EClinicalMedicine 42 (2021) 101220

Contents lists available at ScienceDirect

EClinicalMedicine

journal homepage: https://www.journals.elsevier.com/eclinicalmedicine

Original Research

Application of deep learning upon spinal radiographs to predict progression in adolescent idiopathic scoliosis at first clinic visit

Hongfei Wang[1], Teng Zhang[1], Kenneth Man-Chee Cheung[1], Graham Ka-Hon Shea[1,*]

[1] *Department of Orthopaedics and Traumatology, The University of Hong Kong, Hong Kong*

ARTICLE INFO

Article History:
Received 27 August 2021
Revised 6 November 2021
Accepted 15 November 2021
Available online xxx

Keywords:
Adolescent idiopathic scoliosis
curve progression
radiomics
deep learning
scoliosis screening

ABSTRACT

Background: Prediction of curve progression risk in adolescent idiopathic scoliosis (AIS) remains elusive. Prior studies have revealed the potential for three-dimensional (3D) morphological parameters to prognosticate progression, but these require specialized biplanar imaging equipment and labor-intensive software reconstruction. This study aimed to formulate a deep learning model with standing posteroanterior (PA) X-rays at first clinic visit to differentiate between progressive (P) and non-progressive (NP) curves.

Methods: For this retrospective cohort study, we identified patients presenting with AIS between October 2015 to April 2020 at our tertiary referral centre. Patients with mild curvatures (11 – 30°) who were skeletally immature (Risser sign of ≤2) were recruited. Patients receiving biplanar X-ray radiographs (EOS™) were divided between a training-cross-validation cohort (328 patients) and independent testing cohort (110 patients). Another 52 patients receiving standard PA spinal X-rays were recruited for cross-platform validation. Following 3D reconstruction, we designated the major curve apex upon PA X-rays as the region of interest (ROI) for machine learning. A self-attentive capsule network was constructed to differentiate between curves manifesting P and NP trajectories. A two-stage transfer learning strategy was introduced to pre-train and fine-tune the model. Model performance (accuracy, sensitivity, specificity) was compared to that of traditional convolutional neural networks (CNNs) and a clinical parameter-based logistic regression model.

Findings: 3D reconstruction identified that apical rotation of the major curve and torsion were significantly different between P and NP curve trajectories. Our predictive model utilizing an ROI centered on the major curve apex achieved an accuracy of 76.6%, a sensitivity of 75.2% and a specificity of 80.2% upon independent testing. Cross-platform performance upon standard standing PA X-rays yielded an accuracy of 77.1%, a sensitivity of 73.5% and a specificity of 81.0%. Errors in prediction occurred when the degree of apical rotation / torsion was discrepant from that of the subsequent curve trajectory but could be rectified by considering serial X-rays. Performance was superior to that of traditional CNNs as well as clinical parameter-based regression models.

Interpretation: This is the first report of automated prediction of AIS curve progression based on radiomics and deep learning, towards directing treatment strategy at first visit. Patients predicted to be at-risk of progression may be counselled to receive early bracing with enforcement of treatment compliance. Over-treatment may be avoided in curves deemed to be non-progressive. Results need to be consolidated in larger sample populations of different ethnicities.

Funding: The Society for the Relief of Disabled Children (SRDC).

1. Introduction

Adolescent idiopathic scoliosis (AIS) is a three-dimensional spinal deformity affecting up to 3% of the population [1]. AIS is diagnosed when Cobb angles (a measure of coronal plane deformity) exceed 10°, and progressive deterioration during puberty occurs in two-thirds of patients [2]. Delaying treatment of scoliosis can lead to increased back pain, significant cosmetic deformity, as well as compromised pulmonary function [3]. A prognostic uncertainty remains in distinguishing between progressive (P) and non-progressive (NP) curve trajectories when skeletally immature patients first present [4].

Currently, patients with a curve ≥ 25° and significant remaining growth potential as reflected by change in body height, skeletal age (Risser stage, DRU grading), and timing of menarche [5] are referred

* Correspondence to: Dr Graham Ka-Hon Shea, MBBS, PhD, FRCSEd (Orthopaedics and Traumatology), Department of Orthopaedics and Traumatology, The University of Hong Kong, Hong Kong
E-mail addresses: tgzhang@hku.hk (T. Zhang), gkshea@hku.hk (G.K.-H. Shea).

https://doi.org/10.1016/j.eclinm.2021.101220

Research in context

Evidence before this study

We searched PubMed on July 15, 2021, for articles that described the application of deep learning algorithms to predict curve progression from X-rays in adolescent idiopathic scoliosis, using the search terms "deep learning" OR "machine learning" AND "adolescent idiopathic scoliosis curve progression" AND "X-rays", with no language or date restrictions. Previous studies were limited to the use of 3D spinal morphological parameters, demographic parameters, and molecular biomarkers. Deep learning-based radiomics had yet to be incorporated for the automated prediction of curve progression.

Added value of this study

This study combines recent advances in machine learning with our understanding on curve morphology in adolescent idiopathic scoliosis (AIS). We reveal that posteroanterior (PA) X-rays of AIS patients carry discriminatory features around the major curve apex which facilitate identification of progressive curves. We developed a deep learning model with the capacity to predict progressive curves from PA x-rays taken at first clinic visit in patients with AIS.

Implications of all the available evidence

Our platform provides an automated means to predict for curve progression upon diagnosis of AIS in a cost-effective and non-invasive manner that does not require specialized imaging equipment. A radiomics-based prediction model holds immense value for AIS screening programmes by facilitating 1) early referral to specialist centers, and 2) pre-emptive bracing of at-risk curves. This is a necessary advancement beyond the current practice to observe mild curves in skeletally immature patients.

for bracing. A management dilemma remains in skeletally immature patients with a curve < 25 °. As their subsequent curve trajectories remain unknown, they are observed with X-rays repeated at next follow-up for Cobb angle measurements. Bracing remains reactionary to documented curve progression beyond a defined threshold indicated for brace prescription. Therefore, it is of significant clinical importance to be able to prognosticate curve trajectories when patients first present to the specialist clinic for management, often after an initial diagnosis from school-age screening programmes. Promisingly, measurement of three-dimensional (3D) parameters of scoliotic curves have been described to identify curves at risk of progression [6,7]. However, barriers to application remain in the prerequisite for expensive biplanar X-ray systems followed by time-consuming and error-prone software reconstruction of each individual vertebra [8].

Aforementioned findings on curve morphology have yet to be applied upon recent advances in radiomics, which promise automated prediction of curve progression [9]. Pre-labelled datasets (i.e. presenting X-rays of curves with progressive and non-progressive trajectories) may be utilized for training and validation of machine learning models that extract and process imaging features towards achieving correct classification. Performance may then be assessed upon an independent testing dataset. Prediction models that mimic the layered and interconnected architecture of neurons within the human brain are referred to as artificial neural networks. These neural networks have the capacity for deep learning, which is a subset of machine learning allowing for features and patterns to be 'learnt' by software towards continued refinement in order to improve prediction accuracy. The objective of our study was to develop an artificial neural network capable of differentiating between P and NP curve trajectories from spinal radiographs obtained at first clinic visit.

We hypothesized that uniplanar posteroanterior (PA) X-rays, which convey features of three-dimensional spinal morphology of prognostic relevance such as asymmetry of pedicle shadows and spinous processes to reflect rotation [10] as well as rib vertebral angles [11] would suffice for prediction of progressive curves. To this end, capsule neural networks (CapsNet) possess distinct advantages in comparison to traditional convolutional neural networks (CNNs) in detecting spatial relationships among features to account for geometric transformations, possessing a property known as equivariance, [12] which improves upon model performance. By incorporating a non-iterative and highly parallelizable self-attention routing algorithm, Efficient-CapsNet is another recent advancement in model architecture that reduces model complexity whilst achieving similar results [13].

We first validated differences in 3D spinal parameters reported in a Canadian study to be statistically different between progressive and non-progressive curves in our study population [14]. We identified that morphology of the major curve apex demonstrated the strongest correlation with progression risk, which was thereby designated as the region of interest (ROI) for deep learning. We subsequently developed an Efficient-CapsNet-based radiomics method for automated classification of progressive and non-progressive AIS. A curvature magnitude-sensitive transfer learning scheme was applied, with the model pre-trained upon skeletally mature curves and those having progressed to require surgery. Fine-tuning of the pre-trained model was conducted upon X-rays of mild curves at first presentation in skeletally immature individuals, which was the population of interest for prediction.

To the best of our knowledge this is the first attempt at automated prediction of AIS curve progression that is based on radiomics and deep learning, without necessitating software reconstruction of curve morphology. Advantages of CapsNet in overcoming feature invariance is highlighted in the clinical context of scoliosis as a three-dimensional deformity. Towards clinical application, performance upon a proprietary biplanar imaging platform was extensible across standard standing PA X-rays.

2. Methods

2.1. Patient recruitment

This manuscript adhered to STROBE guidelines. We recruited a total of 490 patients with clinic attendance at the Duchess of Kent Children's Hospital between October 2015 to April 2020 for this retrospective cohort study (Figure 1). The inclusion criteria were (1) diagnosis of AIS, (2) Cobb angle between 11° and 30° upon X-rays at first visit, (3) Risser sign of ≤2 to demonstrate growth potential, and (4) regular follow-up concluding at skeletal maturity (Risser sign of 4 or more) or upon receiving surgery . For patients with more than one curve, only the major curve (largest Cobb angle) was analysed. Skeletally immature patients were followed up in 3 – 6 month intervals with standing X-rays repeated at every visit.

Progression (P) was defined by an increase ≥ 6° between first visit and skeletal maturity, as well as a Cobb angle ≥ 25° at skeletal maturity. Non-progression (NP) was defined by < 6° of curvature increase, or in those with a Cobb angle < 25° at skeletal maturity. Patients with non-progressive curves according to these definitions yet having received brace treatment were excluded.

Informed consent was obtained from all participants and their parents / legal guardians before examination and measurements were conducted. Ethical approval from the Institutional Review Board

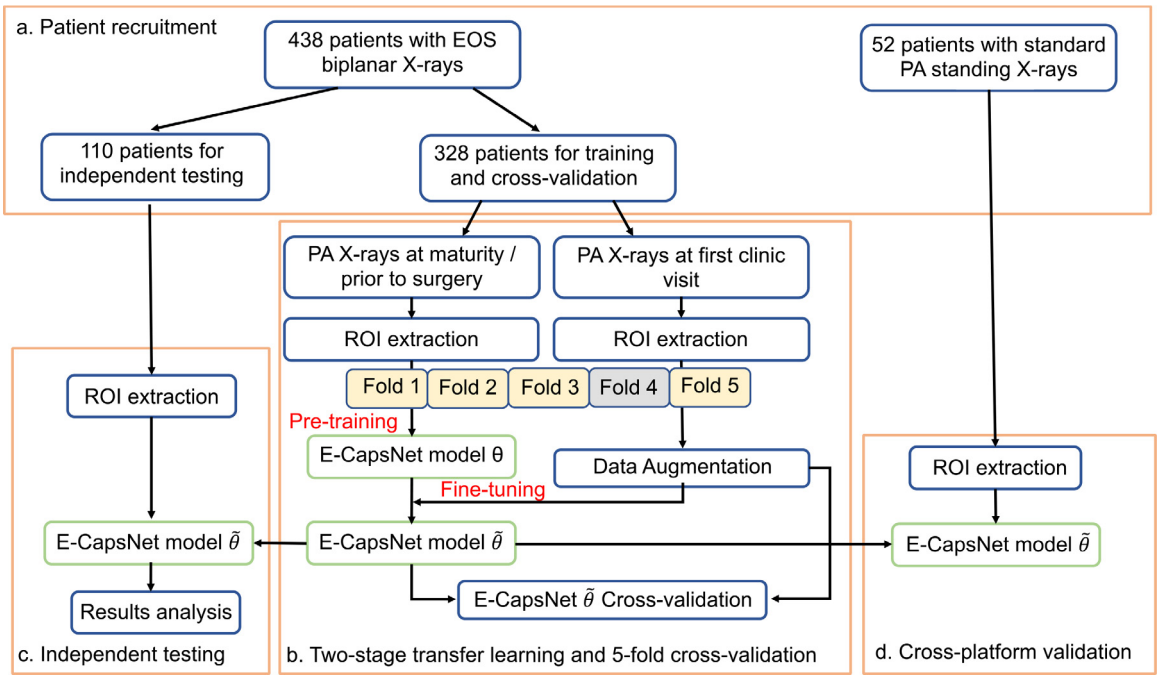


Figure 1. Patient recruitment and workflow (a) Patient recruitment with division into training, validation, and testing datasets. (b) Two-stage transfer learning and cross-validation facilitated hyperparameter searching for model optimisation. (c) Independent testing on the established model with PA films obtained via EOS imaging. (d) Cross-platform validation on standard standing whole spine PA films.

(IRB) of the University of Hong Kong / Hong Kong West Cluster was obtained for this study.

2.2. Reconstruction and morphological parameter analysis

To validate differences in 3D spinal parameters between P and NP curve trajectories upon our local population, [14] we randomly selected 138 patients (65 P, 73 NP) with biplanar EOS X-rays taken upon clinic presentation as an exploration dataset. Reconstruction was conducted by a skilled research assistant using sterEOS software (v1.6). The calculated parameters were divided into three categories - (1) Global descriptors: 3D Cobb angles, thoracic kyphosis (T4–T12) and lumbar lordosis (L1–L5); (2) Regional descriptors: Torsion, which denotes mean of intervertebral axial rotations over the curvature region; (3) Local descriptors: Axial intervertebral rotation of the curve apex, upper, and lower end vertebra.

2.3. Data preprocessing

Amongst our cohort, local descriptors surrounding the major curve apex were highly correlated with curve progression. We therefore selected the apical vertebrae or disc of the major curve, together with at least two adjacent vertebras above and below as well as the lateral rib articulations as the region of interest (ROI). ROIs were extracted within a 150×100-pixel fixed window and saved as single channel grayscale images in JPG formatting (Figure 2).

2.4. Deep capsule network with self-attention routing

Hinton et al. first described the concept of a vectorial organization of neurons to encapsulate both probability and instantiation parameters of a detected feature [15,16]. Recently, Mazzia et al. proposed a highly efficient capsule network (Efficient-CapsNet) with self-attention routing to achieve state-of-the-art results with only 2% of the original CapsNet parameters [13]. This lightweight network allows capsule mechanisms to be applied to complex datasets with more efficient feature encoding, reducing overfitting. Our optimised CapsNet model (Figure 3) starts from an input layer comprising of a 150×100-pixel grayscale image. Primary features were extracted via five convolutional layers. Each output from a convolution layer was followed by a Batch Normalization layer and Rectified Linear Unit (ReLU) activation function, and L2 regularisations were introduced to reduce overfitting [17]. To decrease the number of parameters required for the capsule creation process, there was a depth-wise spatial convolution operation with linear activation following the multilayer convolutional block, mapping extracted features to the primary capsule layer [13,18]. The base expression of features changed from a single neuron to a vectorized capsule. An activation squash function was employed to normalize the probability of a certain entity between zero and one, while vector direction described the entity's attributes. A non-iterative routing algorithm was introduced to exploit a self-attention mechanism. This allowed for reduced numbers of capsules to be routed to the output capsule layer [13,18]. Margin loss was introduced for model optimisation and to determine object class [13,19]. Capsule units in the primary capsule layer were set at 64 to represent characteristic entities. The output capsule layer comprised of two units quantified by 16 vectors to express P and NP classes. Adaptive moment estimation was utilized for training, with batch size set at 30, maximal number of iterations at 260 and initial learning rate at 0.0001. Further details of the CapsNet algorithm and network configuration are respectively shown in Supplementary Figure 1 and Supplementary Table 1.

2.5. Cross-validation and two-stage transfer learning scheme

For each training-cross-validation process, a two-stage transfer learning framework was proposed. PA X-rays upon skeletal maturity (NP group) or immediately prior to treatment (P group) were compiled to pre-train the model, followed by fine-tuning of pre-trained weights upon the dataset of skeletally immature patients with X-rays taken at presentation. Transfer learning allowed for efficient model development with smaller datasets whilst avoiding overfitting and mode collapse. We acquired 328 X-ray for model pre-training and fine-tuning respectively. Flip horizontal was conducted, followed by adjustment for brightness, contrast and sharpness by random factors for data augmentation [20]. There were 2096 paired images for

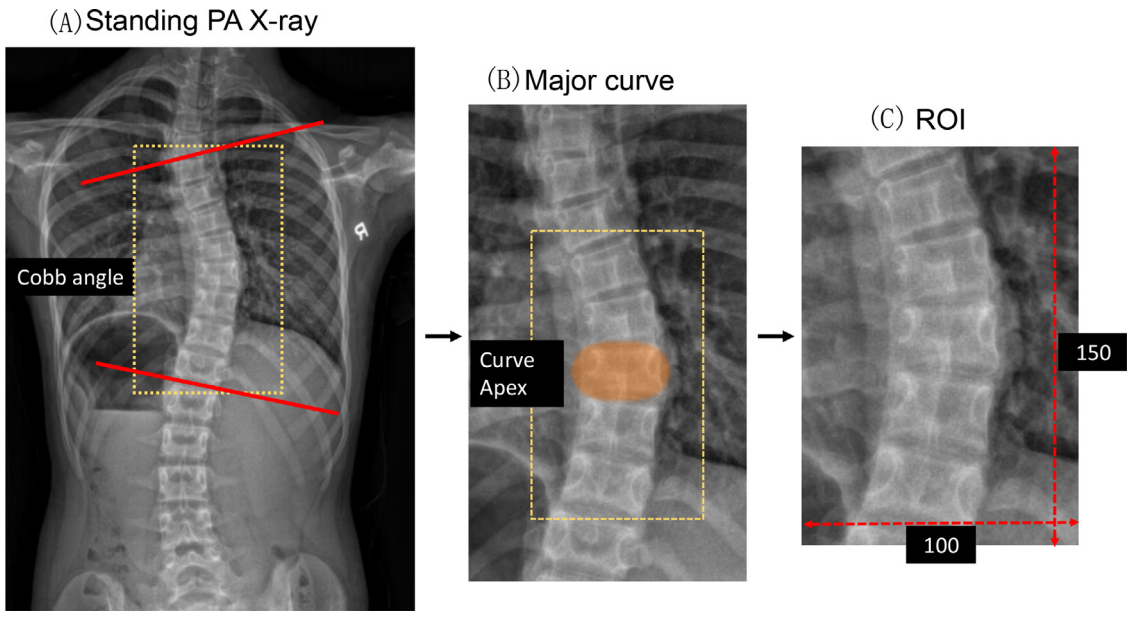


Figure 2. Determining location of the major curve, apex, and region of Interest (ROI) for deep learning (A) The major curve (hatched area) denoted by the largest Cobb angle measured upon standing posteroanterior (PA) X-rays, with upper and lower end vertebra demarcated by red lines. (B) The curve apex (highlighted area) was centered upon in the selection of an ROI measuring 150×100 pixels (C).

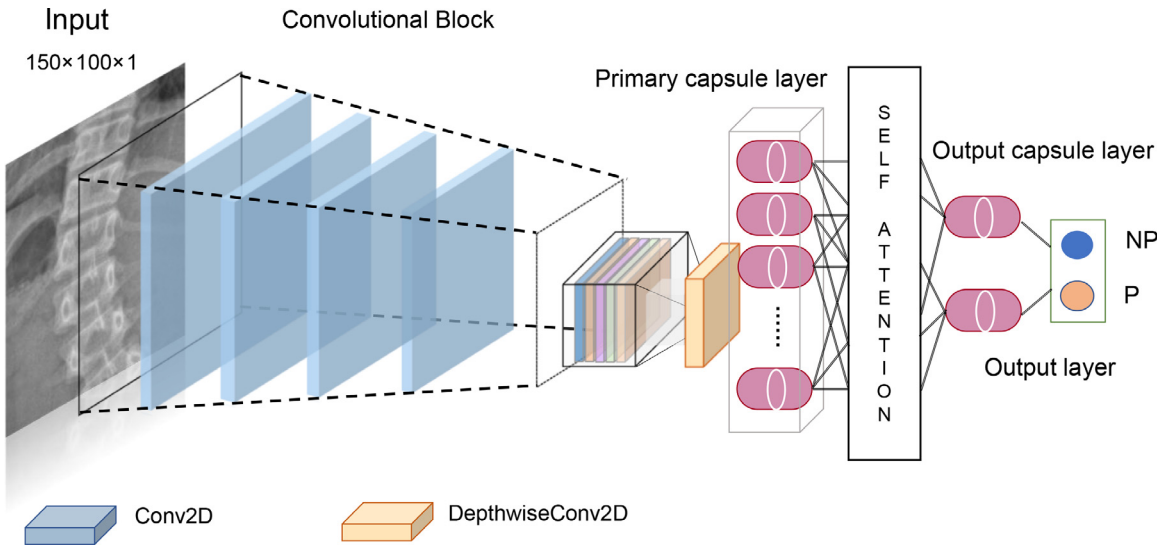


Figure 3. Proposed Efficient CapsNet model with self-attention routing for prediction of progression in adolescent idiopathic scoliosis The model begins with an input layer comprising of a 150×100 pixel grayscale image centered upon the major curve apex. Primary features were extracted via five convolutional layers, followed by a Batch Normalization layer and Rectified Linear Unit (ReLU) activation function, and L2 regularisations were introduced to reduce overfitting. A depth-wise spatial convolutional operation with linear activation followed the multilayer convolutional block, mapping extracted features to the primary capsule layer. A non-iterative routing algorithm was introduced to exploit a self-attention mechanism to efficiently rout reduced numbers of capsules to the output capsule layer. The output capsule layer comprised of two units quantified by 16 vectors to express P and NP classes.

training in each fold cross-validation. For optimal hyperparameter searching, the developmental dataset was divided into five parts. Four parts were utilised for training, while the remaining part was utilized for validation purposes [21]. Augmentation was not conducted upon the validation set in each fold of cross-validation. This process was repeated and an averaged performance was calculated, guiding optimization of model hyperparameters. The proposed model was implemented by Tensorflow and Keras frameworks based on Mazzia's open-source work [13]. The training process was conducted on a server equipped with two NVIDIA Tesla T4 GPUs and 128GB RAM.

2.6. Statistical analysis and model evaluation

Student's t-test was conducted upon continuous variables describing patient characteristics, as well as to compare morphological parameters extrapolated following 3D reconstruction from P and NP curves. We utilized Chi-squared test to compare non-continuous patient variables. A p-value < 0.05 was determined as the threshold for statistical significance. Model prediction performance was evaluated by accuracy, sensitivity, and specificity. SPSS software (v26.0) was used for statistical analysis. For comparison with our CapsNet-based model, a radiomics-based CNN model and deep residual network (ResNet) [22] were implemented under the

Table 1
Clinical and radiological characteristics of the overall patient cohort divided into progressors (P) and non-progressors (NP)

Demographics	Combined cohort	P Group	NP Group	p-value
Number of patients	490	258	232	-
Age (years)	12.1±1.4	12.4±1.1	11.8±1.2	0.259
Sex				0.811
Male	118	61	57	
Female	372	197	175	
Maturity				
Risser sign				0.226
Stage 0	291	149	142	
Stage 1	107	56	51	
Stage 2	92	50	42	
Pre-Menarche at first visit	202	109	93	0.869
Coronal deformity				
Initial Cobb angle of the major curve (°)	20.7±4.5	22.2±4.5	19.7±3.8	0.129
Final Cobb angle of the major curve (°)[a]	27.8±10	34.8±8.8	20±4.3	< **0.0001**
C7PL-CSVL (mm)	13.3±7.4	13.6±7.5	12.4±6.9	0.079
Coronal imbalance[b]	Balanced (409)	Balanced (207)	Balanced (202)	0.060
	Imbalanced (81)	Imbalanced (51)	Imbalanced (30)	
Types of scoliotic curve[c]	RT (189)	RT (98)	RT (91)	0.679
	RTL (48)	RTL (23)	RTL (25)	
	LL (31)	LL (21)	LL (10)	
	LTL (97)	LTL (50)	LTL (47)	
	RT-LL (67)	RT-LL (37)	RT-LL (30)	
	LT-RL (24)	LT-RL (14)	LT-RL (10)	
	Triple (7)	Triple (3)	Triple (4)	
	Other (27)	Other (12)	Other (15)	

[a] Final Cobb angle refers to the latest Cobb angles before initiation of bracing / surgery for patients with progressive curve trajectories, and upon latest follow-up after skeletal maturity for non-progressive patients.
[b] Imbalance was measured as discrepancy between a C7 plumb line and center sacral vertical (CSVL) line exceeding 20 millimeters.
[c] RT, right thoracic; RTL, right thoraco-lumbar; LL, left lumbar; LTL, left thoraco-lumbar; RT-LL, right thoracic-left lumbar; LT-RL, left thoracic-right lumbar; other, left thoracic, right lumbar.

same cross-validation and transfer learning scheme. Furthermore, a logistic regression model (LR) for curve progression utilizing clinical parameters of chronological age, gender, Risser sign, curve imbalance, apex location and Cobb angle of the main curve was included [23]. We assumed values of > 0.5 to represent progression in this LR model.

2.7. Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, model development, data interpretation, or writing of the manuscript.

3. Results

3.1. Demographics of the study cohort

As shown in Figure 1, a total of 490 patients were recruited. Of these, 438 received standing EOS X-rays upon presentation, with 328 randomly selected for model development (training-cross-validation cohort), while the remaining 110 served as an independent dataset for testing. Another cohort of 52 patients receiving standard standing PA X-rays at first visit were recruited for cross-platform validation.

Details regarding age, gender, skeletal maturity, curve types and curve magnitude for the overall study population are shown in Table 1. A total of 258 patients (52.6%) demonstrated curve progression while 232 patients (47.4%) were non-progressors. Female patients (75.9%) were predominant in comparison to male patients (24.1%). There were 291 patients (59.4%) at Risser stage 0 at their first clinical visit, 107 patients (21.8%) at stage 1 and 92 patients (18.8%) at stage 2. A total of 202 female patients (54.3%) were pre-menarchal at first visit. There were no statistical differences between P and NP patients in the overall cohort for any of the analysed variables except for final curve magnitude. There were no significant differences between individual patient cohorts utilised for model development, independent testing, and cross-platform validation (Supplementary Table 2).

3.2. Analysis of three-dimensional morphological parameters

Three-dimensional reconstruction was performed upon presenting X-rays for 138 patients comprising of 65 P and 73 NP curves, allowing for comparison of seven morphological parameters (Table 2). Apical vertebral rotation of the major curve demonstrated the most significant difference between P and NP curve trajectories and was 3.0° higher in the P group. This was followed by torsion, denoting cumulative intervertebral axial rotation spanning the curvature, which was 2.8° higher in the P group. Therefore, our ROI selection was directed towards the major curve apex on standing PA X-rays.

3.3. Performance of CapsNet in comparison to alternative prediction models

The prediction task was performed with our Efficient-CapsNet model as well as more traditional algorithms. Mean values for model

Table 2
3D parameters at first visit in progressive and non-progressive curves

3D Parameter (°)	P group	NP group	p-value
3D Cobb angle	23.7 ± 6.2	24.1 ± 5.3	0.714
Kyphosis (T4-T12)	22.3 ± 8.5	21.4 ± 9.2	0.635
Lordosis (L1-L5)	40.7 ± 9	40.1 ± 11.7	0.227
Apical vertebral rotation	7.3 ± 4.9	4.3 ± 3	**0.006**
Upper curve intervertebral rotation	2.8 ± 1.3	2.7 ± 1.6	0.106
Lower curve intervertebral rotation	2.9 ± 1.7	3.1 ± 1.9	0.143
Torsion	6.1 ± 3	3.3 ± 2.1	**0.020**

3D reconstruction of presenting X-rays demonstrated that apical vertebral rotation and curve torsion were significantly increased in progressive curves (P group) in comparison to non-progressive curves (NP group).

Table 3
Performance comparison between prediction models

Data	Model	Accuracy (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)
Independent Testing (n=110)	LR model	59.0% (56.1-60.4%)	56.2% (54.4-57.1%)	62.4% (59.7-64.0%)
	CNN	56.3% (54.4-57.2%)	54.2% (52.7-55.3%)	59.4% (56.1-62.2%)
	ResNet	61.7% (59.2-63.0%)	58.3% (56.6-60.1%)	62.5% (59.4-64.6%)
	Efficient CapsNet	76.6% (74.9-78.0%)	75.2% (73.3-76.3%)	80.2% (78.8-81.4%)
Cross-platform Testing (n=52)	LR model	58.3% (57.2-61.1%)	55.4% (53.0-56.9%)	63.3% (60.2-65.1%)
	CNN	55.9% (54.1-58.0%)	54.4% (51.6-56.1%)	60.1% (57.7-62.5%)
	ResNet	60.2% (58.6-63.7%)	59.8% (57.6-61.9%)	61.7% (59.4-63.9%)
	Efficient CapsNet	77.1% (75.8-78.1%)	73.5% (71.9-75.6%)	81.0% (79.2-81.9%)

Comparison in model accuracy, sensitivity, and specificity upon PA X-rays captured via EOS imaging apparatus (Independent Testing), and standard PA standing X-rays (Cross-platform Testing). Efficient CapsNet outperformed a residual neural network (ResNet), convolutional neural network (CNN), and a logistic regression (LR) – based model utilising clinical and radiological parameters. CI = confidence interval.

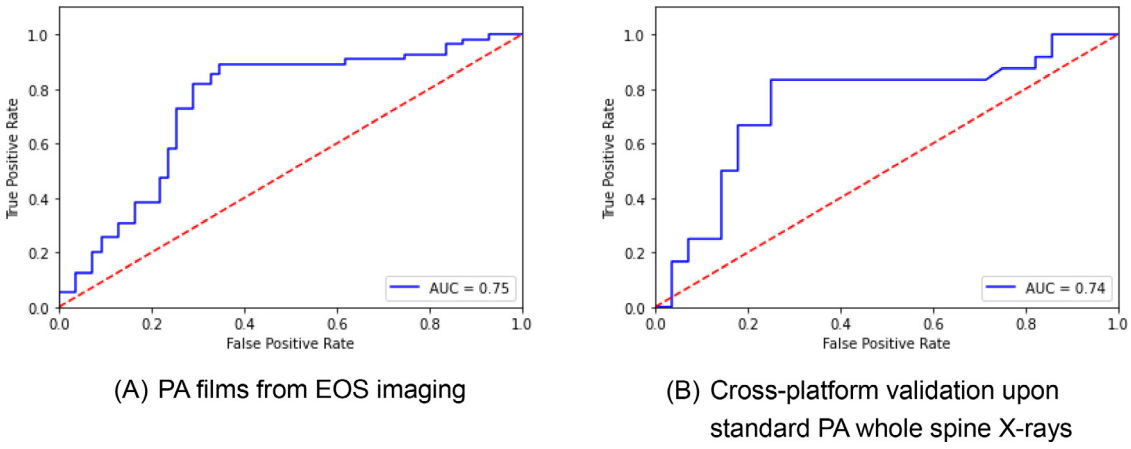


Figure 4. Receiver operating characteristic (ROC) curve for independent testing and cross-platform validation

performance of five repeated experiments are shown in Table 3. The Efficient-CapsNet model achieved an accuracy of 82.1%, sensitivity of 80.8% and specificity of 88.6% on the validation folds. For independent testing on a cohort of 110 patients (55 P and 55 NP), the model achieved an accuracy of 76.6%, sensitivity of 75.2% and specificity of 80.2%. A corresponding ROC (receiver operator characteristic) curve with area under the curve (AUC) of 0.75 is shown in Figure 4A. Upon independent testing, Efficient-CapsNet outperformed CNN (accuracy of 55.7%, sensitivity of 53.3%, and specificity of 60.2%) and ResNet classification models (accuracy of 60.0%, sensitivity of 56.4%, specificity of 60.6%). The clinical and radiological parameters-based regression model also fared poorly in comparison (accuracy of 59.1%, sensitivity of 56.3%, specificity of 61.8%).

Cross-platform validation was conducted to test model performance upon standard standing PA whole spine X-ray images. The Efficient-CapsNet model achieved an accuracy of 77.1%, sensitivity of 73.5% and specificity of 81.0%. A corresponding ROC curve exhibited an AUC of 0.74 as shown in Figure 4B. Again, performance of Efficient-CapsNet was superior to CNN (accuracy of 54.7%, sensitivity of 52.1% and specificity of 55.3%), ResNet (accuracy 58.6%, sensitivity of 57.4%, and specificity of 58.8%), and logistic regression-based classification models (accuracy of 57.7%, sensitivity of 57.1%, specificity of 58.3%).

3.4. Analysis of patients with erroneous trajectory predictions

Three-dimensional curve reconstruction and parameter analysis was performed upon 13 false negative (FN) and 11 false positive (FP) results present within the 110 patients from the independent testing cohort. Values for apical intervertebral rotation (3.5°±2.9°) and torsion (3.9°±2.6°) for false negatives were significantly lower in comparison to the progressive group (Table 4). Conversely amongst false positives, apical vertebral rotation (7.4°±5.8°) and torsion (5.4°±3.3°) significantly exceeded that of non-progressive curves (Table 5). Representative case examples are illustrated in Figure 5. To improve

Table 4
Analysis of 3D parameters in false negatives in comparison to progressive curves

3D Parameter (°)	False negatives	P group	p-value
3D Cobb angle	23.1±3.5	23.7±6.2	0.872
Apical vertebral rotation	3.5±2.9	7.3±4.9	**0.023**
Torsion	3.9±2.6	6.1±3	**0.036**

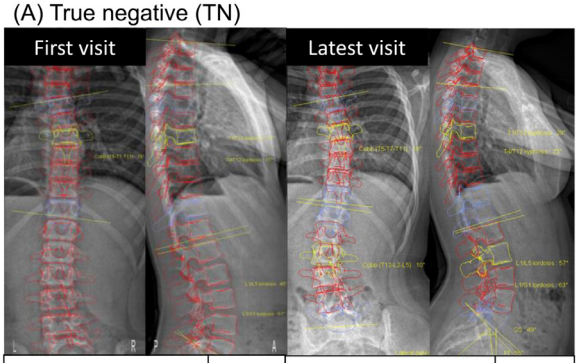
3D reconstruction of false negative results subject to prediction by the CapsNet model demonstrated significantly decreased apical vertebral rotation and torsion as compared to curves correctly predicted to be progressive in trajectory (P group). N = 13 for false negatives, N = 65 for correctly labelled progressive curves within the exploration dataset.

Table 5
Analysis of 3D parameters in false positives in comparison to non-progressive curves

3D Parameter (°)	False positives	NP group	p-value
3D Cobb angle	22.7±4.4	24.1±5.3	0.721
Apical vertebral rotation	7.4±5.8	4.3±3	**0.015**
Torsion	5.4±3.3	3.3±2.1	**0.034**

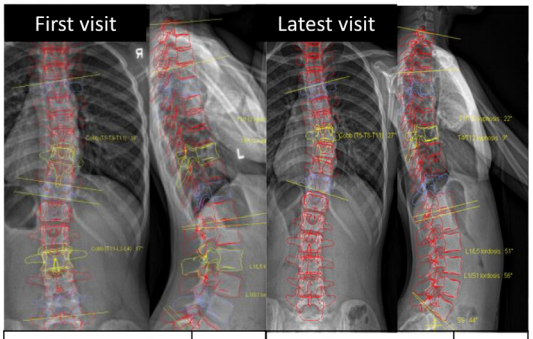
3D reconstruction of false positive results subject to prediction by the CapsNet model demonstrated significantly increased apical vertebral rotation and torsion as compared to curves correctly predicted to be non-progressive in trajectory (NP group). N = 11 for false positives, N = 73 for correctly labelled non-progressive curves within the exploration dataset.

(A) True negative (TN)



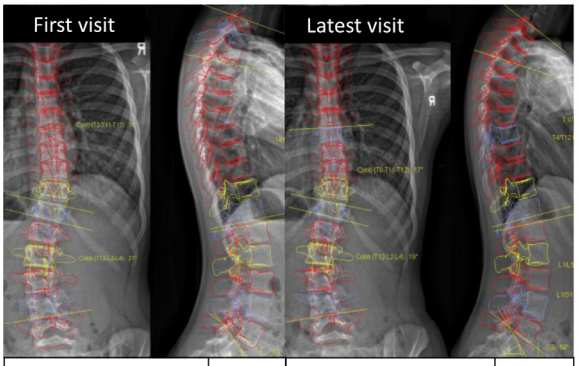
Cobb (T5-T7-T11)	18.2°	Cobb (T5-T7-T11)	17.9°
Apex axial rotation	0.9°	Apex axial rotation	0.6°
Torsion	2.1°	Torsion	1.8°

(B) False negative (FN)



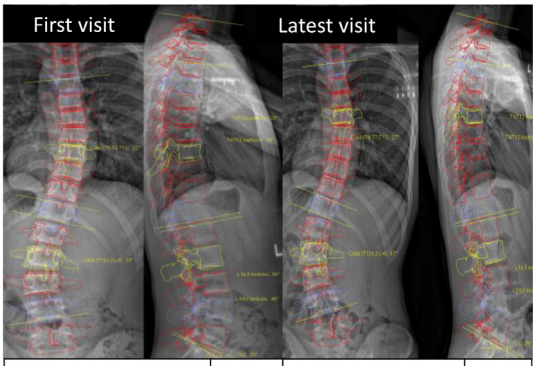
Cobb (T5-T9-T11)	18.3°	Cobb (T5-T9-T11)	27°
Apex axial rotation	0.8°	Apex axial rotation	7.7°
Torsion	0.6°	Torsion	6.8°

(C) False positive (FP)



Cobb (T12-L2-L4)	21°	Cobb (T12-L2-L4)	19.5°
Apex axial rotation	6.7°	Apex axial rotation	8.9°
Torsion	5.6°	Torsion	7.5°

(D) True positive (TP)



Cobb (T12-L2-L4)	23°	Cobb (T12-L2-L4)	37°
Apex axial rotation	7.5°	Apex axial rotation	11°
Torsion	7.4°	Torsion	10.1°

Figure 5. Case illustrations of four classes of patients subject to our predictive model. The true negative patient (A) exhibited minimal apical axial rotation (0.9°) and torsion (2.1°) upon first visit and was correctly classified as a non-progressor. In contrast a false positive patient (C) exhibited increased apical axial rotation (6.7°) and torsion (5.6°) that persisted at latest visit despite there being no progression to the coronal curve magnitude. A false negative patient (B) demonstrated comparatively limited apical rotation (0.8°) and torsion (0.6°) at presentation which 'corrected' by an increased magnitude in combination with Cobb angles at latest visit. A true positive patient (D) demonstrated significant axial rotation (7.4°) and torsion (10.1°) upon presentation, and at next follow-up Cobb angles had increase by 14°.

model performance, we show in Supplementary Table 3 that a 'false negative' patient demonstrated marked increase in apical rotation as well as torsion at a juncture preceding Cobb angle deterioration. Our model accurately predicted progression upon the 2nd set of X-rays when early bracing could still be initiated.

4. Discussion

The objective of our present work was to utilize radiomics to predict curve progression in skeletally immature patients suffering from AIS. We first validated associations between 3D parameters and AIS curve progression [6,14]. This formed the basis for deep learning upon a region of interest centered upon the major curve apex. Following optimisation, our CapsNet-based deep learning model achieved an accuracy, sensitivity, and specificity approaching 80% across different imaging platforms. The capacity for early prognostication of curve trajectories promises to direct pre-emptive treatment. Widespread application of our findings may be feasible not only at first clinic presentation subsequent to referral, but also upon diagnosis at school-age screening programmes, in the absence of specialized equipment and expertise.

Whilst guidelines for intervention exist for 'larger' skeletally immature curves, namely bracing at Cobb angles >25° and surgery when curves exceed 50°, [24] our study was specifically designed to address the management conundrum for those with milder curves, in whom the present practice of watchful waiting remains inadequate [25]. Expert opinion amongst an international multidisciplinary group has deemed treatment of the 20° curve in the skeletally immature to be equipoise, and thus in desperate need of further study [4]. Careful labelling of our dataset was a prerequisite to correct recognition of curve progression within this specific patient cohort. Firstly, we avoided labelling curves deteriorating ≥ 6° yet attaining a magnitude of < 25° at maturity (i.e. from 11° to 24°) as progressive since such curves have minimal effect on cosmesis and function, and observation alone would suffice. Secondly, we excluded from analysis patients that did not reach these thresholds for progression only as a response to brace treatment, as intervention affected natural curve trajectories. The impact of our findings is to allow for triage of patients for referral following diagnosis, as well as to identify those at-risk of progression to receive pre-emptive bracing. Patients at our locale await several months following diagnosis to be reviewed by an orthopaedist, and as curve magnitude alone is an indicator for earlier appointments, those with mild yet progressive curves are attended to in a delayed manner. The effectiveness of bracing is well supported in mild curvatures of < 25° and therefore may be offered as a 'prophylactic' measure after appropriate counselling of at-risk individuals [26]. On the contrary, close observation may be recommended in confidence for non-progressive curves to avoid over prescription as

bracing is a potential source of discomfort, skin irritation, and psychological distress.

Logistic regression-based models utilizing discrete clinical and radiological parameters were proposed early on to prognosticate curve trajectories yet fared poorly upon our dataset [23]. This was expected as indices of age, skeletal maturity, and curve descriptors were of no discriminatory value at baseline. The performance of 3D parameter-based prediction [6] (75% sensitivity, 94% specificity), albeit exceeding our present model, is tempered by the requirement for expensive EOS biplanar imaging apparatus and software reconstruction that is both time consuming as well as prone to measurement error in being a semi-automated process. Most studies utilizing machine learning for AIS have focused upon attaining Cobb angles measurements and classification by curve type [27]. One related study utilized Independent Component Analysis to predict evolution in scoliotic curve morphology over time, which nevertheless was dependent upon 3D reconstruction for feature engineering and simulation [7]. These highlight the advantages of our radiomics-based approach in being non-invasive, cost-effective, and automated after the ROI is specified. The excellent performance of CapsNet compared to other classes of neural networks may be attributed to its capacity to preserve and detect spatial hierarchies between features – unique in the context of scoliosis as a complex multiplanar deformity – and used as a means to supersede 3D reconstruction. Scoliosis screening programs around the world rely on standard (not biplanar) standing X-ray images of the whole spine for definitive diagnosis, following forward bend test and/or Moiré topography for initial screening [25]. Cross-platform validation ensured that our model may be seamlessly incorporated for prognostication upon initial diagnosis.

Our analysis of cases with erroneous prediction of curve trajectories provided valuable insight into discriminatory features identified by the trained model. Apical rotation and torsion were mismatched in comparison to curve trajectory in false positive and false negative cases. Our results come full circle as it is implied that the same 3D parameters that we deemed to be significantly different between P and NP curves in our groundwork were extrapolated via deep learning. It may be argued that reducing the number of false negatives (type II error) is most pertinent to real world application as bracing is relatively safe and well tolerated. A means to rectify this was to apply our model upon serial X-rays.

A limitation to our study was restricting analysis to the major curve alone. The effect of double and triple major curves upon progression risk remains unclear. Further study is indicated in consideration of whole spine X-ray images as input, as well as upon corresponding lateral projections. Utilizing serial images as we described in our analysis of erroneous prediction is another approach that promises improved prognostic accuracy. Our model did not incorporate clinical parameters relating to growth rate nor more comprehensive measures of skeletal maturity (Saunders staging, DRU staging) as the accuracy of Risser scoring, which was part of our inclusion criteria, has been challenged [28]. Finally, it is essential that results are consolidated with larger study populations, as well as to establish external validity across different ethnicities as our cohort was predominantly Chinese.

In summary, we demonstrate that a CapsNet-based deep learning model utilizing X-rays at first visit predicted subsequent curve trajectories. The application of CapsNet was essential to overcoming feature invariance as AIS is a complex three-dimensional deformity. In the context of screening, prediction of progressive curves facilitates early referral. At the specialist clinic one may counsel progressors to receive pre-emptive bracing. This is a necessary advancement upon the present practice of watchful waiting to reveal curve progression, which inevitably delays treatment.

5. Contributors

Conception and design of study - HW, KC, GS; Collection and analysis of data – HW, TZ; Verification of underlying data – HW, GS; Drafting of manuscript – HW, TZ, KC, GS; Authors with access and responsibility to raw data – HW, TZ, GS; Decision to submit manuscript for publication – HW, TZ, KC, GS.

Declaration of Competing Interest

KC has previously received research funding and honoraria from AO spine, Nuvasive, Medtronic, Globus, ICISO, the Japanese Orthoapedic Association, and the University of North Carolina. He is a member of the AOSpine Knowledge Forum Steering Committee, as well as a Nominating Committee Member of the Scoliosis Research Society. All the other authors report no conflicts.

Data sharing statement

The core codes used in this study are available at https://github.com/whongfeiHK/AIS-curve-prediction-Deep-Learning. All raw data used in the study are available from the corresponding author on reasonable request.

Funding

This study was supported by the Society for the Relief of Disabled Children (SRDC).

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2021.101220.

References

[1] Force UPST. Screening for Adolescent Idiopathic Scoliosis: US Preventive Services Task Force Recommendation Statement. JAMA 2018;319(2):165–72.
[2] Cheng JC, Castelein RM, Chu WC, et al. Adolescent idiopathic scoliosis. Nature reviews disease primers 2015;1(1):1–21.
[3] DA J. Natural history of adolescent idiopathic scoliosis: a tool for guidance in decision of surgery of curves above 50°. J Child Orthop 2013;7(1):37–41.
[4] Roye BD, Simhon ME, Matsumoto H, et al. Establishing consensus on the best practice guidelines for the use of bracing in adolescent idiopathic scoliosis. Spine Deform 2020;8(4):597–604.
[5] Schlenzka D, Yrjönen T. Bracing in adolescent idiopathic scoliosis. J Child Orthop 2013;7(1):51–5.
[6] Nault M-L, Beausćjour M, Roy-Beaudry M, et al. A predictive model of progression for adolescent idiopathic scoliosis based on 3D spine parameters at first visit. Spine 2020;45(9):605–11.
[7] Garcia-Cano E, Cosio FA, Duong L, et al. Prediction of spinal curve progression in adolescent idiopathic scoliosis using random forest regression. Computers in biology and medicine 2018;103:34–43.
[8] Melhem E, Assi A, El Rachkidi R, Ghanem I. EOS® biplanar X-ray imaging: concept, developments, benefits, and limitations. J Child Orthop 2016;10(1):1–14.
[9] Keek SA, Leijenaar RT, Jochems A, Woodruff HC. A review on radiomics and the future of theranostics for patient selection in precision medicine. The British journal of radiology 2018;91(1091):20170926.
[10] Lam GC, Hill DL, Le LH, Raso JV, Lou EH. Vertebral rotation measurement: a summary and comparison of common radiographic and CT methods. Scoliosis 2008;3(1):16.
[11] Mehta MH. THE RIB-VERTEBRA ANGLE IN THE EARLY DIAGNOSIS BETWEEN RESOLVING AND PROGRESSIVE INFANTILE SCOLIOSIS. The Journal of Bone and Joint Surgery British volume 1972;54-B(2):230–43.
[12] Jiang X, Wang Y, Liu W, Li S, Liu J. Capsnet, cnn, fcn: Comparative performance evaluation for image classification. International Journal of Machine Learning and Computing 2019;9(6):840–8.
[13] Mazzia V, Salvetti F, Efficient-CapsNet Chiaberge M. Capsule Network with Self-Attention Routing. arXiv preprint arXiv:210112491 2021.
[14] Nault M-L, Mac-Thiong J-M, Roy-Beaudry M, et al. Three-dimensional spinal morphology can differentiate between progressive and nonprogressive patients with

adolescent idiopathic scoliosis at the initial presentation: a prospective study. Spine 2014;39(10):E601.
[15] Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. arXiv preprint arXiv:171009829 2017.
[16] Hinton GE, Sabour S, Frosst N. Matrix capsules with EM routing. International conference on learning representations; 2018:2018.
[17] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. International conference on machine learning; 2015: PMLR; 2015. p. 448–56.
[18] Choi J, Seo H, Im S, Kang M. Attention routing between capsules. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops; 2019; 2019. 0-.
[19] Wu Y, Wu Y, Gong R, et al. Rotation consistent margin loss for efficient low-bit face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020; 2020. p. 6866–76.
[20] Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. 2018 international interdisciplinary PhD workshop (IIPhDW); 2018: IEEE; 2018. p. 117–22.
[21] Browne MW. Cross-validation methods. Journal of mathematical psychology 2000;44(1):108–32.
[22] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016; 2016. p. 770–8.
[23] Peterson L-E, Nachemson AL, Bradford D, et al. Prediction of progression of the curve in girls who have adolescent idiopathic scoliosis of moderate severity. Logistic regression analysis based on data from The Brace Study of the Scoliosis Research Society. Journal of Bone and Joint Surgery-Series A 1995;77(6):823–7.
[24] Negrini S, Donzelli S, Aulisa AG, et al. 2016 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. Scoliosis and Spinal Disorders 2018;13(1):3.
[25] Dunn J, Henrikson NB, Morrison CC, Blasi PR, Nguyen M, Lin JS. Screening for Adolescent Idiopathic Scoliosis: Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 2018;319(2):173–87.
[26] Lateur G, Grobost P, Gerbelot J, Eid A, Griffet J, Courvoisier A. Efficacy of nighttime brace in preventing progression of idiopathic scoliosis of less than 25°. Orthopaedics & Traumatology: Surgery & Research 2017;103(2):275–8.
[27] Chen K, Zhai X, Sun K, Wang H, Yang C, Li M. A narrative review of machine learning as promising revolution in clinical practice of scoliosis. Annals of Translational Medicine 2021;9(1).
[28] Hacquebord JH, Leopold SS. In brief: The Risser classification: a classic tool for the clinician treating adolescent idiopathic scoliosis. Clin Orthop Relat Res 2012;470(8):2335–8.