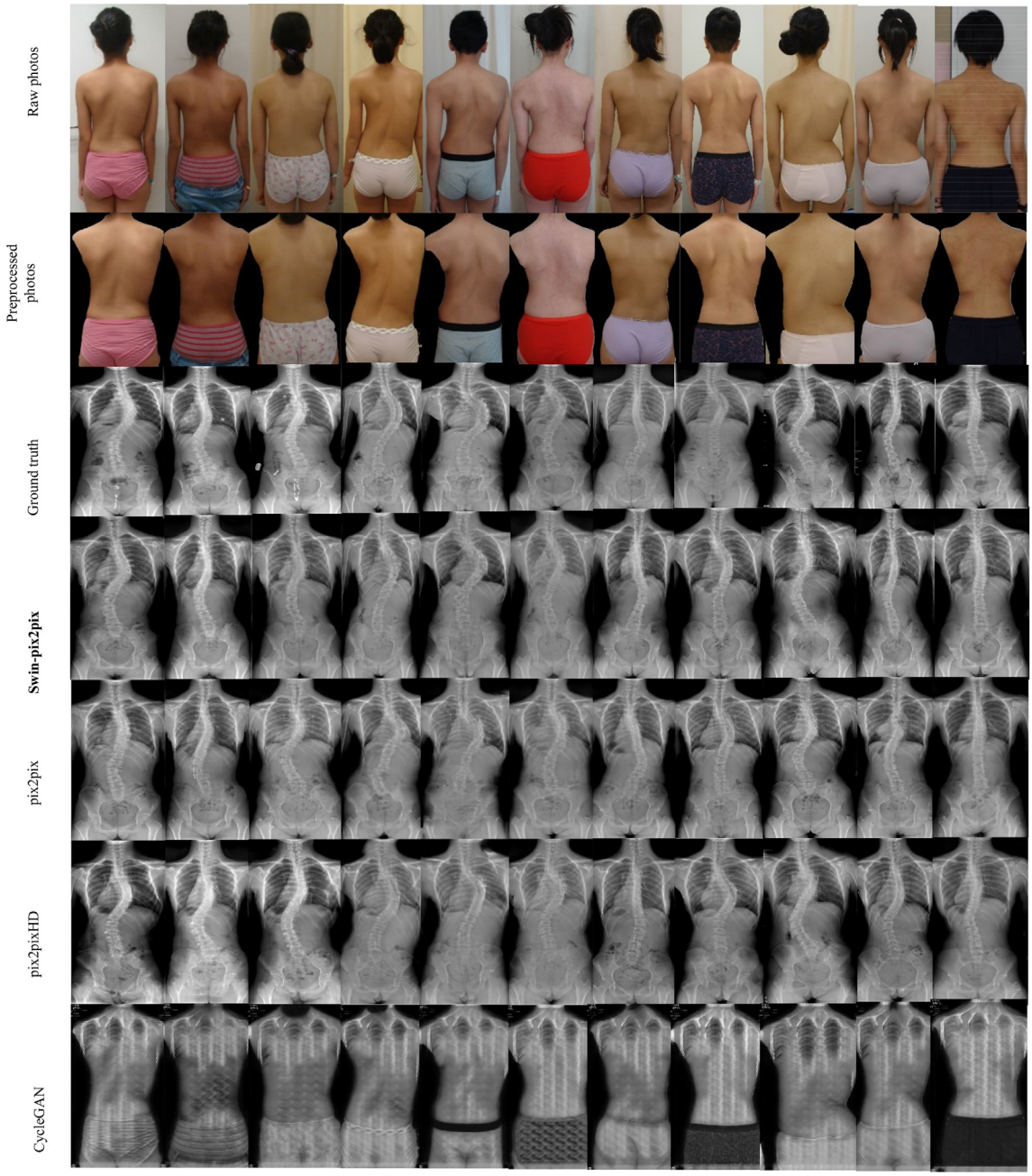

Fig. 4: **Comparison of representative VXI generated by swin-pix2pix and three baseline models.** Each column of images is from the same patient. The first row (Raw photos) shows the collected original back 2D-RGB images. The second row (Preprocessed photos) displays the back 2D-RGB images after preprocessing with the YOLOv8 model. The third row (ground truth) presents the spine X-ray images post-preprocessing with the YOLOv8 model. The fourth to the seventh rows depict the VXI generated by our Swin-pix2pix, pix2pix, pix2pixHD, and CycleGAN, respectively. Abbreviations: YOLOv8: You Only Look Once version 8; VXI: Virtual X-ray Image.

We measured the main and secondary Cobb angles of spinal curvature in the VXI produced by the three groups of models with internal, external and prospective test datasets (Fig. 5, Fig. 6) and calculated the MAE, root mean squared error (RMSE), coefficient of determination (R^2), and correlation coefficient (r) values between the VXI and the GT. Detailed results are shown in Table 2 (eAppendix 6). Compared to the two baseline models, Swin-pix2pix achieved the best results across all three datasets. Specifically, it performed the best on the

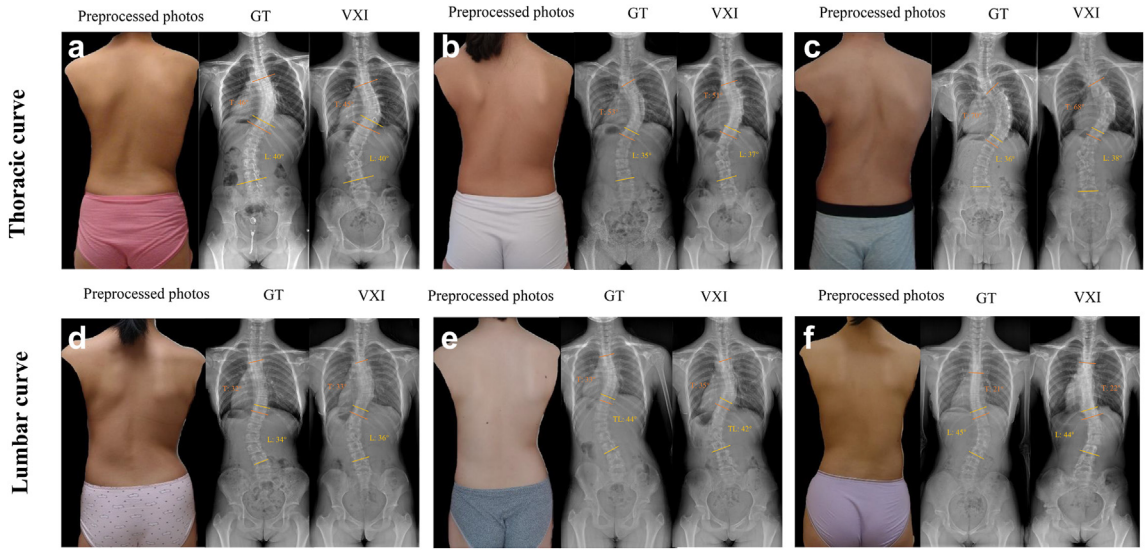

Fig. 5: **Measurement results of representative VXI generated by Swin-pix2pix.** Each subplot represents a case of a patient with AIS, with the first row for patients whose main curve is located at T (thoracic curve) and the second row for patients whose main curve is located at TL/L (thoracolumbar/lumbar curve). Each subplot, from left to right, displays the preprocessed photos, GT, and VXI. The images show that GT and VXI have almost identical main and secondary Cobb angles and consistent main curve positions and scoliosis classifications, indicating that our Swin-pix2pix generated VXI can accurately assess the type and severity of scoliosis in patients with AIS. Abbreviations: GT: Ground Truth X-ray Image; VXI: Virtual X-ray Image; T: Thoracic Curve; TL/L: Thoracolumbar/Lumbar Curve.

prospective test set, with the MAE and RMSE between the GT and VXI for the main curve being 3.2° and 4.1°, respectively, and for the secondary curve being 3.1° and 4.1°, respectively. In the internal test set, the MAE and RMSE between the GT and VXI for the main curve were 3.3° and 4.5°, respectively, and for the secondary curve were 3.3° and 4.3°, respectively. In the external test set, the MAE and RMSE between the GT and VXI for the main curve were 3.8° and 5.1°, respectively, and for the secondary curve were 3.9° and 4.9°, respectively. Compared to the prospective dataset, the Swin-pix2pix model's performance was slightly less impressive on the retrospective test sets. This may be due to the lack of strict control over factors such as imaging equipment, shooting distance, shooting height, and lighting conditions. Additionally, we conducted linear regression analysis and Bland-Altman analysis on the internal test set, external test set, and prospective test set. The Swin-pix2pix model exhibited the highest R^2 and the smallest standard deviation (SD) in all three datasets. This indicates that the model demonstrated the best goodness-of-fit and consistency between the GT and the generated VXI for the scoliosis Cobb angle in all three datasets (Fig. 7).

A multivariate linear regression analysis was conducted on the MAE to assess the impact of demographic information on the accuracy of VXI generated by Swin-pix2pix. Among the six included variables (age, gender, Risser sign, BMI, Cobb angle, curve classification), gender, BMI, and Cobb angle were identified as independent risk factors affecting the accuracy of VXI. Females, larger BMI, and larger Cobb angles were associated with increased discrepancies between VXI and GT (Supplementary eTable S4).

Performance on scoliosis severity grading, main curve position, and curve classification

For scoliosis severity grading, we presented the quantitative performance of three models and two spine surgery experts across the five grades and total (Level 1: 0–19°; Level 2: 20–39°; Level 3: 40–59°; Level 4: 60–79°; Level 5: ≥80°) (Supplementary eTable S5). We evaluated the predictive performance for scoliosis severity grading, main curve position, and curve classification on the internal test dataset, external test dataset, and prospective test dataset, recording the evaluation results using confusion matrices (Fig. 8, Fig. 9). Overall, the model's performance on the prospective test set was superior to that on the internal test set, both of which outperformed the external test set. Given that the prospective test set was collected in a real-world setting, we use it as an example. For scoliosis severity grading, we presented the quantitative performance of three models and two spine surgery experts across the five grades and total (Level 1: 0–19°; Level 2: 20–39°; Level 3: 40–59°; Level 4: 60–79°; Level 5: ≥80°) (Supplementary eTable S5). Total results show that, for the main curve, Swin-pix2pix achieved the best results in accuracy (0.93 [0.91, 0.95]), sensitivity (0.81 [0.75, 0.87]), specificity (0.95 [0.94, 0.96]), PPV (0.86 [0.81, 0.89]), and NPV

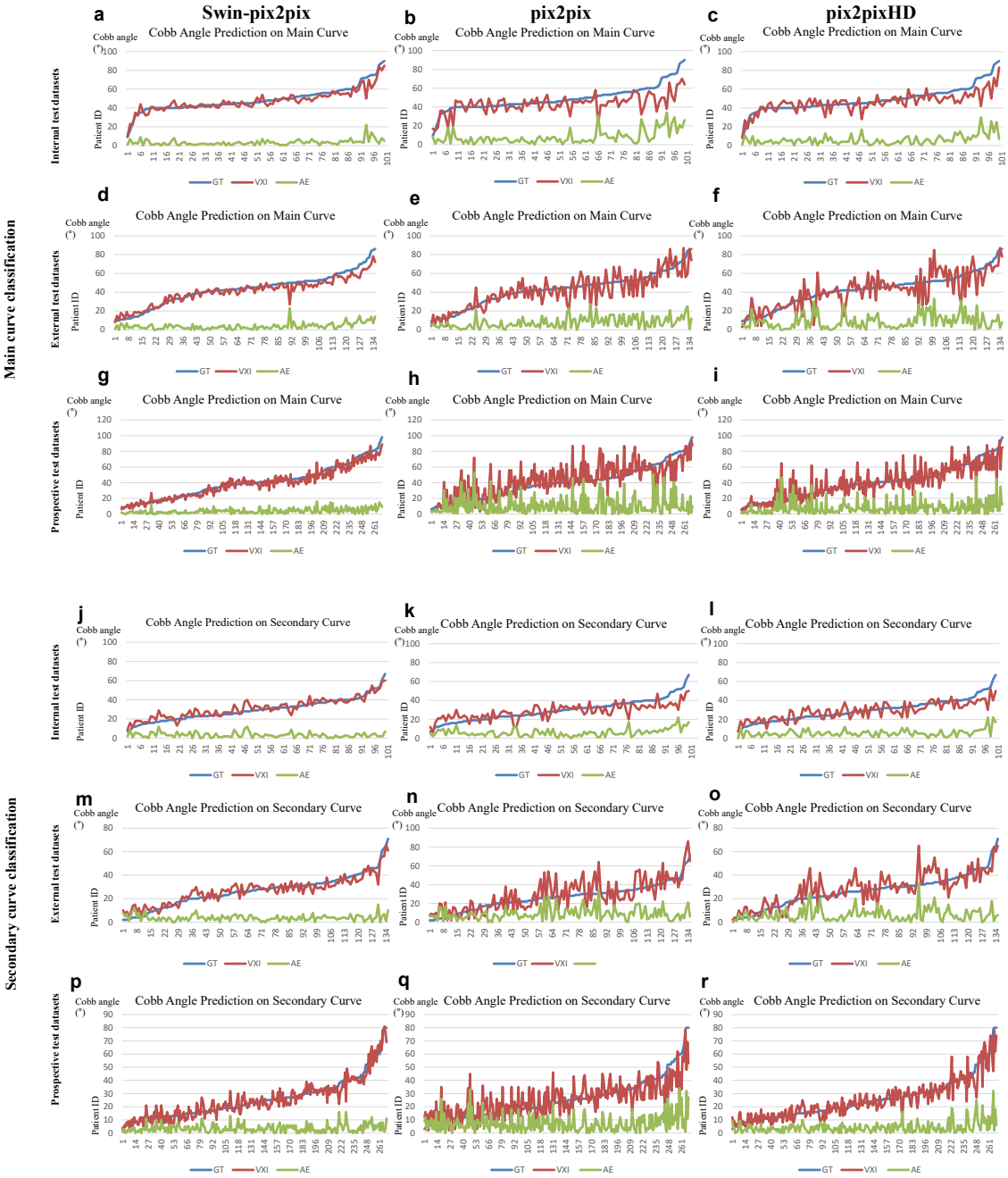

Fig. 6: **Cobb angle assessment of VXI generated by Swin-pix2pix and two baseline models.** The first to third columns in the figure respectively represent the assessment results of Swin-pix2pix, pix2pix, and pix2pixHD. The top three rows show the results for the main curve, and the bottom three rows show the results for the secondary curve, each containing the results for the internal test set, external test set, and prospective test set. In each subplot, the x-axis represents the participant ID number, the y-axis represents the Cobb angle, the blue line indicates the Cobb angle for GT, the red line represents the Cobb angle for VXI, and the green line represents the Absolute Error (AE) between the two. The figure demonstrates that for both the main and secondary curves, the Cobb angle curves for GT and VXI in our Swin-pix2pix model almost overlap, with AE significantly lower than that of pix2pix and pix2pixHD. When the Cobb angle increases, the AE in VXI of all three models also increases. Abbreviations: GT: Ground Truth X-ray Image; VXI: Virtual X-ray Image.

Evaluation metrics	Model	Test datasets	Cobb angle (degree)		MAE	RMSE	R^2	r
			GT	VXI				
Main curve	Swin-pix2pix	Internal	49.1 ± 13.1	47.6 ± 11.4	3.3	4.5	0.88	0.95
		External	43.0 ± 15.6	41.0 ± 14.9	3.9	5.2	0.91	0.97
		Prospective	40.6 ± 20.3	39.3 ± 18.9	3.2	4.1	0.94	0.95
	pix2pix	Internal	49.1 ± 13.1	43.5 ± 10.0	8.1	11.0	0.29	0.69
		External	43.0 ± 15.6	42.1 ± 15.8	8.7	11.9	0.53	0.71
		Prospective	40.6 ± 20.3	35.4 ± 19.4	7.6	10.4	0.44	0.77
	pix2pixHD	Internal	49.1 ± 13.1	46.1 ± 10.4	6.1	8.1	0.61	0.81
		External	43.0 ± 15.6	41.8 ± 14.8	6.5	8.2	0.67	0.78
		Prospective	40.6 ± 20.3	37.8 ± 20.5	5.9	7.5	0.67	0.86
Secondary curve	Swin-pix2pix	Internal	29.8 ± 11.4	31.5 ± 10.3	3.3	4.3	0.86	0.94
		External	26.5 ± 12.8	27.0 ± 11.6	3.9	4.9	0.88	0.91
		Prospective	25.4 ± 10.4	25.5 ± 10.3	3.1	4.1	0.91	0.93
	pix2pix	Internal	29.8 ± 11.4	28.5 ± 11.9	6.0	7.3	0.59	0.78
		External	26.5 ± 12.8	29.2 ± 13.2	8.7	7.9	0.33	0.73
		Prospective	25.4 ± 10.4	26.2 ± 12.3	5.7	7.4	0.57	0.77
	pix2pixHD	Internal	29.8 ± 11.4	29.0 ± 8.9	5.5	6.7	0.65	0.81
		External	26.5 ± 12.8	28.3 ± 13.6	6.3	7.2	0.60	0.79
		Prospective	25.4 ± 10.4	25.0 ± 9.6	4.9	5.8	0.83	0.85

GT: Ground truth image; VXI: Virtual X-ray image; MAE: Mean absolute error; RMSE: Root mean squared error; R^2: Coefficient of determination; r: Pearson correlation coefficient.

Table 2: **Evaluation metrics on cobb angle prediction between VXI from three models and GT.**

(0.95 [0.94, 0.97]), surpassing pix2pix, pix2pixHD, and both spine surgery experts. Assessing the secondary curve is more challenging than the main curve; nonetheless, Swin-pix2pix achieved the best results in accuracy (0.89 [0.87, 0.91]), sensitivity (0.79 [0.61, 0.86]), specificity (0.91 [0.72, 0.93]), PPV (0.78 [0.60, 0.86]), and NPV (0.92 [0.72, 0.93]), with accuracy, specificity, and NPV close to those of human 1, and other metrics significantly better than the baseline models. For the main curve position, we presented the quantitative performance for T, TL/L, and total (Table 3). Total results indicate that Swin-pix2pix achieved the best results in accuracy (0.93 [0.90, 0.95]), sensitivity (0.94 [0.90, 0.97]), specificity (0.92 [0.86, 0.97]), PPV (0.96 [0.93, 0.98]), and NPV (0.88 [0.81, 0.94]) for thoracic curve, slightly higher than the two baseline models. For curve classification, we displayed the quantitative performance across the six grades and total (Table 3), with Swin-pix2pix achieving the best total accuracy (0.97 [0.96, 0.98]), sensitivity (0.90 [0.83, 0.96]), specificity (0.98 [0.97, 0.99]), PPV (0.87 [0.80, 0.93]), and NPV (0.98 [0.97, 0.99]), where sensitivity and PPV showed a noticeable advantage over the two baseline models.

Comparison of Swin-pix2pix performance across three external test datasets

We compared the results among the external paired datasets from the three spine deformity centres to further evaluate the performance of the Swin-pix2pix model. We input the back 2D-RGB images from these three hospitals into the trained Swin-pix2pix and compared the output VXI with the collected GT, with test results shown in Supplementary eFig. S7. We measured the main and secondary Cobb angles of spinal curvature in the VXI generated from the three datasets (Supplementary eFig. S8) and calculated the MAE, RMSE, R^2, and r values between the VXI and GT, with detailed results presented in Supplementary eTable S6. The predictive MAE for the main curve on PUMCH, XHASJU, and SAHZU were 4.6°, 2.4°, and 3.8°, respectively, and for the secondary curve were 3.5°, 3.4°, and 4.1°, respectively.

We also assessed the predictive performance for scoliosis severity grading, main curve position, and curve classification on the three external test sets, recording the evaluation results using confusion matrices (Supplementary eFigs. S9 and S10). For scoliosis severity grading, the accuracy rates for the main curve predictions on PUMCH, XHASJU, and SAHZU were 0.89 [0.82, 0.95], 0.94 [0.91, 0.96], and 0.92 [0.87, 0.95], respectively, and for the secondary curve were 0.98 [0.97, 0.99], 0.94 [0.91, 0.96], and 0.96 [0.94, 0.97], respectively (Supplementary eTable S7). For the main curve position, the accuracy rates on PUMCH, XHASJU, and SAHZU were 0.84 [0.77, 0.92], 0.79 [0.67, 0.92], and 0.86 [0.81, 0.90], respectively. For curve classification, the accuracy rates on PUMCH, XHASJU, and SAHZU were 0.92 [0.88, 0.97], 0.91 [0.87, 0.94], and 0.94 [0.92, 0.96], respectively (Supplementary eTable S8).

Discussion

This study introduced an X-ray image generation system based on back 2D-RGB images, comprising an image

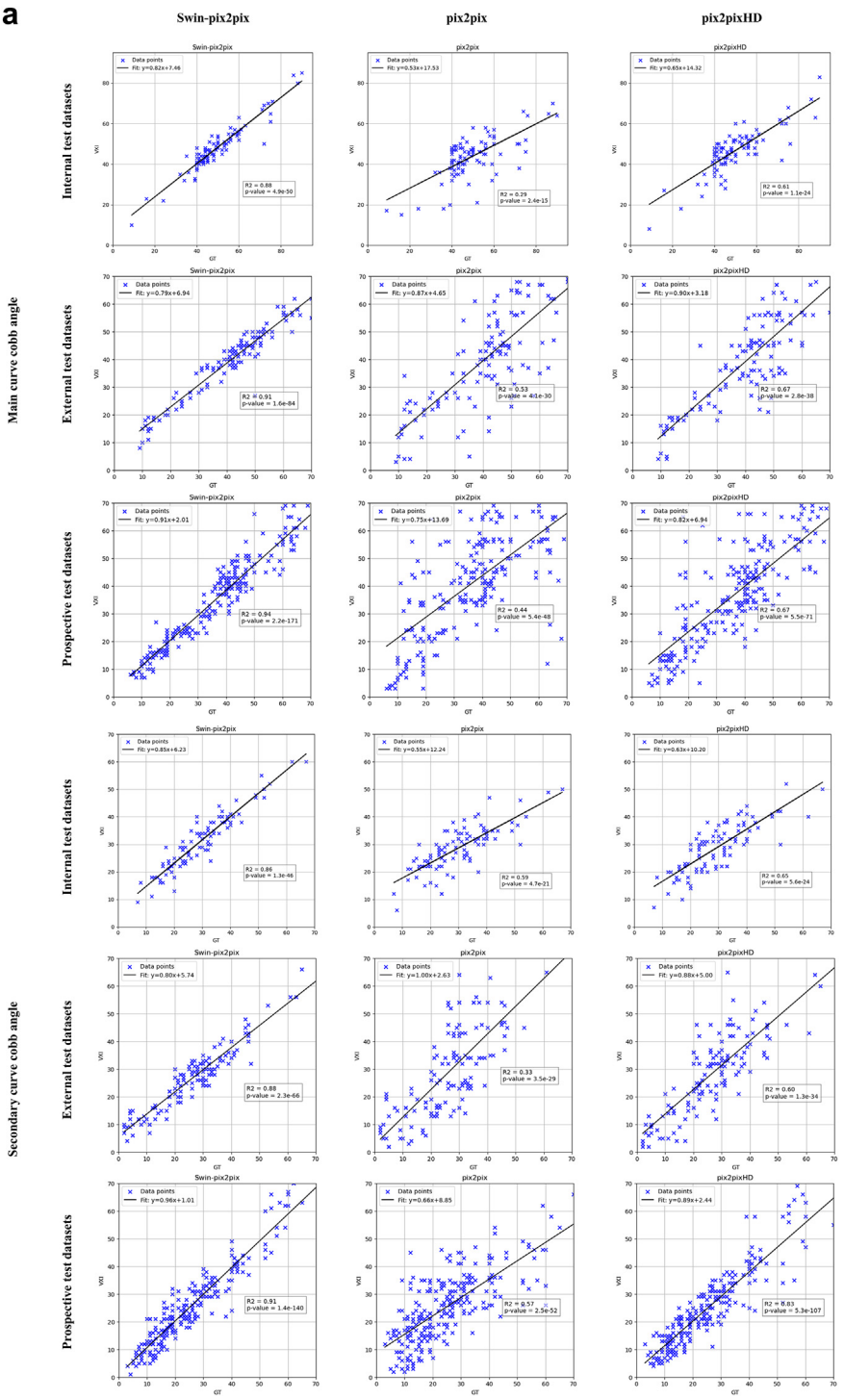

Fig. 7: **Linear regression analysis and bland–altman analysis of cobb angle.** a. Linear Regression Analysis of Cobb Angle: The first to third columns in the figure respectively represent the assessment results of Swin-pix2pix, pix2pix, and pix2pixHD. The top three rows show the results for the main curve, and the bottom three rows show the results for the secondary curve, each containing the results for the internal test set, external test set, and prospective test set. In each subplot, the x-axis corresponds to the Cobb angle measurements from the GT, and the y-axis to those from the VXI. The blue dots denote individual data points for each participant, with the black line indicating the regression line. b. Bland-Altman Analysis of Cobb Angle: The first to third columns in the figure respectively represent the results from Swin-pix2pix, pix2pix, and pix2pixHD. The top three rows show the results for the main curve, and the bottom three rows show the results for the secondary curve, each containing the results for the internal test set, external test set, and prospective test set. Each subplot's x-axis shows the average Cobb angle

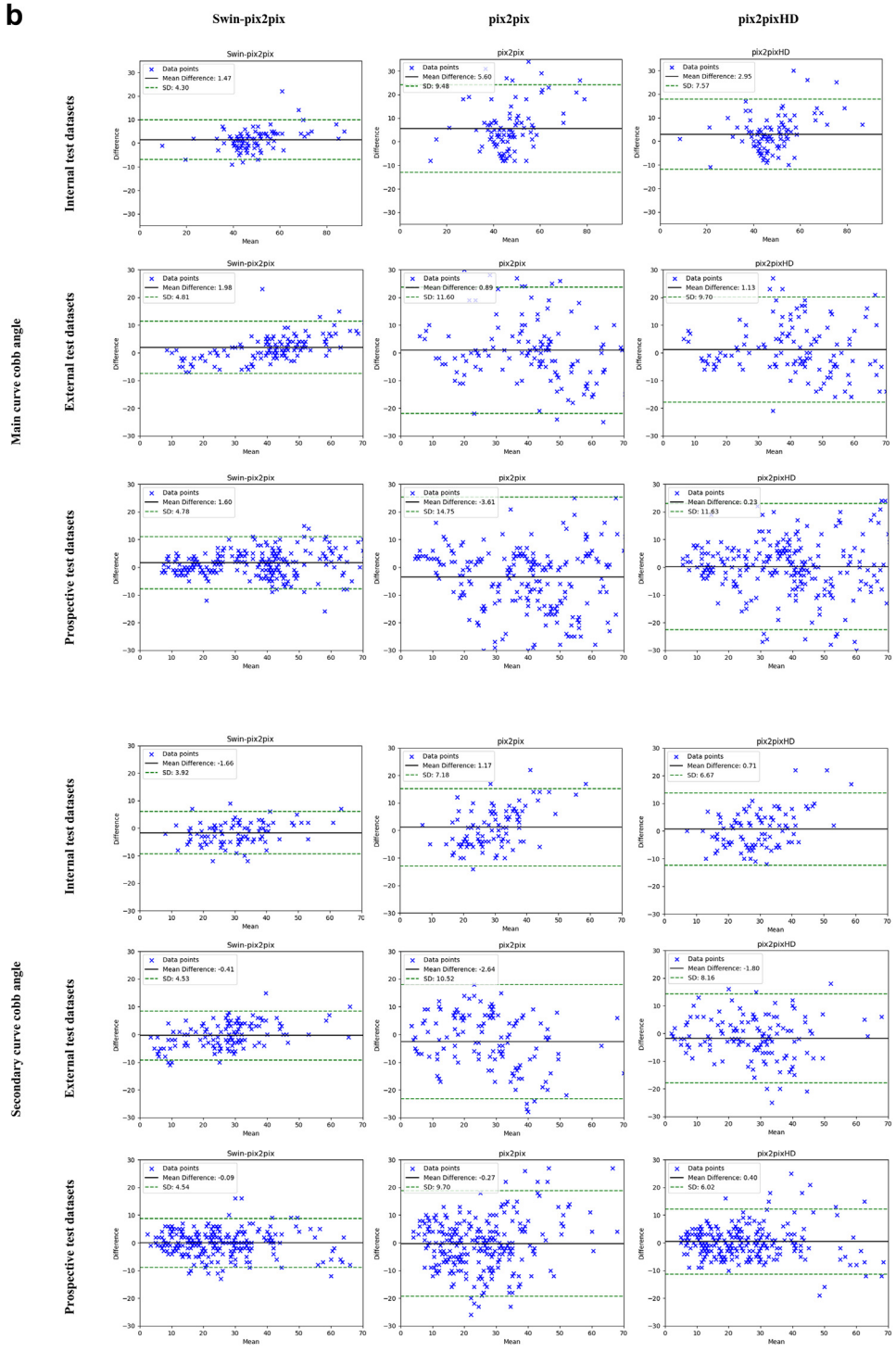

Fig. 7: (continued)

measurements between GT and VXI, while the y-axis shows the difference between the GT and VXI measurements. The figures illustrate that in all three datasets, for both the main and secondary curves, the VXI produced by Swin-pix2pix shows the best fitting and consistency with the GT Cobb angle measurements. Abbreviations: GT: Ground Truth X-ray Image; VXI: Virtual X-ray Image.

Fig. 8: **Confusion matrices for scoliosis severity grading.** The first to third columns in the figure represent the evaluation results for Swin-pix2pix, pix2pix, and pix2pixHD, respectively, while the fourth and fifth columns correspond to the evaluation results of two spine surgery experts. The top three rows show the assessment results for the main curve, and the bottom three rows show the results for the secondary curve, each containing the results for the internal test set, external test set, and prospective test set. In each subplot, the x-axis represents the predicted labels, and the y-axis represents the true labels. Scoliosis severity grading categorizes the Cobb angle into five levels: Level 1: 0–19°; Level 2: 20–39°; Level 3: 40–59°; Level 4: 60–79°; Level 5: ≥80°.

preprocessing module and an X-ray image generation module. The image preprocessing module is capable of removing background noise and optimizing image quality. The X-ray generation module can efficiently generate corresponding VXI based on input back segmentation images. Our system demonstrated good generalizability and robustness across different datasets, indicating its applicability to various real-world scenarios.

Given that scoliosis has become the third major health threat to adolescents after obesity and myopia, the Chinese Ministry of Education announced in 2021 the inclusion of spinal examinations in the routine physical examinations for primary and secondary schools. Following this, the National Health Commission released the "Technical Guidelines for the Prevention and Control of Abnormal Spinal Curvature in Children and Adolescents" to guide the screening of scoliosis. The cumulative radiation exposure caused by traditional radiographic examination methods have been significantly correlated with an increased risk of cancer, especially in children and adolescents, who are

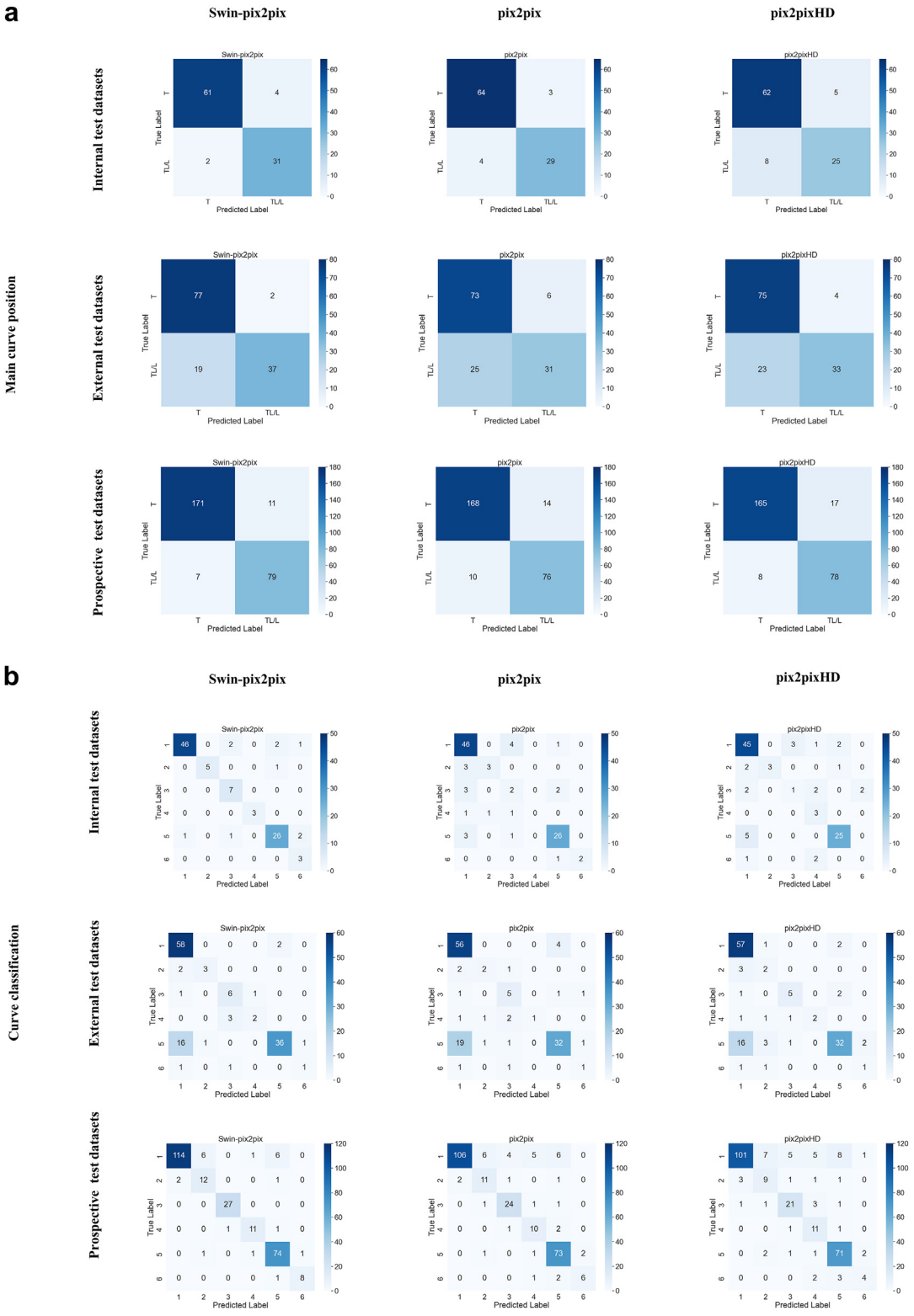

***Fig. 9*: Confusion matrices for main curve position and curve classification.** a. Displays the results for predicting the main curve position in the three datasets by Swin-pix2pix and the two baseline models. b. Shows the results for curve classification in the three datasets by Swin-pix2pix and the two baseline models. In both a and b, the first to third columns represent the evaluation results of Swin-pix2pix, pix2pix,

more sensitive to radiation exposure due to higher cellular metabolic activity.[25–27] Thus, current screening methods still primarily rely on surface appearance screening. However, the lack of specialized spinal surgeons in many regions limits the widespread adoption of scoliosis screening.

Previous studies applying deep learning methods to evaluate AIS on back appearance 2D-RGB images have shown promising results. Yang et al.[28] developed a deep learning-based AIS classification system that uses Faster-RCNN to automatically locate regions of interest and then applies a Resnet model to identify and classify image features. Patients were divided into four groups based on the size of the Cobb angle for scoliosis, and the system was validated on three tasks: 1) binary classification of scoliosis (≥10°) vs. non-scoliosis (0–9°); 2) binary classification of brace treatment (20°–44°) vs. surgical treatment (≥45°); and 3) multi-class classification among the four groups. The system achieved an 80.0% accuracy rate for task 3 on an internal dataset. The system also outperformed human experts in external dataset comparisons, though accuracy remained lower for multi-class tasks, and the system only output scoliosis severity labels without aiding in the visual assessment of curvature morphology. Zhang et al.[29] further conducted research and developed a multi-layer convolutional neural network featuring an attention mechanism and a multi-task strategy. The researchers categorized patients based on the size of the scoliosis Cobb angle into three groups: ① no or mild (≤20°); ② moderate (>20°, ≤40°); ③ severe (>40°); and according to curve type into: single T, single TL/L, and mixed curve; and based on whether the scoliosis Cobb angle progression was >5° within six months into: progressive and nonprogressive. In the prospective test set, the model's accuracy for predicting no or mild and severe cases was 73.26% and 79.84%, respectively. The accuracy for predicting the three types of curves ranged between 72.51% and 74.07%, and the accuracy for predicting scoliosis progression was 70.49%, indicating a further improvement in the model's performance. The two aforementioned studies use classification models to evaluate the severity of scoliosis, which have great application value in large-scale screenings. However, the evaluation of spine deformity should not only focus on the Cobb angle. Shoulder balance, pelvic balance, and coronal balance are also important indicators that affect scoliosis progression and treatment outcomes[30–32] (these anatomical points are also the labels selected in our study for the registration of photos and X-rays). The generated X-rays provide doctors with richer disease information to enhance the comprehensiveness of AIS assessments. Zhang et al.[18] continued by collecting RGBD images of patients with AIS, predicting anatomical landmarks with HRNet, and generating RCI through CycleGAN, demonstrating strong correlations between RCI and GT Cobb angles. This groundbreaking study was the first to generate virtual X-rays, significantly improving the accuracy and reliability of AIS diagnosis. However, the study did not compare its results with other cGAN models or expert evaluations and lacked sufficient external validation. Additionally, the high cost of the depth camera equipment and the associated learning curve for its use limited its widespread adoption.

Our study aims to generate X-rays equivalent to those produced by depth cameras using a 2D camera, enabling almost everyone to obtain an accurate AIS assessment for free. This approach offers more options for economically underdeveloped countries and regions and has the potential to become the mainstream method for AIS screening and follow-up in the future. However, we found that the CycleGAN model, which performs excellently on RGBD images, did not perform well in our 2D-RGB scenario. The model attempts to learn the mapping between 2D-RGB images and X-ray images, but its performance is limited due to the absence of one-to-one correspondence. It is believed that the challenge in evaluating AIS on back appearance 2D-RGB images lies in the fact that the Cobb angle of spinal curvature is often less apparent on RGB images than on X-rays due to the obstruction of back soft tissues, creating a complex nonlinear mapping relationship between 2D-RGB images and X-rays. This poses a significant challenge to the model's architecture design.

Classic Generative Adversarial Networks (GANs) consist of a Generator and a Discriminator, where through their mutual competition, the Generator eventually produces high-quality data.[33,34] Conditional GANs (cGANs), an extension of GANs, incorporate additional conditional information, enabling the generated data to be not only visually realistic but also to meet specific conditions or attributes.[35] Pix2pix was the first framework to successfully apply cGAN to a variety of image-to-image translation tasks, with many subsequent studies like cycleGAN and pix2pixHD improving upon it.[19] Pix2pix is widely applicable, learning the conditional information provided by paired data to accurately capture the mapping between input and output in an end-to-end manner, thus

and pix2pixHD, respectively, while the first to third rows represent the internal test set, external test set, and prospective test set. In each subplot, the x-axis represents the predicted labels, and the y-axis represents the true labels. In the main curve position, T indicates a thoracic curve; TL/L denotes a thoracolumbar/lumbar curve. Curve classification categorizes the morphology of scoliosis into six levels, as detailed in the methods section.

Evaluation metrics	Model	Test datasets	Main curve position			Curve classification						
			T	TL/L	Total	1	2	3	4	5	6	Total
Accuracy	Swin-pix2pix	Internal	0.92			0.94	0.99	0.97	1.00	0.93	0.97	0.97
		External	0.84			0.84	0.98	0.96	0.97	0.85	0.98	0.93
		Prospective	0.93			0.94	0.96	0.99	0.99	0.96	0.99	0.97
	pix2pix	Internal	0.93			0.85	0.96	0.89	0.97	0.92	0.99	0.93
		External	0.77			0.79	0.96	0.94	0.97	0.80	0.97	0.91
		Prospective	0.91			0.91	0.96	0.97	0.96	0.94	0.98	0.95
	pix2pixHD	Internal	0.87			0.84	0.97	0.91	0.95	0.92	0.95	0.92
		External	0.80			0.81	0.93	0.96	0.98	0.81	0.97	0.91
		Prospective	0.91			0.89	0.94	0.95	0.95	0.93	0.97	0.93
Sensitivity	Swin-pix2pix	Internal	0.91	0.94	0.92	0.90	0.83	1.00	1.00	0.87	1.0	0.93
		External	0.97	0.66	0.82	0.97	0.60	0.75	0.40	0.67	0.33	0.62
		Prospective	0.94	0.92	0.93	0.96	0.80	1.00	0.85	0.96	0.89	0.90
	pix2pix	Internal	0.95	0.88	0.92	0.90	0.50	0.29	0.00	0.87	0.67	0.54
		External	0.92	0.55	0.74	0.93	0.40	0.63	0.20	0.59	0.33	0.51
		Prospective	0.91	0.88	0.90	0.83	0.73	0.89	0.77	0.95	0.67	0.81
	pix2pixHD	Internal	0.92	0.76	0.84	0.88	0.50	0.14	1.00	0.83	0.00	0.56
		External	0.95	0.59	0.77	0.95	0.40	0.63	0.4	0.59	0.33	0.55
		Prospective	0.91	0.91	0.91	0.80	0.60	0.78	0.85	0.92	0.44	0.73
Specificity	Swin-pix2pix	Internal	0.94	0.91	0.92	0.98	1.00	0.97	1.00	0.96	0.97	0.98
		External	0.66	0.97	0.74	0.73	0.99	0.97	0.99	0.98	0.99	0.94
		Prospective	0.92	0.94	0.93	0.99	0.97	0.99	0.99	0.95	0.99	0.98
	pix2pix	Internal	0.88	0.95	0.92	0.80	0.99	0.93	1.00	0.94	1.00	0.94
		External	0.55	0.92	0.82	0.68	0.98	0.96	1.00	0.94	0.98	0.92
		Prospective	0.88	0.91	0.90	0.99	0.97	0.98	0.97	0.94	0.99	0.97
	pix2pixHD	Internal	0.76	0.92	0.84	0.80	1.00	0.97	0.95	0.96	0.98	0.94
		External	0.59	0.95	0.77	0.71	0.95	0.98	1.0	0.95	0.98	0.93
		Prospective	0.91	0.91	0.91	0.97	0.96	0.97	0.95	0.93	0.99	0.96
PPV	Swin-pix2pix	Internal	0.97	0.83	0.90	0.98	1.00	0.70	1.00	0.90	0.50	0.85
		External	0.80	0.95	0.88	0.74	0.75	0.60	0.67	0.95	0.50	0.70
		Prospective	0.96	0.88	0.92	0.98	0.63	0.96	0.85	0.89	0.89	0.87
	pix2pix	Internal	0.94	0.91	0.92	0.82	0.75	0.25	0.00	0.87	1.00	0.61
		External	0.74	0.84	0.79	0.70	0.50	0.50	1.00	0.86	0.33	0.65
		Prospective	0.94	0.84	0.90	0.98	0.58	0.80	0.55	0.86	0.75	0.75
	pix2pixHD	Internal	0.88	0.83	0.86	0.81	1.00	0.25	0.38	0.89	0.00	0.56
		External	0.77	0.89	0.83	0.72	0.25	0.71	1.0	0.89	0.33	0.65
		Prospective	0.95	0.82	0.89	0.96	0.47	0.72	0.48	0.84	0.57	0.67
NPV	Swin-pix2pix	Internal	0.84	0.97	0.90	0.91	0.99	1.00	1.00	0.94	1.00	0.97
		External	0.95	0.80	0.88	0.96	0.98	0.98	0.98	0.81	0.98	0.95
		Prospective	0.88	0.96	0.92	0.91	0.99	0.99	0.99	0.98	0.99	0.98
	pix2pix	Internal	0.91	0.94	0.92	0.89	0.97	0.95	0.97	0.94	0.99	0.95
		External	0.84	0.74	0.79	0.93	0.98	0.98	0.97	0.78	0.98	0.94
		Prospective	0.84	0.94	0.90	0.87	0.98	0.99	0.99	0.98	0.99	0.97
	pix2pixHD	Internal	0.83	0.88	0.86	0.87	0.97	0.94	1.00	0.93	0.97	0.95
		External	0.89	0.77	0.83	0.95	0.98	0.98	0.98	0.78	0.98	0.93
		Prospective	0.82	0.95	0.89	0.84	0.98	0.97	0.99	0.97	0.98	0.96

Table 3: **Quantitative performance on main curve position and curve classification.**

generating realistic, relevant images. Although later models like CycleGAN have made significant progress in image-to-image translation tasks with unpaired data, experimental results sometimes show a difficult-to-overcome gap compared to training results based on paired data.[24] The performance of the cGAN generator directly affects the quality and accuracy of the generated images. Pix2pix's generator uses the UNet-256 segmentation model, which can generally learn and understand the structural information in images well. Haiderbhai et al.[36] developed pix2xray based on the pix2pix architecture, successfully synthesizing X-ray images from 2D-RGB images of hand gestures. Lu et al.[37] were the first to integrate the advantages of Swin

Transformer into both the encoder and decoder of the standard UNet, enhancing the representation of semantic features.[37,38] Inspired by this, we integrated Swin blocks into the encoder of pix2pix's generator, not only improving 1) the ability to capture details and contextual information in images; 2) the ability to capture features at different scales; 3) the ability to effectively process long-distance pixel dependencies but also helping to lighten the network structure and increase computational efficiency. Through the visualized feature maps, it can be observed that compared to pix2pix, Swin-pix2pix demonstrates better performance in extracting features of the spine and trunk edges (eAppendix 7, Supplementary eFig. S13).

To our knowledge, in this study, we built a paired dataset of routine back 2D-RGB images and standing full-spine X-rays of patients with AIS, and for the first time developed an innovative Swin-pix2pix network structure that integrates the self-attention mechanism, successfully training a VXI generation system. The greatest advantage of this system is its ability to accurately generate VXI using routine back 2D-RGB images, overcoming traditional screening's limitations in personnel and location. Since no special equipment is required, screening can be conducted in medical institutions, schools, communities, and homes, significantly increasing the prevalence of scoliosis screening. Based on the excellent performance of our Swin-pix2pix system, we developed an end-to-end workstation that allows uploading back 2D-RGB images using personal computers or smartphones and obtaining generated VXI within seconds (eAppendix 4, Supplementary eFigs. S4 and S5). We recorded a video showcasing the system development environment and its real-world application: a doctor with no prior training in using the system takes a photograph of the patient's back. Without any complex operations, the photo is uploaded to the website, and a generated VXI image is obtained. The collected photo and VXI image are then uploaded to our developed database for further evaluation, details of which can be found in the Supplementary materials.

From our observations, regardless of demographic differences, we believe the quality of 2D-RGB images, standing posture, and lighting environment significantly impact VXI generation. Specifically, these factors directly affect the YOLOv8 model's segmentation of the torso outline during preprocessing and the Swin-pix2pix's ability to capture back texture features during generation. Increasing the internal structural similarity between input and target images, thus generating more accurate spinal curvature. In our study, our training set comprises retrospective 2D-RGB photo data without strict environmental control, which poses both challenges and opportunities. These photos cover a variety of everyday scenes and commonly used photography equipment, allowing the model to adapt to the heterogeneity of multi-centre data and enhancing its robustness. Furthermore, rich 3D information can be embedded even in 2D photos including depth information and rotational-related information, which helps minimize rotational errors and optimizes the representation of scoliosis forms. However, we must acknowledge that data heterogeneity does affect the model's performance. We found that the results from PUMCH closely matched our expectations, less than the systemic error produced by manual Cobb angle measurement,[39] due to good lighting, appropriate shooting angles, and better consistency with the training dataset. On the other hand, the dataset from SAHZU performed poorly because the photos were taken in poor lighting conditions and from a greater distance, which significantly differed from the scenarios in the training set. This presents a higher demand for the model's generalization ability. The fact that the model performed better on the prospective test dataset than on all other test datasets further confirms the above findings. We are exploring ways to improve the model to enhance the quality and accuracy of image generation for such data.

Overall, our Swin-pix2pix model demonstrated universality and robustness in comprehensive evaluations, achieving satisfactory synthetic results, which can be attributed to the use of paired datasets and the model's lightweight and refined architectural innovations. These innovations allow the model to perform exceptionally well in processing more complex scenarios. Our experimental results also prove that the generated VXI is clinically relevant for predicting the severity and type of spinal curvature in patients with AIS. Thus, our study empirically demonstrated the significant advantages of applying the Transformer model to pix2pix for the first time. The innovative applications of these models not only enhanced the practicality of the outcomes but also achieved diagnostic levels comparable to professional medical practitioners. Notably, the recent success of the model Sora in the field of video further confirms the effectiveness of applying Transformer modules in generative networks. The foundation of the Sora model is the Diffusion Transformer (DiT),[40] an advanced algorithm that integrates diffusion models with Transformer architecture. DiT effectively preserves the spatial information and coherence of images, enabling the generative model to capture both local details and overall structure more accurately, thereby significantly enhancing the model's overall performance.

Despite the encouraging results of this study, there are some limitations. First, Our data volume and quality still need improvement. In the future, we will establish strict data collection standards and continue increasing the data volume of normal adolescents, mild scoliosis, and severe scoliosis cases. Second, all participants in this study were Han Chinese from different regions, and the system trained on this dataset may not accurately apply to other ethnic groups. Lastly, despite strict quality control measures, we cannot completely

eliminate the error caused by trunk rotation during imaging, which may potentially affect the experimental results to some extent.

In summary, we have developed the first X-ray image generation system based on back 2D-RGB images using deep learning technology. This system can efficiently and radiation-free assess the severity and type of scoliosis in patients with AIS, performing superior to all known cGAN models and spinal surgery experts. We believe this system has the potential for widespread screening and assessment of scoliosis in the future, significantly reducing the rates of missed diagnoses and misdiagnoses.

Contributors

Z.Z.Z., W.J.L. initiated the project and the collaboration. Z.H., N.L., E.C.C. developed the network architectures, training, and testing setup. Z.L., X.D.Q., J.L., Y.Q. designed the clinical setup. Y.C., S.R.W., J.L.Y., Z.W.W., X.P.C., A.Y.L. created the data set and defined clinical labels. Y.M.W., W.J.L. contributed to the software engineering. J.C.C., K.G.Y. contributed clinical expertise. W.Y.L., X.L.L. contributed algorithmic expertise. Z.H., N.L., Z.Z.Z., W.J.L. wrote the paper. Z.H., N.L., Y.C. have access to and verify the underlying study data.

Data sharing statement

After publication, partial data supporting the research results can be made available for academic purposes upon written request to the corresponding author. The source code and trained models of Swin-pix2pix can be requested from the corresponding author for non-commercial purposes after patent approval (Patent Application No.: 202311748636.X). Images and sensitive patient privacy information will not be disclosed. Additionally, access to this data will require signing a data transfer or access agreement, subject to case-by-case review by Nanjing Drum Tower Hospital.

Declaration of interests

All authors declare there is no conflict of interest.

Acknowledgements

This work was supported by the National Key R&D Program of China (2023YFC2507700), the Natural Science Foundation of Jiangsu Province (BK20230147), the China Postdoctoral Science Foundation (2022M711581), the Nanjing Medical Science and Technology Development Foundation (YKK22098), the Jiangsu Provincial Key Research and Development Program (BE2023658), and the Jiangsu Provincial Medical Innovation Centre of Orthopedic Surgery (CXZX202214). The manuscript submitted does not contain information about medical device(s)/drug(s). No relevant financial activities outside the submitted work.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2024.102779.

References

1 Weinstein SL, Dolan LA, Cheng JC, Danielsson A, Morcuende JAJTl. Adolescent idiopathic scoliosis. *Lancet*. 2008;371(9623):1527–1537.
2 Cheng JC, Castelein RM, Chu WC, et al. Adolescent idiopathic scoliosis. *Nat Rev Dis Primers*. 2015;1(1):1–21.
3 Kuznia AL, Hernandez AK, Lee LU. Adolescent idiopathic scoliosis: common questions and answers. *Am Fam Physician*. 2020;101(1):19–23.
4 Choudhry MN, Ahmad Z, Verma R. Adolescent idiopathic scoliosis. *Open Orthop J*. 2016;10:143.
5 Weinstein SL, Dolan LA, Spratt KF, Peterson KK, Spoonamore MJ, Ponseti IVJJ. Health and function of patients with untreated idiopathic scoliosis: a 50-year natural history study. *JAMA*. 2003;289(5):559–567.
6 Weinstein SL, Dolan LA, Wright JG, Dobbs MB. Effects of bracing in adolescents with idiopathic scoliosis. *N Engl J Med*. 2013;369(16):1512–1521.
7 Weinstein S, Ponseti IJ. Curve progression in idiopathic scoliosis. *J Bone Joint Surg Am*. 1983;65(4):447–455.
8 Sanders JO, Newton PO, Browne RH, Katz DE, Birch JG, Herring JA. Bracing for idiopathic scoliosis: how many patients require treatment to prevent one surgery? *J Bone Joint Surg Am*. 2014;96(8):649–653.
9 Chung N, Cheng Y-H, Po H-L, et al. Spinal phantom comparability study of Cobb angle measurement of scoliosis using digital radiographic imaging. *J Orthop Translat*. 2018;15:81–90.
10 Côté P, Kreitz BG, Cassidy JD, Dzus AK, Martel J. A study of the diagnostic accuracy and reliability of the Scoliometer and Adam's forward bend test. *Spine (Phila Pa 1976)*. 1998;23(7):796–802.
11 Fong DYT, Lee CF, Cheung KMC, et al. A meta-analysis of the clinical effectiveness of school scoliosis screening. *Spine (Phila Pa 1976)*. 2010;35(10):1061–1071.
12 Dunn J, Henrikson NB, Morrison CC, Blasi PR, Nguyen M, Lin JSJJ. Screening for adolescent idiopathic scoliosis: evidence report and systematic review for the US preventive services task force. *JAMA*. 2018;319(2):173–187.
13 Development of 3-D ultrasound system for assessment of adolescent idiopathic scoliosis (AIS): and system validation. In: Cheung C-WJ, Law S-Y, Zheng Y-P, eds. *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE; 2013.
14 Hong Y, Hwang U, Yoo J, Yoon SJACS. How generative adversarial networks and their variants work: An overview. *ACM Comput Surv*. 2019;52(1):1–43.
15 Chen Y, Lin Y, Xu X, et al. Multi-domain medical image translation generation for lung image classification based on generative adversarial networks. *Comput Methods Programs Biomed*. 2023;229:107200.
16 Sun H, Wang F, Yang Y, et al. Transfer learning-based attenuation correction for static and dynamic cardiac PET using a generative adversarial network. *Eur J Nucl Med Mol Imaging*. 2023;50(12):3630–3646.
17 Zhao H, Li H, Maurer-Stroh S, Cheng L. Synthesizing retinal and neuronal images with generative adversarial nets. *Med Image Anal*. 2018;49:14–26.
18 Meng N, Wong K-YK, Zhao M, Cheung JP, Zhang TJE. Radiograph-comparable image synthesis for spine alignment analysis using deep learning with prospective clinical validation. *EClinicalMedicine*. 2023;61:102050.
19 Image-to-image translation with conditional adversarial networks. In: Isola P, Zhu J-Y, Zhou T, Efros AA, eds. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
20 Swin transformer: hierarchical vision transformer using shifted windows. In: Liu Z, Lin Y, Cao Y, et al., eds. *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.
21 Swinir: image restoration using swin transformer. In: Liang J, Cao J, Sun G, Zhang K, Van Gool L, Timofte R, eds. *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.
22 Swin transformer v2: scaling up capacity and resolution. In: Liu Z, Hu H, Lin Y, et al., eds. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.
23 High-resolution image synthesis and semantic manipulation with conditional gans. In: Wang T-C, Liu M-Y, Zhu J-Y, Tao A, Kautz J, Catanzaro B, eds. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
24 Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Zhu J-Y, Park T, Isola P, Efros AA, eds. *Proceedings of the IEEE international conference on computer vision*. 2017.
25 Kwan ML, Miglioretti DL, Bowles EJ, et al. Quantifying cancer risk from exposures to medical imaging in the Risk of Pediatric and Adolescent Cancer Associated with Medical Imaging (RIC) Study: research methods and cohort profile. *Cancer Causes Control*. 2022;33(5):711–726.
26 Thierry-Chef I, Harbron R, Hauptmann M, et al. *Risk of hematological malignancies from CT radiation exposure in children, adolescents and young adults*. 2023.
27 Rodemann HP, Blaese MA, eds. Responses of normal cells to ionizing radiation. *Semin Radiat Oncol*. 2007;17(2):81–88. Elsevier.
28 Yang J, Zhang K, Fan H, et al. Development and validation of deep learning algorithms for scoliosis screening using back images. *Commun Biol*. 2019;2(1):390.

29 Zhang T, Zhu C, Zhao Y, et al. Deep learning model to classify and monitor idiopathic scoliosis in adolescents using a single smartphone photograph. *JAMA Netw Open*. 2023;6(8):e2330617.
30 Lang C, Huang Z, Zou Q, Sui W, Deng Y, Yang JJTSJ. Coronal deformity angular ratio may serve as a valuable parameter to predict in-brace correction in patients with adolescent idiopathic scoliosis. *Spine J*. 2019;19(6):1041–1047.
31 Yang Y, Yang M, Zhao J, Zhao Y, Yang C, Li MJESJ. Postoperative shoulder imbalance in adolescent idiopathic scoliosis: risk factors and predictive index. *Eur Spine J*. 2019;28:1331–1341.
32 Sato T, Yonezawa I, Matsumoto H, et al. Surgical Predictors for prevention of postoperative shoulder imbalance in Lenke Type 2A adolescent idiopathic scoliosis. *Spine (Phila Pa 1976)*. 2022;47 (4):E132–E141.
33 Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems*. vol. 27. 2014.
34 Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Adv Neural Inform Process Syst*. 2020;63(11):139–144.
35 Mirza M, Osindero S. *Conditional generative adversarial nets*. 2014.
36 Haiderbhai M, Ledesma S, Lee SC, et al. pix2xray: converting RGB images into X-rays using generative adversarial networks. *Int J Comput Assist Radiol Surg*. 2020;15:973–980.
37 Lin A, Chen B, Xu J, et al. Ds-transunet: dual swin transformer u-net for medical image segmentation. *IEEE Trans Instr Meas*. 2022;71:1–15.
38 Vaswani A, Shazeer N, Parmar N, et al. *Attention is all you need*. 2017;vol. 30.
39 Shea KG, Stevens PM, Nelson M, Smith JT, Masters KS, Yandow S. A comparison of manual versus computer-assisted radiographic measurement: intraobserver measurement variability for Cobb angles. *Spine (Phila Pa 1976)*. 1998;23(5):551–555.
40 Scalable diffusion models with transformers. In: Peebles W, Xie S, eds. *Proceedings of the IEEE/CVF international conference on computer vision*. 2023.