An intelligent composite model incorporating global / regional X-rays and clinical parameters to predict progressive adolescent idiopathic scoliosis curvatures and facilitate population screening

Hongfei Wang,[a] Teng Zhang,[a] Changmeng Zhang,[a] Liangyu Shi,[a] Samuel Yan-Lik Ng,[a] Ho-Cheong Yan,[a] Karen Ching-Man Yeung,[b] Janus Siu-Him Wong,[a] Kenneth Man-Chee Cheung,[a] and Graham Ka-Hon Shea[a,*]

[a]Department of Orthopaedics and Traumatology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, China
[b]Department of Orthopaedics and Traumatology, Tuen Mun Hospital, Hong Kong, China

eBioMedicine 2023;95: 104768
Published Online xxx
https://doi.org/10.1016/j.ebiom.2023.104768

Summary

Background Adolescent idiopathic scoliosis (AIS) affects up to 5% of the population. The efficacy of school-aged screening remains controversial since it is uncertain which curvatures will progress following diagnosis and require treatment. Patient demographics, vertebral morphology, skeletal maturity, and bone quality represent individual risk factors for progression but have yet to be integrated towards accurate prognostication. The objective of this work was to develop composite machine learning-based prediction model to accurately predict AIS curves at-risk of progression.

Methods 1870 AIS patients with remaining growth potential were identified. Curve progression was defined by a Cobb angle increase in the major curve of ≥6° between first visit and skeletal maturity in curves that exceeded 25°. Separate prediction modules were developed for i) clinical data, ii) global/regional spine X-rays, and iii) hand X-rays. The hand X-ray module performed automated image classification and segmentation tasks towards estimation of skeletal maturity and bone mineral density. A late fusion strategy integrated these domains towards the prediction of progressive curves at first clinic visit.

Findings Composite model performance was assessed on a validation cohort and achieved an accuracy of 83.2% (79.3–83.6%, 95% confidence interval), sensitivity of 80.9% (78.2–81.9%), specificity of 83.6% (78.8–84.1%) and an AUC of 0.84 (0.81–0.85), outperforming single modality prediction models (AUC 0.65–0.78).

Interpretation The composite prediction model achieved a high degree of accuracy. Upon incorporation into school-aged screening programs, patients at-risk of progression may be prioritized to receive urgent specialist attention, more frequent follow-up, and pre-emptive treatment.

Funding Funding from The Society for the Relief of Disabled Children was awarded to GKHS.

Keywords: Adolescent idiopathic scoliosis; Prognosis; Machine learning; Skeletal maturity; Bone mineral density

Introduction

Adolescent idiopathic scoliosis (AIS) is a complex three-dimensional spinal deformity diagnosed when coronal angulation of the spine (Cobb angle) exceeds 10° upon standing spinal radiographs.[1] AIS has an overall population prevalence of up to 5%,[2] and deterioration of the curve following presentation occurs in as many as two-thirds of patients.[3] Severe curves >50° are recommended to receive surgery as they tend to progress beyond skeletal maturity.[4] Skeletally immature patients with curves ≥25° are prescribed bracing to prevent or delay curve progression. As there is prognostic uncertainty in minor curvatures <25° and remaining skeletal growth, doctors adopt a strategy of watchful waiting.[5] Patients with minor curvatures comprise the majority of referrals in countries with school-aged screening programs.[6] The cost-effectiveness of universal screening for AIS has been

*Corresponding author. Department of Orthopaedics and Traumatology, The University of Hong Kong, China.
E-mail address: gkshea@hku.hk (G.K.-H. Shea).

Research in context

Evidence before this study

We searched PubMed on May 5, 2023, using the search terms "deep learning" OR "machine learning" AND "adolescent idiopathic scoliosis curve progression" without language or date restrictions. Previous studies applying artificial intelligence to predict curve progression were predominantly limited to single predictive modalities which included spinal radiographs, 3D spinal parameters, demographic or clinical characteristics, and molecular biomarkers. Approaches utilising information fusion of readily available radiological and clinical parameters had yet to be reported.

Added value of this study

This study applied machine learning techniques upon multiple clinical and radiological domains representing risk factors for scoliosis curve progression. All included parameters were readily available, and automated analysis of hand X-rays provided a means to estimate both bone mineral density as well as skeletal age. The resultant composite deep learning model achieved performance figures surpassing that of comparable single and bimodal prediction platforms.

Implications of all the available evidence

Our platform promises point-of-care prognostication of AIS curvatures. The potential for pre-emptive treatment and a customized surveillance plan, particularly in locales with population-wide screening, represents a substantial advancement in personalized medicine.

challenged when accurate prognostication following diagnosis remains elusive.[3]

Machine learning approaches have been utilised to predict AIS curvature progression. A random forest regression method applied upon reconstructed 3D models predicted for changes in spinal morphology over time.[7] Others have incorporated biomechanical analysis[8] and dynamic patient-specific parameters[9] together with imaging features to predict curve progression. Computer vision has the advantage of automating extraction of imaging features for modeling via big data-driven training approaches. As such, we previously utilised an attention-based capsule neural network (Efficient-CapsNet) upon the major curve apex of spinal radiographs of AIS patients to classify progressive (P) and non-progressive (NP) curves at first clinic visit.[10] Model prediction performance resulted from the recognition of increased vertebral rotation and torsion as indicators of progression risk.[10,11] Similarly, a deep convolutional neural network (DCNN) has been applied upon whole spinal X-rays for AIS prognostication.[12] Nevertheless in these prior paradigms, point-of-care application remains hindered by the necessity for 3D reconstruction, and by mediocre performance when models fail to incorporate risk features for progression beyond spinal radiographic features, which include age, gender, skeletal maturity, and bone mineral density.[2,13,14]

We hypothesized that improved prediction of curve trajectory at first clinic visit following scoliosis screening would result from the development of a composite fusion model accounting for such diverse risk factors. Hand X-rays convey information relevant to the estimation of skeletal maturity as well as bone mineral density. Distal radius and ulna (DRU) grading is an ordinal system that has excellent sensitivity in prognosticating skeletal growth and AIS progression risk,[15] and may be tackled as an image classification problem. Additionally, the 2nd metacarpal cortical index (2MCI) has been validated to be a simple estimate of T-scores when routine DXA scanning is infeasible for AIS prognostication.[16,17]

To formulate a composite prediction platform, we generated an automated pipeline for DRU grading and 2MCI quantification upon left hand X-rays. These were fused to selected clinical data and attention-based convolution neural networks (Attention-ResNet), the latter trained to recognize imaging features upon whole spine X-rays and the major curve apex. Accurate prognostication of scoliosis curvatures is essential towards the formulation of a patient-specific management plan following presentation.

Methods

Ethical statement

Ethical approval was obtained from the Institutional Review Board (IRB) of The University of Hong Kong/Hong Kong West Cluster (Reference number UW22-691). Informed consent was waived as anonymised retrospective patient data was utilised.

Patient recruitment

Three patient cohorts were identified for this study (Supplementary Figure S1). The first cohort ('curve progression cohort') contained 710 patients receiving follow-up in 3- to 6-month intervals, enabling labelling of major curve trajectories from first clinic presentation until skeletal maturity. Additional inclusion criteria consisted of (1) diagnosis of AIS, (2) Cobb angle between 11° and 30° upon standing posteroanterior X-rays at first visit, (3) DRU grading $\leq$ R9U8 to demonstrate growth potential, and (4) regular follow-up concluding at skeletal maturity (R11U9) or upon receiving surgery. This cohort was utilised to develop the spinal X-ray radiomics modules and identified from amongst scoliosis clinic attendees between October 2015 and

September 2021, of which more than 90% were referrals from a two-tiered school-aged screening program.[18] We measured the Cobb angle upon standing whole spine radiographs of the largest coronal curvature by determining the upper and lower end vertebras and drawing a line parallel to the respective upper and lower endplates, as has been described in the Spinal Deformity Study Group radiographic manual. Curve 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 a Cobb angle <25° at skeletal maturity. Patients with NP curves according to these definitions who received bracing were also excluded. With a 7:3 division ratio, 490 patients who presented to the clinic before 15 October 2019 were designated for model training and 5-fold cross-validation, while the remaining 220 patients served as an independent testing set.

The second cohort ('hand X-ray cohort') consisted of 1160 patients aged 8–15 with left hand X-rays utilised for training and testing of the automated DRU grading and 2MCI calculation pipelines. These patients were divided 6:1:3 into a training set (n = 696), validation set (n = 116) and testing set (n = 348). Hand X-rays of poor quality (over/underexposure) or exhibiting bony deformities were excluded. A third cohort ('hand X-ray/DXA cohort') consisted of 326 adult patients with DXA scans and left-hand X-rays obtained within a 1-month period. Similarly, patients with hand X-rays of poor imaging quality (over/underexposure) or exhibiting bony deformities/fractures were excluded. After exclusion, 193 patients remained to validate the correlation between DXA T-scores and automated 2MCI.

Data pre-processing

Conventionally, 2nd metacarpal index (2MCI) and DRU grading are manually measured (Fig. 1A and B). In preparation for automated hand X-ray analysis, an experienced orthopaedic researcher used Roboflow to label regions of interest (ROIs) corresponding to i) the 2nd to 4th metacarpals, and ii) distal radial and ulnar physis. Pixel-level segmentation labels of the second metacarpus and the corresponding intramedullary mid diaphysis were subsequently labelled. Skeletal maturity indices (DRU and Sanders staging) from both the hand X-ray cohort as well as curve progression cohort were labelled by two experienced orthopaedic researchers.

Features contained within the major curve apex of posteroanterior (PA) spinal radiographs predict curve progression due to their capacity to convey rotation and torsion. On the other hand, whole spine X-rays facilitate assessment of global spinal imbalance as a risk factor for curve progression.[12] Therefore, we extracted a regional spinal X-ray ROI (300 × 200-pixel fixed window) centred upon the apical vertebrae/disc of the major curve as a, together with at least two adjacent vertebras above and below with lateral rib articulations (Fig. 1C). We also extracted a global spinal X-ray ROI (300 × 300-pixel fixed window) covering T1 to the sacrum together with clavicles, ribs, and pelvis. All ROI images were saved as single channel grayscale image in JPG formatting.

Information related to the clinical parameter-based prediction module (age, gender, menarche status, Sanders stage) was retrieved from electronic patient records (EPR), whilst Cobb angles and coronal shift (C7 plumb line to central sacral vertical line; C7PL-CSVL) were measured by two orthopaedic surgeons.

Pipeline for automated analysis of hand X-rays

Automated calculation of second metacarpal index (2MCI)

The first step was automated extraction of the 2nd to 4th metacarpals (Supplementary Figure S2A) upon pre-labelled left-hand X-rays via an open-source object detection framework (Yolov5, ultralytics.com/yolov5). We then implemented an attention mechanism-based encoder-decoder deep structure (Attention U-Net) to achieve pixel-level segmentation of the second metacarpal (Supplementary Figure S2B), containing both cortical and intramedullary regions. The encoder consisted of five convolution layers followed by a pooling dimension reduction operation. Multi-scale feature-maps captured contextual information and were merged through skip connections to combine coarse- and fine-level dense predictions. Attention gating was introduced before the concatenation operation to disambiguate irrelevant responses and preserve activations salient to second metacarpal segmentation. The decoder increased feature map dimensions through up-sampling operations to extract semantic information from combined feature maps and generated a segmented mask.

Quantification of 2MCI (Supplementary Figure S2C) proceeded from the outputted mask via a zero-padding resize operation of metacarpal crops to generate 400 × 400-pixel inputs. The optimized Attention-U-Network generated segmentation results of the second metacarpal with zero-pixel value black backgrounds. Rule-based image rotation was applied so the diaphyses were orientated vertically. The middle-third of the 2nd metacarpal was extracted and resized into a 128 × 128-pixel ROI. Another pretrained Attention-U-Net received the mid-diaphyseal ROIs to predict segmentation masks of the intramedullary portion. 2MCI could then be calculated as a ratio of mid-metacarpal total and intramedullary area according to the formula:

$$2MCI = 1 - \frac{number\ of\ pixels\ in\ mid\text{–}diaphysis\ intramedullary\ segment}{number\ of\ pixels\ in\ total\ mid\text{–}diaphysis\ segment}$$

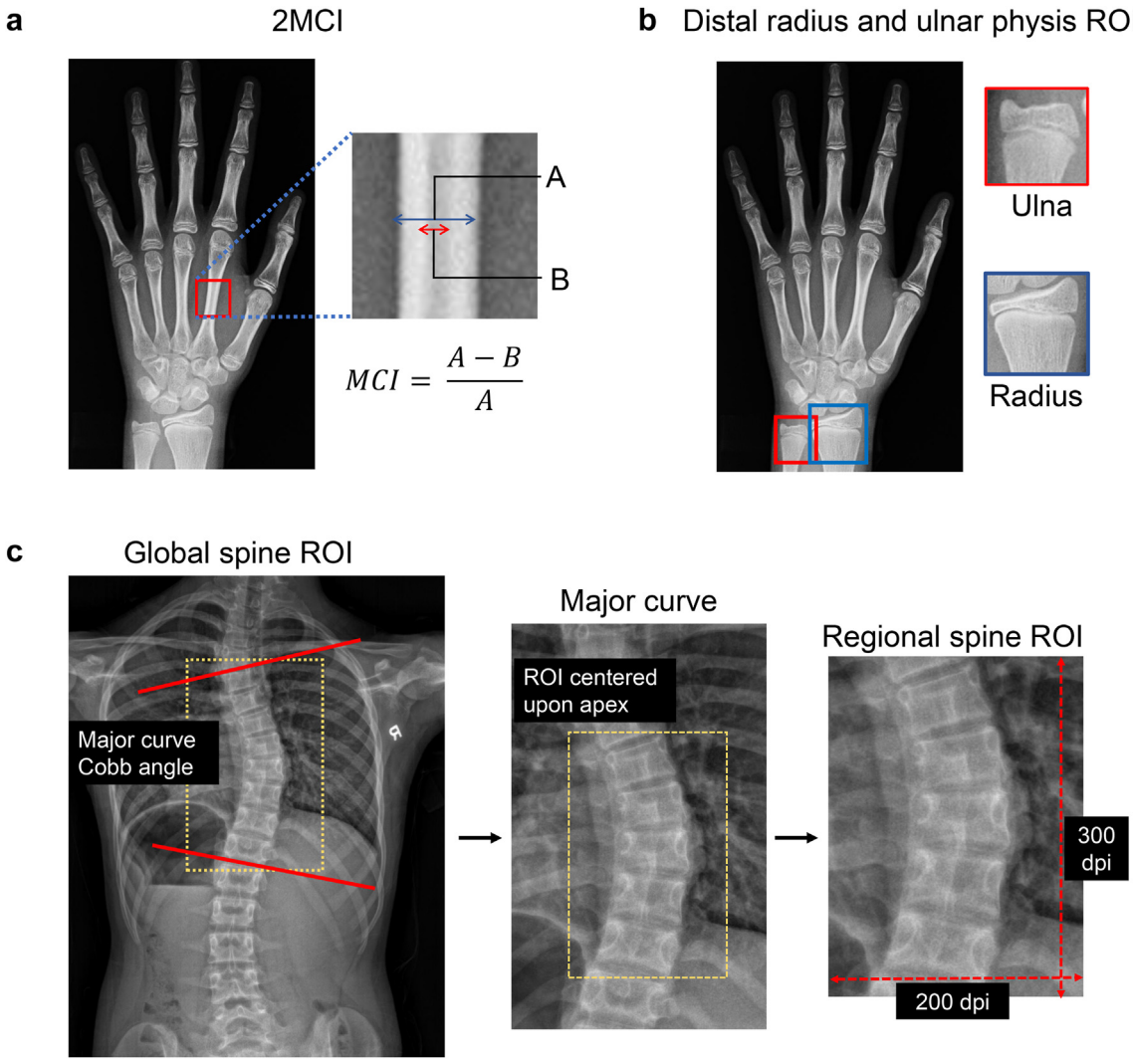

Fig. 1: Definition of 2nd metacarpal index (2MCI) and X-ray regions of interest (ROI).

Automated grading of distal radius and ulna physeal maturity

We first trained an open-source object detection framework (Yolov5, ultralytics.com/yolov5) to automatically extract the distal radius and ulna regions upon prelabeled left-hand X-rays (Supplementary Figure S2A), and thereafter, applied a Residual Convolutional Neural Network (ResNet) for assessment of DRU grading (Fig. 2A). This was approached as a regression problem with continuous output since X-ray changes were graduated. After feature mapping by fully connected layers, distal radius and ulna grading estimates were attained as numerical outputs rounded to the nearest integer.

Composite model architecture

Automated assessment of 2MCI and DRU was achieved as described previously. With regards to spinal radiographs, two residual attention networks (Attention-ResNet-56)[19] were implemented to receive the whole spine and major curve apex ROIs of P and NP curvatures (Fig. 1C). Image encoding blocks in the pretrained Attention-ResNet models were extracted with individual network weights generated for both the whole spine and curve apex ROIs (Fig. 2C–E).

With regards to the composite model, values achieved for 2MCI and DRU (Fig. 2A) were integrated with clinical variables (Fig. 2F) and then encoded by a multilayer perceptron (MLP). A late fusion strategy was employed to transfer pretrained weights (see below) from global and regional spine X-rays. High dimensional features from global and regional spinal X-rays as well as MLP-encoded variables were flattened and catenated into a single vector. Fully connected neural networks with attention activation were

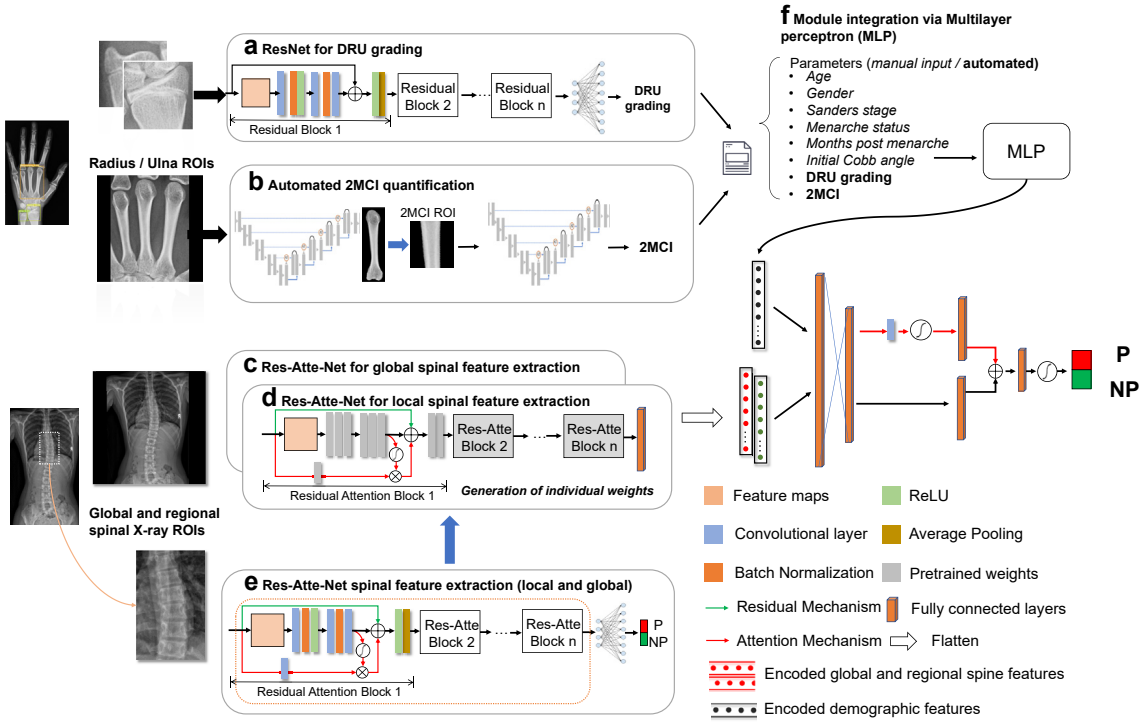

Fig. 2: Overview of composite prediction model architecture. PA = posteroanterior; P = progressive; NP = non-progressive; DRU = distal radius and ulnar; MCI = metacarpal index; MLP = Multilayer perceptron.

subsequently employed to prognosticate progression/non-progression.

Composite model training and validation

A three-stage training/validation/testing scheme was implemented on the object detection model, 2MCI calculation model, and DRU staging model. Furthermore, a two-stage transfer learning strategy was proposed for training upon the attention-ResNet-based spinal radiograph module. Global or regional ROIs of spine X-rays taken upon skeletal maturity (NP group) or immediately prior to treatment (P group) were compiled to pre-train the attention-ResNets, followed by fine-tuning of pre-trained weights on the dataset of skeletally immature patients with X-rays taken at first clinic visit. Five-fold cross-validation was employed for optimal hyperparameter searching. There were 3136 paired data for training in each fold of cross-validation, which included original images and those generated by modification of brightness, contrast, sharpness, as well as flip horizontal images. Regarding composite model training, image feature encoding weights were retained from pretrained values, and only MLP weights for feature fusion were optimized for under the data augmentation and 5-fold cross-validation strategy described above. The proposed model was implemented by Tensorflow and Keras frameworks and based upon Python 3.8. The training process was conducted on a server equipped with two NVIDIA Tesla T4 GPUs and 128 GB RAM. In order to assess erroneous predictions, 3D reconstruction was conducted by a, orthopaedic researcher using sterEOS software (v1.6) upon biplanar X-rays obtained at first clinic visit.[20]

Statistical analysis

Wilcoxon rank sum test was conducted upon continuous and ordinal variables, whilst Chi-squared test was conducted upon categorical variables. Logistic regression was performed for covariate adjustment towards prognostication of curve trajectory using 2MCI. Spearman correlation test was conducted upon estimated 2MCI and the manually measured ground truth amongst the adult population. SPSS (version 28.0.1) was used for statistical analysis. All statistical tests were two sided and a *P*-value <0.01 was determined as the threshold for statistical significance relating to model performance, which was relaxed to <0.05 for analysis of clinical parameters resulting erroneous trajectory predictions.

The segmentation model for 2MCI calculation was evaluated by IoU measures (Intersection Over Union) in comparison to manually measured ground truths. The DRU grading model was evaluated by mean absolute percentage error in comparison to manually graded ground truths. Single modality and composite model prediction performance was evaluated via measures of accuracy, sensitivity, specificity, and area under the

curve (AUC) upon receiver operator characteristic (ROC) curves. Five-times repeated model training using optimized hyper-parameters together with bootstrap sampling (n = 10,000) on the independent testing set was employed to generate 95% confidence intervals.

Role of funders

Funding provided by The Society for the Relief of Disabled Children enabled for the recruitment of graduate students and research assistants to perform the research. The funders had no role in study design, data collection, data analyses, interpretation, nor writing of report.

Results

Demographics of the study cohort

Clinical details regarding the 'curve progression' cohort are summarized in detail upon Table 1. Average age at first clinic presentation was 12.82 ± 1.44 and the cohort was female dominant (72.5%). According to our study criteria, 307 out of 710 patients (43.2%) demonstrated curve progression, and final Cobb angles differed by more than 14° (35.12° ± 7.58° in progressors vs. 20.24° ± 4.39° in non-progressors, $P < 0.001$, Wilcoxon rank sum test). The NP group was significantly more skeletally mature than the P group in accordance with DRU and Sanders grading ($P < 0.001$, Wilcoxon rank sum test) but not Risser score ($P = 0.052$, Wilcoxon rank sum test). The P group was more likely to be pre-menarchal at first clinic visit ($P < 0.001$, Chi-squared test), demonstrated lower values for manually measured 2MCI ($P < 0.001$, Wilcoxon rank sum test), and exhibited greater coronal shift ($P < 0.001$, Wilcoxon rank sum test) than the NP group.

Performance of automated hand X-ray pipeline

Quantification of 2MCI

The pretrained Attention-U-Net achieved an IoU of 94.7% (94.2–95.5%; 95% CI) for second metacarpal segmentation in comparison to the ground truth. An IoU of 98.6% (97.6–98.9%, 95% CI) and 97.7% (96.5%–98.2%) was respectively attained in segmenting the mid-

Variable	Combined Cohort	P group	NP group	P value
Number of patients	710	307	403	
Age	12.82 ± 1.44	12.23 ± 1.34	13.27 ± 1.30	**<0.001**
Gender				0.048
Male	195	72	123	
Female	515	235	280	
Skeletal Maturity				
Distal radius grade	7.76 ± 0.92	7.4 ± 0.94	8.0 ± 0.81	**<0.001**
Distal ulna grade	6.48 ± 0.86	6.15 ± 0.8	6.73 ± 0.8	**<0.001**
Sanders score	4.52 ± 1.89	3.49 ± 1.56	5.22 ± 1.77	**<0.001**
Risser sign	1.47 ± 0.77	1.39 ± 0.79	1.71 ± 0.6	0.052
Menarche status				**<0.001**
Pre (at first visit)	162	130	32	
Post (at first visit)	353	105	248	
Months post menarche	13.73 ± 9.59	10.43 ± 7.84	14.42 ± 9.8	0.045
Second metacarpal index				
Manually measured	0.49 ± 0.12	0.47 ± 0.09	0.51 ± 0.11	**<0.001**
Automated	0.48 ± 0.06	0.47 ± 0.06	0.50 ± 0.06	**<0.001**
Coronal deformity				
Initial Cobb angle (°)[a]	20.54 ± 4.08	21.95 ± 4.08	19.46 ± 3.74	**<0.001**
Final Cobb angle (°)[b]	26.67 ± 9.49	35.12 ± 7.58	20.24 ± 4.39	**<0.001**
C7PL-CSVL (mm)	13.3 ± 6.9	13.6 ± 7.7	12.8 ± 4.2	**<0.001**
Types of scoliotic curve	RT (253)	RT (107)	RT (146)	0.558
	RTL (75)	RTL (34)	RTL (41)	
	LL (50)	LL (23)	LL (27)	
	LTL (134)	LTL (56)	LTL (78)	
	RT-LL (103)	RT-LL (44)	RT-LL (59)	
	LT-RL (39)	LT-RL (17)	LT-RL (22)	
	Triple (14)	Triple (7)	Triple (7)	
	Other (42)	Other (19)	Other (23)	

Bold indicates $P < 0.01$. RT: right thoracic curve; RTL: right thoracolumbar; LL: left lumbar; LTL: left thoracolumbar; RT-LL: right thoracic-left lumbar; LT-RL, left thoracic-right lumbar. [a]Initial Cobb angle of the major curve. [b]Final Cobb angle of the major curve upon post-maturity follow-up in NP cases, and prior to initiating of bracing or surgery for P cases.

Table 1: **Characteristics of the patient cohort recruited for composite model development.**

diaphyseal region and intramedullary portion of the mid-diaphysis. Pixel segmentation error predominantly occurred over the metacarpal head and base regions. As a proof-of-principle that automated 2MCI calculations approximated BMD, we compared automated 2MCI results to DXA scores in an adult population. Spearman correlation testing (Supplementary Figure S3) revealed moderate correlation with T-scores at the left femoral neck (n = 189, r = 0.460; $P < 0.001$), left total hip (n = 190, r = 0.470; $P < 0.001$) and left trochanter (n = 103, r = 0.400; $P < 0.001$). Subsequently, 2MCI scores were ascertained for P and NP patients. 2MCI scores adjusted by logistic regression for gender, age and DRU staging indicated that it was an independent factor for curve progression (0.47 ± 0.06 in P group, 0.50 ± 0.06 in NP group; $P = 0.002$, Wilcoxon rank sum test), with an adjusted odds ratio of 0.01 (0.001–0.19, 95% CI).

Distal radius and ulna physeal grading

Classification performance of the ResNet DRU assessment model was evaluated in the curve progression cohort compared to the manually labelled ground truth. The model achieved an accuracy of 86.1% (84.3–87.0%; 95% CI), specificity of 81.8% (80.1–82.9%), and sensitivity of 85.2% (83.9–86.1%) for radius staging, and an accuracy of 85.4% (84.7–87.0%; 95% CI), specificity of 75.6% (74.3–76.4%), and sensitivity of 91.25% (89.6–92.4%) for ulnar staging. This corresponded to a mean absolute percentage error of 5.35 ± 2.83 and 5.92 ± 3.9 for distal radius and ulna grading respectively. Confusion matrices of automated DRU grading in comparison to manual labels are shown in Supplementary Figure S4.

Evaluation of composite prediction model performance

Prediction task performance was evaluated upon the composite platform in comparison to single and bimodal prediction domains. Mean values for model performance following five repeated experiments are summarized in Table 2, with corresponding receiver operator characteristic (ROC) curves shown in Fig. 3.

Regarding single modality prediction, performance of a logistical regression model utilising clinical data alone (age, gender, menarche status, Sanders stage, automated DRU grading/2MCI) demonstrated an accuracy of 69.5% (66.2–71.9%; 95% CI), sensitivity of 68.8% (66.5–72.1%), specificity of 71.4% (69.5–73.2%), and AUC of 0.65 (0.63–0.69). Attention ResNet applied upon the whole spine ROI achieved an accuracy of 73.4% (71.3–74.8%; 95% CI), sensitivity of 68.7% (64.6–69.2%), specificity of 74.0% (72.6–76.7%), and AUC of 0.74 (0.69–0.77). Attention Resnet applied upon the major curve apex achieved an accuracy of 77.6% (74.5–78.2%; 95% CI), sensitivity of 78.4% (74.0–76.2%), specificity of 75.2% (75.8–79.6%) and AUC of 0.78 (0.75–0.79). We also evaluated the performance of our previously described Efficient CapsNet model,[10] which achieved an accuracy of 71.8% (70.1–72.7%; 95% CI), sensitivity of 70.2% (69.1–72.4%), specificity of 72.7% (70.7–73.8%) and AUC of 0.73 (0.71–0.74) upon the dataset.

Incorporation of an additional prognosticative domain improved upon model performance. The combination of global spine ROI and clinical data with a late fusion strategy achieved an accuracy of 77.8% (75.7–78.2%; 95% CI), sensitivity of 75.1% (76.2–78.6%), specificity of 79% (77.1–79.7%) and AUC of 0.78 (0.74–0.81). The combination of the regional spine ROI (major cure apex) together with clinical data achieved an accuracy of 80.1% (78.6–81.2%; 95% CI), sensitivity of 78.1% (76.2–78.6%), specificity of 82.2% (80.4–82.7%) and AUC of 0.81 (0.78–0.83). Multiscale performance (combination of global and regional spine ROIs) did not result in much improvement compared to the prior combination of clinical and radiological domains, and only achieved an accuracy of 76.4% (74.3–77.6%; 95% CI), sensitivity of 74.3% (72.7–75.1%), specificity of 78.1% (77.2–79.6%) and AUC of 0.79 (0.75–0.81).

The composite prediction model combining multiscale spinal X-rays, hand X-rays, and clinical data achieved the best performance figures. Model accuracy was

Prediction platform	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
Single modality prediction				
Clinical parameters	69.5% (66.2–71.9%)	68.8% (66.5–72.1%)	71.4% (69.5–73.2%)	0.65 (0.63–0.69)
Atte-ResNet upon whole spine	73.4% (71.3–74.8%)	68.7% (64.6–69.2%)	74.0% (72.6–76.7%)	0.74 (0.69–0.77)
Atte-ResNet upon curve apex	77.6% (74.5–78.2%)	75.2% (74.0–76.2%)	78.4% (75.8–79.6%)	0.78 (0.75–0.79)
Efficient-CapsNet	71.8% (70.1–72.7%)	70.2% (69.1–72.4%)	72.7% (70.7–73.8%)	0.73 (0.71–0.74)
Bimodal prediction				
Atte-Resnet (whole spine) + clinical parameters	77.8% (75.7–78.2%)	75.1% (76.2–78.6%)	79.0% (77.1–79.7%)	0.78 (0.74–0.81)
Atte-Resnet (Whole spine + curve apex)	76.4% (74.3–77.6%)	74.3% (72.7–75.1%)	78.1% (77.2–79.6%)	0.79 (0.75–0.81).
AtteResnet (curve apex) + clinical parameters	80.1% (78.6–81.2%)	78.1% (76.2–78.6%)	82.2% (80.4–82.7%)	0.81 (0.78–0.83)
Composite prediction				
Definitive composite model	83.2% (79.3–83.6%)	80.9% (78.2–81.9%)	83.6% (78.8–84.1%)	0.84 (0.81–0.85)

Table 2: **Performance comparison between single, bimodal, and composite prediction models.**

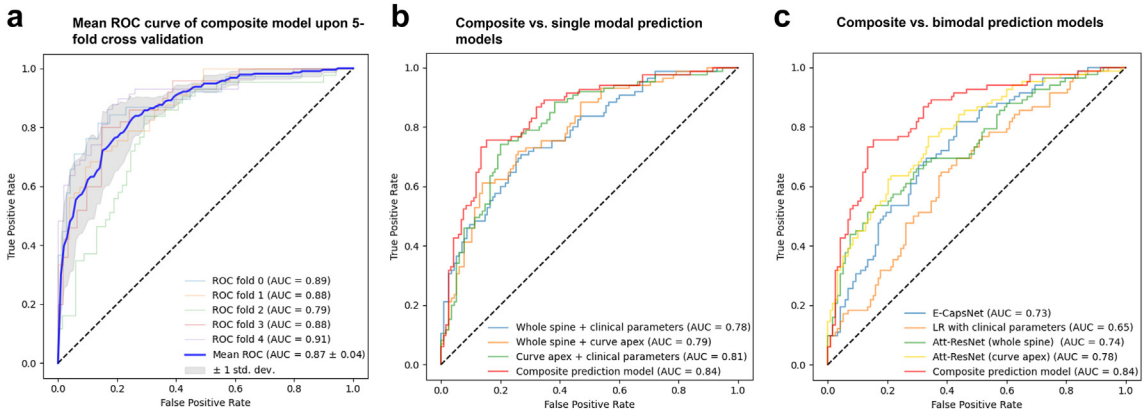

Fig. 3: ROC curves for composite prediction model in comparison to single modal and bimodal prediction models.

83.2% (79.3–83.6%; 95% CI), sensitivity was 80.9% (78.2–81.9%), specificity was 83.6% (78.8–84.1%) and AUC was 0.84 (0.81–0.85). ROC (receiver operator characteristic) curves upon 5-fold cross validation achieved an AUC of 0.87 (0.83–0.91).

Analysis of improved prediction performance upon the composite model

We identified false positive (FP) and false negative (FN) results based on three single modality prediction models (clinical parameters, global spine X-ray ROI and regional spine X-ray ROI) which were correctly labelled by composite prediction. The top 15% of corrected predictions (with greatest variation in output estimators between single modal and composite model) were identified from the independent testing dataset as well as each validation fold. Clinical parameters were retrieved from these patients, and 3D reconstruction of the spinal column was performed to facilitate analysis of radiological variables.

As shown in Supplementary Table S1, FN results from the clinical parameters-based model differed from the NP group in apical vertebral rotation and torsion. Correct labelling by the composite model indicated the capacity of an additional radiomics-based module to recognize and weigh these imaging factors towards accurate prognostication. When attention ResNet alone was applied upon global spine ROIs, FNs resulted when imaging features failed to account for significantly lower age and skeletal maturity (reflected by DRU and Sanders scores). Upon ResNet analysis of the curve apex alone, FNs exhibited a significant reduction in 2MCI, implicating osteopenia as an unrecognized risk factor. As summarized in Supplementary Table S2, the analysis of FPs similarly demonstrated the capacity of the composite predictive model to identify and weigh multiple discordant prediction domains and in so doing improve upon overall performance.

Discussion

AIS is a complex disease with genetic, metabolic, biomechanical, and environmental theories having been proposed to explain its aetiology.[2] The objective of our present work was to utilise machine learning upon multiple clinical and radiological risk factors towards accurate prediction of AIS curve trajectories at first clinic visit. Our platform facilitates point-of-care prognostication to direct the management of AIS in skeletally immature patients. At-risk patients may receive urgent referral to specialist centres for close radiological surveillance or pre-emptive treatment. On the contrary, patients with NP trajectories may receive less frequent clinical surveillance towards safe and efficient allocation of healthcare resources.

The latest recommendation statement from The United States Preventive Services Task Force (USPST) in 2018 upgraded the evidence for scoliosis screening from “D” (discouraged) to “I” (uncertainty about the balance of benefit and harms of the service).[3] Literature supporting the diagnostic accuracy of simple screening tests as well as the efficacy of brace treatment or scoliosis-specific exercises was tempered by lack of direct evidence regarding the cost-effectiveness and long-term beneficial outcomes for AIS treated in adolescence.[21] In response to these findings, many have indicated a pressing need for further research on scoliosis prognostication towards guiding treatment initiation.[22] To the best of our knowledge, this is the first study to apply advances in digital medicine for scoliosis prognostication that has achieved performance figures that may facilitate triage and direct early intervention. In the context of screening programs, at-risk patients may next receive pre-emptive bracing[21] or exercise treatment[23] as evidence-based interventions, and surveillance intervals may be adjusted such that patients at low-risk of progression are scheduled for longer follow-up intervals due to concerns about the hazards of regular

radiation exposure.[24] Conversely in locales that lack screening and where experience in managing scoliosis may be lacking, our prognostic platform guides the non-specialist as to whether urgent attention is required. These amount to providing personalized care across a variety of healthcare contexts.

Introducing attention mechanisms to convolution neural networks (CNN) represents a recent advancement in computer vision model architectures to suppress irrelevant regions in an input image while highlighting salient features.[25] Both attention-based CNNs and deep neural networks were employed in our prediction modules. The radiomics component of the composite model was achieved by integration of multiple pretrained feature extraction networks followed by fine-tuning. Hand X-rays provided an additional dimension for analysis in estimating skeletal maturity as well as bone mineral density. Whilst osteopenia has been associated with AIS progression risk,[14] 2MCI had not previously been utilised towards AIS prognostication. Our composite model achieved an accuracy, sensitivity, and specificity that outperformed single modal and bimodal prediction models, as well our previously described CapsuleNet-based architecture.[10] Compared to CapsuleNet, ResNet is more favourable towards composite integration[26] whilst avoiding overfitting. In contrast to other imaging-based prognostication systems, there was no need to rely upon biplanar imaging systems,[27] 3D reconstruction,[27] DXA scanning[14] and operator-dependent ultrasonography.[28] Despite extensive work performed on the genetics of AIS,[29–31] this has failed to yield loci of sufficient reproducibility and penetrance for clinical prognostication.

Our model would benefit from external validation since training was performed on a predominantly Southern Chinese population. Fine-tuning would likely be required upon patients of other ethnicities due to differences in body build and pubertal development.[32] Prospective studies would consolidate findings and have important ramifications as to whether population-wide screening should be more widely adopted. An underlying assumption towards pre-emptive intervention in response to model prediction is that curves deemed to be progressive would respond to bracing or exercise; whilst intuitive, this needs to be demonstrated. Our findings are also specific to AIS alone, as the pathophysiology and natural history of early onset scoliosis, for example due to congenital and neuromuscular causes, is distinct.[33] In future model re-iterations, an objective should be to reveal the timing, magnitude, and rate of curve deterioration.

Conclusion

We demonstrated that a composite machine learning model could achieve accurate prediction of AIS curve trajectories. This promises to facilitate point-of-care prognostication to guide management. The potential for pre-emptive treatment and a personalized surveillance plan represents a substantial advancement in disease management.

Contributors

Conception and design of study–HW, KC, TZ, GS; Collection and analysis of data—HW, CZ, LS, SYLN, YHC, KCMY, JSHW, GS; Verification of underlying data—HW, GS; Drafting of manuscript—HW, GS. All authors read and approved the final version of the manuscript.

Data sharing statement

The core codes used in this study are available at https://github.com/whongfeiHK/AIS-composite-model. All raw data used in the study are available from the corresponding author upon reasonable request.

Declaration of interests

KC has previously received research funding and honoraria from AO spine, Nuvasive, Medtronic, Globus, ICISO, the Japanese Orthopaedic 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.

Acknowledgements

The authors would like to thank The Society for the Relief of Disabled Children for their generous research grant supporting this project.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2023.104768.

References

1 Kim H, Kim HS, Moon ES, et al. Scoliosis imaging: what radiologists should know. *Radiographics*. 2010;30(7):1823–1842.
2 Cheng JC, Castelein RM, Chu WC, et al. Adolescent idiopathic scoliosis. *Nat Rev Dis Prim*. 2015;1:15030.
3 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):173–187.
4 Weinstein SL, Dolan LA, Spratt KF, Peterson KK, Spoonamore MJ, Ponseti IV. Health and function of patients with untreated idiopathic scoliosis: a 50-year natural history study. *JAMA*. 2003;289(5):559–567.
5 Negrini S, Aulisa AG, Aulisa L, et al. 2011 SOSORT guidelines: orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth. *Scoliosis*. 2012;7(1):3.
6 Altaf F, Drinkwater J, Phan K, Cree AK. Systematic review of school scoliosis screening. *Spine Deform*. 2017;5(5):303–309.
7 García-Cano E, Arámbula Cosío F, Duong L, et al. Prediction of spinal curve progression in adolescent idiopathic scoliosis using random forest regression. *Comput Biol Med*. 2018;103:34–43.
8 Tajdari M, Pawar A, Li H, et al. Image-based modelling for Adolescent Idiopathic Scoliosis: mechanistic machine learning analysis and prediction. *Comput Methods Appl Mech Eng*. 2021;374:113590.
9 Tajdari M, Tajdari F, Shirzadian P, et al. Next-generation prognosis framework for pediatric spinal deformities using bio-informed deep learning networks. *Eng Comput*. 2022;38(5):4061–4084.
10 Wang H, Zhang T, Cheung KM, Shea GK. Application of deep learning upon spinal radiographs to predict progression in adolescent idiopathic scoliosis at first clinic visit. *eClinicalMedicine*. 2021;42:101220.
11 Nault ML, Mac-Thiong JM, 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 (Phila Pa 1976)*. 2014;39(10):E601–E606.
12 Yahara Y, Tamura M, Seki S, et al. A deep convolutional neural network to predict the curve progression of adolescent idiopathic scoliosis: a pilot study. *BMC Muscoskel Disord*. 2022;23(1):610.
13 Lenz M, Oikonomidis S, Harland A, et al. Scoliosis and Prognosis—a systematic review regarding patient-specific and radiological

predictive factors for curve progression. *Eur Spine J*. 2021;30(7):1813–1822.
14 Hung VWY, Qin L, Cheung CSK, et al. Osteopenia: a new prognostic factor of curve progression in adolescent idiopathic scoliosis. *J Bone Joint Surg Am*. 2005;87(12):2709–2716.
15 Cheung JP, Cheung PW, Samartzis D, Cheung KM, Luk KD. The use of the distal radius and ulna classification for the prediction of growth: peak growth spurt and growth cessation. *Bone Joint J*. 2016;98-b(12):1689–1696.
16 Evans RA, McDonnell GD, Schieb M. Metacarpal cortical area as an index of bone mass. *Br J Radiol*. 1978;51(606):428–431.
17 Ma D, Jones G. The association between bone mineral density, metacarpal morphometry, and upper limb fractures in children: a population-based case-control study. *J Clin Endocrinol Metab*. 2003;88(4):1486–1491.
18 Fong DY, Cheung KM, Wong YW, et al. A population-based cohort study of 394,401 children followed for 10 years exhibits sustained effectiveness of scoliosis screening. *Spine J*. 2015;15(5):825–833.
19 Residual attention network for image classification. In: Wang F, Jiang M, Qian C, et al., eds. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
20 Illés T, Somoskeöy S. The EOS™ imaging system and its uses in daily orthopaedic practice. *Int Orthop*. 2012;36(7):1325–1331.
21 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.
22 Sarwark JF, Davis MM. Evolving recommendations for scoliosis screening: a compelling need for further research. *JAMA*. 2018;319(2):127–129.
23 Dimitrijević V, Viduka D, Šćepanović T, et al. Effects of Schroth method and core stabilization exercises on idiopathic scoliosis: a systematic review and meta-analysis. *Eur Spine J*. 2022;31(12):3500–3511.
24 Doody MM, Lonstein JE, Stovall M, Hacker DG, Luckyanov N, Land CE. Breast cancer mortality after diagnostic radiography: findings from the U.S. Scoliosis Cohort Study. *Spine (Phila Pa 1976)*. 2000;25(16):2052–2063.
25 Wang W, Shen J. Deep visual attention prediction. *IEEE Trans Image Process*. 2017;27(5):2368–2378.
26 Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. *Nat Rev Cancer*. 2022;22(2):114–126.
27 Nault ML, Beauséjour M, Roy-Beaudry M, et al. A predictive model of progression for adolescent idiopathic scoliosis based on 3D spine parameters at first visit. *Spine (Phila Pa 1976)*. 2020;45(9):605–611.
28 Lam TP, Hung VW, Yeung HY, et al. Quantitative ultrasound for predicting curve progression in adolescent idiopathic scoliosis: a prospective cohort study of 294 cases followed-up beyond skeletal maturity. *Ultrasound Med Biol*. 2013;39(3):381–387.
29 Ogura Y, Kou I, Miura S, et al. A functional SNP in BNC2 is associated with adolescent idiopathic scoliosis. *Am J Hum Genet*. 2015;97(2):337–342.
30 Kou I, Takahashi Y, Johnson TA, et al. Genetic variants in GPR126 are associated with adolescent idiopathic scoliosis. *Nat Genet*. 2013;45(6):676–679.
31 Khanshour AM, Kou I, Fan Y, et al. Genome-wide meta-analysis and replication studies in multiple ethnicities identify novel adolescent idiopathic scoliosis susceptibility loci. *Hum Mol Genet*. 2018;27(22):3986–3998.
32 Ramnitz MS, Lodish MB. Racial disparities in pubertal development. *Semin Reprod Med*. 2013;31(5):333–339.
33 Karol LA. The natural history of early-onset scoliosis. *J Pediatr Orthop*. 2019;39(Issue 6, Supplement 1 Suppl 1):S38–s43.