Conquering the Cobb Angle: A Deep Learning Algorithm for Automated, Hardware-Invariant Measurement of Cobb Angle on Radiographs in Patients with Scoliosis

Abhinav Suri, BA, MPH • Sisi Tang, BA[1] • Daniel Kargilis, BA[2] • Elena Taratuta, MD • Bruce J. Kneeland, MD • Grace Choi, BA • Alisha Agarwal, BA • Nancy Anabaraonye, BA • Winnie Xu, BA • James B. Parente, BA • Ashley Terry, BA • Anita Kalluri, BA • Katie Song, BA • Chamith S. Rajapakse, PhD

Scoliosis is a disease estimated to affect more than 8% of adults in the United States. It is diagnosed with use of radiography by means of manual measurement of the angle between maximally tilted vertebrae on a radiograph (ie, the Cobb angle). However, these measurements are time-consuming, limiting their use in scoliosis surgical planning and postoperative monitoring. In this retrospective study, a pipeline (using the SpineTK architecture) was developed that was trained, validated, and tested on 1310 anterior-posterior images obtained with a low-dose stereo-radiographic scanning system and radiographs obtained in patients with suspected scoliosis to automatically measure Cobb angles. The images were obtained at six centers (2005–2020). The algorithm measured Cobb angles on hold-out internal ($n = 460$) and external ($n = 161$) test sets with less than 2° error (intraclass correlation coefficient, 0.96) compared with ground truth measurements by two experienced radiologists. Measurements, produced in less than 0.5 second, did not differ significantly ($P = .05$ cutoff) from ground truth measurements, regardless of the presence or absence of surgical hardware ($P = .80$), age ($P = .58$), sex ($P = .83$), body mass index ($P = .63$), scoliosis severity ($P = .44$), or image type (low-dose stereoradiographic image vs radiograph; $P = .51$) in the patient. These findings suggest that the algorithm is highly robust across different clinical characteristics. Given its automated, rapid, and accurate measurements, this network may be used for monitoring scoliosis progression in patients.

Supplemental material is available for this article.

©RSNA, 2023

Scoliosis is a common deformation of the spine, estimated to affect 2%–4% of adolescents and more than 8% of adults in the United States (1,2). The deformation is caused by lateral curvature (or multiple curvatures) of the spine and can result in chronic back pain, respiratory distress or infection, and increased risk of fractures (3–5). Treatments include both surgical and nonoperative interventions (6,7).

Currently, a physical examination is used to preliminarily screen individuals for radiologic evaluation if scoliosis is suspected (8). Upon referral to a radiologist, patients undergo standing anterior-posterior radiography for spinal curvature quantification with use of the Cobb angle, which is the angle between the maximally tilted superior and inferior vertebral endplates (9). A Cobb angle of 10° or greater indicates scoliosis and represents the current reference standard in diagnosis (9). Additionally, rapid changes in Cobb angle may indicate the urgent need for surgical intervention. Individuals can be referred for surgical correction of scoliosis based on Cobb angle grading and a number of other clinical factors (10). Following surgery, postoperative patient management includes monitoring the Cobb angle to determine whether the rate of scoliotic curvature development decreases or stops (11). Furthermore, Cobb angles

have been shown to have intra- and interobserver reliability values of ±9.6° and ±11.8°, respectively (12). With the aid of computerized programs (requiring manual input), interclass correlation coefficients have been reported to be as high as 0.96 when measuring a coronal Cobb angle (13).

Considering the large volume of radiographs obtained and high associated human effort for annotation, tools for automated Cobb angle measurements pose potential clinical benefit. Studies on accurate automated Cobb angle measurements have been performed, but prior implementations have included the use of precropped images and excluded or did not reference patients with hardware present, limiting clinical applicability (14–23). Given that implanted hardware is often present in patients with prior spinal alignment surgeries, an ideal solution to automated Cobb angle measurements would be both hardware-invariant and readily applicable to baseline radiographs for screening.

Herein, we report the development and application of a region-based neural network architecture that is hardware-invariant (ie, capable of measuring Cobb angles even if extensive hardware is present). We also report the performance of the network across a diverse patient sample with varying clinical characteristics. Last, we compare this

Summary
The proposed neural network automatically and accurately measured the Cobb angle on radiographs with various imaging conditions (surgical hardware obstructing the spine, inconsistent region imaged) in patients with scoliosis.

Key Points
- The SpineTK neural network architecture measured Cobb angles on radiographs in under 0.5 second with less than 2° mean error compared with ground truth measurements made by radiologists, each with 20 years of experience.
- There was no evidence of a difference in network measurement error after stratifying images based on presence or absence of surgical hardware and different patient characteristics (age, sex, body mass index, and scoliosis severity; all $P > .05$).
- This network can be used to monitor scoliosis progression in patients and highlights the utility of landmark detection networks (such as SpineTK) in musculoskeletal radiology applications.

Keywords
Cobb Angle, Convolutional Neural Network, Deep Learning Algorithms, Pediatrics, Machine Learning Algorithms, Scoliosis, Spine

network's Cobb angle measurements to measurements made by two highly experienced radiologists.

Materials and Methods

Deep Learning System Overview

A neural network architecture was designed to perform two tasks: *(a)* detect vertebral bodies and *(b)* locate the four corner points on each vertebral body (corresponding to the top and bottom endplates). The vertebral body bounding boxes could then be fit to a 6th degree polynomial (24). Vertebral bodies closest to the inflection points of the curve were considered end points of the curve in Cobb angle calculations (the endplates of these vertebrae form the Cobb angle). This network architecture, called SpineTK, was studied for use in the task of vertebral deformity quantification and was retrained for the purpose Cobb angle measurement (25,26). An overview of the architecture of the system can be found in Figure 1 along with the full pipeline for Cobb angle measurement.

### Dataset and Annotation Description

In this retrospective study, radiographs were used to train, evaluate, and test our network. Institutional review board approval was obtained for the study, and imaging studies were downloaded and analyzed in a Health Insurance Portability and Accountability Act–compliant manner. Low-dose stereoradiographic scanning system images (hereafter, EOS images) (EOS Imaging) were obtained from 2005 to 2020 at the University of Pennsylvania radiology centers (n = 5) using associated anterior-posterior EOS scanners. The images were obtained in individuals with indications shown in the flow diagram in Figure S1. All radiographs were acquired in different individuals. The

network was trained and validated on EOS images (*n* = 509 for training, *n* = 180 for validation). For the testing dataset, we used 460 EOS images and an additional 161 radiographs from an external public dataset (27); we added the external test dataset to include radiographs (after the network was trained). The resulting split is noted in Table 1 (47.7% reserved for testing data). Sample size for this study was chosen based on all images available in our laboratory (with the minimum number of samples in the test set calculated to be 310, assuming the following noninferiority conditions: $P = .05$, $1 – \beta = 0.90$, SD = 1.5, $d = 0.5$). The 161 radiographs in the external dataset came from the Jordan University of Science and Technology (27 images were excluded from original dataset because they already had written or digital Cobb angle markings [*n* = 20] or were cell phone images of radiographs [*n* = 7]). Final average image dimensions for each dataset are reported in Table 1. Information about imaging conditions for the radiographs is limited, except that individuals were imaged when standing and that digital stitching was used to image the full spine.

Each source Digital Imaging and Communications in Medicine image from the EOS scans was loaded into memory (using Pydicom; *https://pydicom.github.io*) and to the 0–255 range (see Appendix S1, section 1, for details). All images were resized to 800 × 800 pixels, as was used in the original SpineTK publication (25), to avoid retraining the backbone feature extractor network. These images were then split into training (*n* = 509), validation (*n* = 180), and internal testing (*n* = 460) datasets. Radiographs (*n* = 161) were already formatted appropriately and were added as a separate external dataset (621 total images in the testing dataset; these images were not used to train the network). Training and validation data were used in a fivefold cross-validation scheme to determine the best fold for final testing (hyperparameters chosen by means of grid search). Information on aggregate dataset characteristics can be found in Table 1 (training and validation characteristics reported on the fold used to ultimately test the network). To increase the size of the network and expand the set of conditions the network was exposed to, the training data were expanded by a factor of 10 with random augmentations (see Appendix S1, section 1, for list; there were 5090 images after augmentation).

Four annotators (G.C., A.A., N.A., and W.X., research assistants with hundreds to thousands of hours of musculoskeletal annotation experience) and two musculoskeletal radiologists (E.T. and B.J.K., each with at least 20 years of experience) produced Cobb angle measurements and key point annotations for each image. Further details about annotation procedures can be found in Appendix S1, section 1. Results from a reproducibility study (evaluating agreement between annotator and radiologist measurements) can be found in Appendix S1, section 2. This substudy showed that the agreement between annotators and experienced radiologists was 0.94 (using a two-way random mixed effects model), indicating a high degree of agreement between the two annotator types (intraclass correlation coefficient >0.90 indicates excellent agreement) (28). All test set annotations made by the annotators were checked and adjusted (if needed) by the expert radiologists (43 adjusted of 621 images).

SpineTK Network (simplified)
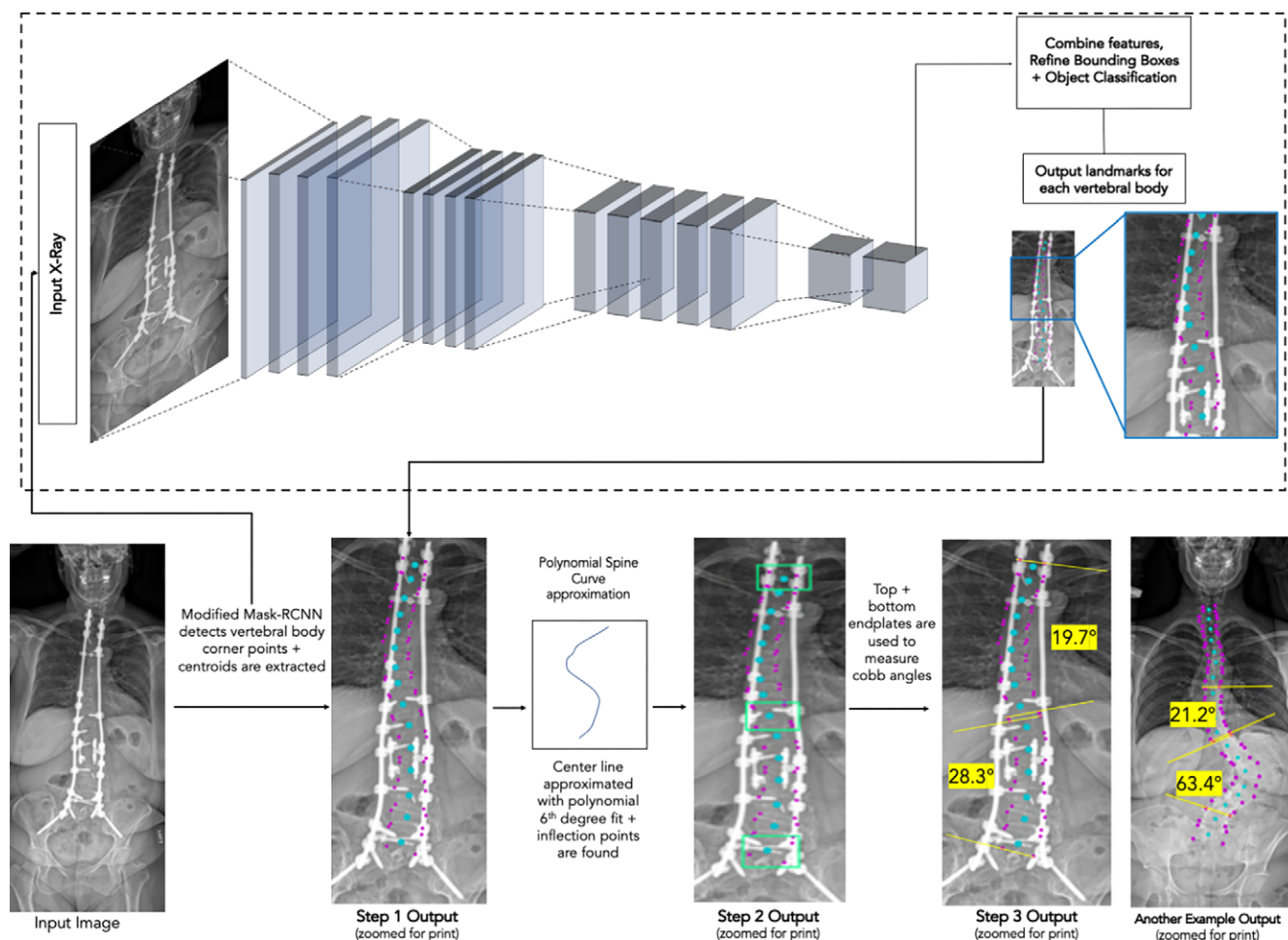
Figure 1: Overview of network and Cobb angle pipeline. Starting at the bottom left: An anteroposterior radiographic image is input into the algorithm. The image is then processed by the SpineTK neural network (modified mask–region-based convolutional neural network [Mask-RCNN] architecture). Further explanation of the structure of the network (convolutions, etc), along with details on how to calculate Cobb angles from the vertebral body corner points, can be found in Appendix S1. Vertebral body corner points and centroids can then be located (under "step 1 output"). The centroids are used to fit a curve, which is then used to find the points of infection ("step 2 output"). The endplates of vertebrae with centroids closest to the points of inflection are used to measure the Cobb angle ("step 3 output" and "another example output").

Neural network training.— The SpineTK network architecture was trained on a Google CoLab instance using the aforementioned training data in a fivefold cross-validation scheme (25). To prevent overfitting, early stopping was used to cut short training time only if the change in loss at 100 epoch intervals was 10% or less (final epoch count on fold used to train network = 2100). Network hyperparameters (batch size, number of images generated per batch) were also tuned (with use of grid search) to minimize the total loss in the network when averaged across the fivefold validation loss. Further details about the network design can be found in Appendix S1, section 3. Additionally, training and validation loss curves on the best fold (defined as the fold with the lowest total network loss) are provided in Figure S2. The full code for all trained use cases of the network is available at the following repository with documentation: *https://github.com/abhisuri97/SpineTK*.

Statistical analysis.— The primary metric of evaluation on the holdout test set was the difference between the Cobb angle pre-dicted by the network versus the angle found by human anno-tators. Additionally, we reported the intraclass correlation coef-ficient and coefficient of variation by using the two-way random mixed-effects model (intraclass correlation coefficient >0.90 is deemed excellent agreement; coefficient of variation values closer to 0 indicate a high degree of agreement) (28). We also stratified the angle measurement error metric by age category (by decade), sex, body mass index (BMI), hardware presence (yes or no), and severity of scoliosis. Furthermore, we reported Cobb angle measurements in comparison with those generated by the mus-culoskeletal radiologist annotator (radiologist expert measure-ment) to measure the network's ability to match an experienced radiologist's measurements. Significance of the error differences when stratifying by clinical characteristics were reported using a standard two-sample *t* test (in the case of sex and hardware stratification) or analysis of variance (in the case of age category, severity of scoliosis, and BMI), with the null hypothesis that the difference in angle measurements between groups is not signifi-cant. *P* < .05 was considered to indicate a statistically significant

Table 1: Demographic Characteristics of the Dataset

Characteristic	Training	Validation	Internal Test Set	External Test Set	Overall
			Dataset		
No. of images*	509 (38.9)	180 (13.7)	460 (35.1)	161 (12.3)	1310
AP EOS image dimensions[†]					
Height (mm)	768.1 ± 82.3	780.5 ± 63.2	780.3 ± 71.8	NA	777.8 ± 69.1
Width (mm)	335.4 ± 18.3	310.2 ± 21.6	332.2 ± 13.9	NA	330.8 ± 20.3
Imaging modality					
No. of EOS scans	509	180	460	0	1149
No. of radiographs	0	0	0	161	161
No. of female patients	325 (63.9)	115 (63.9)	314 (68.3)	121 (75.2)	875 (66.8)
No. of patients with hardware	297 (58.3)	117 (65.0)	253 (55.0)	35 (21.7)	702 (53.6)
Spinal level					
Lumbar	129 (25.3)	33 (18.3)	114 (24.8)	17 (10.6)	293 (22.3)
Thoracic	148 (29.1)	48 (26.7)	113 (24.6)	18 (11.2)	327 (24.9)
Cervical	20 (3.9)	36 (20.0)	26 (5.7)	0 (0)	82 (6.3)
No. of fused levels[†]	5.4 ± 2.3	5.6 ± 2.8	6.1 ± 3.9	5.1 ± 3.4	5.5 ± 2.9
Age group					
Overall (y)[‡]	59.3 ± 16.2 (12–93)	54.6 ± 15.9 (11–89)	58.3 ± 25.1 (3–94)	16.3 ± 4.1 (5–24)	54.3 ± 20.7 (3–94)
≤20 years	10 (2.0)	20 (11.1)	50 (10.9)	151 (93.8)	231 (17.6)
21–30 years	22 (4.3)	17 (9.4)	45 (9.8)	10 (6.2)	94 (7.2)
31–40 years	28 (5.5)	15 (8.3)	39 (8.5)	0	82 (6.3)
41–50 years	44 (8.6)	13 (7.2)	46 (10.0)	0	103 (7.9)
51–60 years	117 (23.0)	26 (14.4)	72 (15.7)	0	215 (16.4)
61–70 years	157 (30.8)	44 (24.4)	114 (24.8)	0	315 (24.0)
>70 years	131 (25.7)	45 (25.0)	94 (20.4)	0	270 (20.6)
Body mass index[§]					
Overall[†]	28.1 ± 8.2	28.9 ± 4.6	28.3 ± 4.9	…	28.1 ± 6.9
<18.5	8 (1.6)	5 (2.8)	37 (8.0)	…	50 (3.8)
18.5 up to 25.0	161 (31.6)	54 (30.0)	171 (37.2)	…	386 (29.5)
25.0 up to 30.0	187 (36.7)	55 (30.6)	153 (33.3)	…	395 (30.2)
≥30.0	153 (30.1)	66 (36.7)	99 (21.5)	…	318 (24.3)
Ground truth major Cobb angle (degrees)[†]	22.9 ± 15.4	28.3 ± 21.0	25.8 ± 18.3		25.9 ± 19.7
Severity of scoliosis					
No scoliosis (<10°)	109 (21.4)	34 (18.9)	80 (17.4)	30 (18.6)	253 (19.3)
Mild scoliosis (10° up to 20°)	131 (25.7)	59 (32.8)	108 (23.5)	28 (17.4)	326 (24.9)
Moderate scoliosis (20° up to 40°)	176 (34.6)	48 (26.7)	179 (38.9)	59 (36.6)	462 (35.3)
Severe scoliosis (≥40°)	93 (18.3)	39 (21.7)	93 (20.2)	44 (27.3)	269 (20.5)

Note.—Unless otherwise specified, data are numbers of patients, with percentages in parentheses. For the training, validation, test, and overall datasets, we provide the full breakdown of characteristics available from our records and image parameters. For hardware data, we further present in which region of the spine the hardware was predominantly present (eg, rod that comprises L5–T12 is counted as lumbar). For continuous variables (age, body mass index [BMI], ground truth major Cobb angle), we binned individuals in line with traditionally used clinical markers (eg, for BMI, the cutoff points for buckets correspond to the underweight, normal, overweight, and obese categories). Note: the public radiograph dataset we used did not contain corresponding information on BMI or original imaging dimensions (before conversion to JPG files); thus, demographic characteristics are only reported for EOS scans. Additionally, the term "radiograph" is equivalent to plain film. AP = anterior-posterior, NA = not applicable.

* Percentages were calculated based on the total number of images in the overall dataset.

[†] Data are means ± SDs.

[‡] Data are means ± SDs, with ranges in parentheses.

[§] BMI was calculated as patient weight in kilograms divided by patient height in meters squared.

difference. Parametric tests were appropriate for use because of the normal distribution of errors (Shapiro-Wilk test [$P < .05$ suggests lack of normality], $P = .31$). The statistical tests were calculated using the statsmodels package (version 0.11.0; *https://www.statsmodels.org/stable/release/version0.11.html*) in Python (29).

Results

Cobb angles were measured by the algorithm on the test set of 621 images (internal: $n = 460$; external: $n = 161$); the mean patient age was 58.3 years ± 25.1 (SD) (68.3% female) in the internal test set and 16.3 years ± 4.1 (75.1% female) in the external test set. Over the entire test set, the Cobb angle measurement error (annotator-measured Cobb angle minus algorithm-measured Cobb angle) was quantified on the major Cobb angle (because it is the one used to determine a clinical diagnosis of scoliosis). The mean absolute Cobb angle measurement error was found to be 1.16° ± 0.90. The mean squared Cobb angle measurement error was 2.18°. Among the test set images annotated by experienced radiologists only ($n = 100$), the mean absolute Cobb angle measurement error was 1.45° ± 1.19, and the mean squared Cobb angle measurement error was 3.54°. Overall agreement between the algorithm predictions and the radiologist raters was 0.96 (95% CI: 0.93, 0.99), with a coefficient of variation of 8.29%. Of note, the network and annotators picked the same end points 92.51% of the time (1149 of 1242 vertebrae endplates).

To quantify whether the Cobb angle measurements produced by the algorithm vary by patient clinical characteristics, we stratified Cobb angle measurement errors by each of the following: presence of large amount of hardware on the image (ie, surgical pins, screws, rods, pacemakers, and other hardware that obstructs part of the spine) versus absence (ie, no hardware or minor hardware, such as electrocardiogram leads), sex of the patient, BMI range, scoliosis angle severity, and age range (see Table 1 or Fig 2 for cut points). We further report results on both our internal test set from our institution and institutional partners (internal test set, all EOS images) and the external test set (separate institution, all radiographs). We found no evidence of a difference in algorithm measurement errors between the hardware (t test pooled test set: $P = .80$), sex (t test pooled test set: $P = .83$), BMI (analysis of variance pooled test set: $df = 3$; $P = .63$), scoliosis or Cobb angle severity (analysis of variance pooled test set: $df = 3$; $P = .44$), and age categories (analysis of variance pooled test set: $df = 6$; $P = .58$). We also found no evidence of a difference in image type (EOS vs radiograph) errors (t test pooled test set: $P = .51$). Table 2 shows the individual errors (mean error ± 1 SD) by category, along with individual P values calculated for the internal and external test datasets for the aforementioned characteristics. Figure 2B–2F shows the distribution of Cobb angle measurement errors. Figure 2G shows a confusion matrix of the overall accuracy of scoliosis degree classification (all categories showed greater than 90% classification accuracy, and two patients were misclassified as not having scoliosis). Per this confusion matrix analysis, the network predicted the correct scoliosis category for 96.6% of patients (600 of 621), with no false-negative interpretations (ie, no individuals with scoliosis were diagnosed as not having scoliosis). Figure 2H shows example results of major Cobb angle measurements on EOS images and radiographs.

Discussion

We presented a deep learning algorithm that can automatically produce accurate Cobb angle measurements from anterior-posterior EOS radiographs. This method had a mean absolute angle measurement error of 1.16°, showed a strong agreement with manually annotated Cobb angle measurements made by experienced radiologists (intraclass correlation coefficient, 0.96, indicating excellent agreement [28] and matching prior studies in this metric [13]), and was statistically invariant across a number of clinical characteristics (hardware presence, sex, BMI, scoliosis severity, age, and image type). Furthermore, this network was highly accurate in detecting scoliosis for all scoliosis grades (greater than 90% accuracy across all severity types). The network can produce accurate measurements in a variety of clinical conditions, such as a high amount of hardware obstructing parts of the spine, severe scoliosis, different imaging positions, and different windows of the body imaged. Last, the network's high performance on the external test set demonstrates its generalizability (the network was trained on EOS images only and still created accurate measurements on radiographs). Furthermore, this network has the ability to aid in curvature calculation for EOS systems among patients who have hardware, where EOS scan–derived angles had previously only been validated in hardware-free individuals (30).

In comparison with other models attempting automated Cobb angle measurement, our network has a number of qualitative advantages in terms of validating its clinical efficacy (Table S1). Notably, the effectiveness of an algorithm in evaluating images in individuals with hardware allows for longitudinal monitoring of patients who have undergone spinal alignment surgery or other musculoskeletal procedures. In addition, because deep learning models are subject to broad variation in performance based on subtle changes in imaging characteristics, accounting for differences in performance across clinical subgroups increases generalizability and potential for robust implementation. This represents the third application of the SpineTK network in evaluating spine-relevant metrics. Furthermore, this study specifically highlights how mask–region-based convolutional neural networks (such as SpineTK) can be applied to various landmark detection tasks and be retrained for clinical use. In our study, we retrained SpineTK on a different imaging modality (EOS imaging and radiography vs MRI, CT, and radiography) and plane of imaging (coronal vs sagittal) and showed that it can still produce accurate landmark locations relevant to the Cobb angle that match or surpass values in the existing literature and provide advantages in terms of invariance across clinical characteristics. Last, we demonstrate that the network can learn to generalize, as suggested by its high performance on data that were not represented in the training set.

Our study had limitations. In a manual analysis of EOS images with 1 SD and greater Cobb angle measurement error in hardware versus nonhardware categories, artifacts from hardware can occlude the spine on images such that the algorithm is unable to detect enough vertebral bodies to form an appropriate curve. However, it is notable that the angle measurement errors did not exceed 5° on EOS images in patients either with or without hardware. When looking at
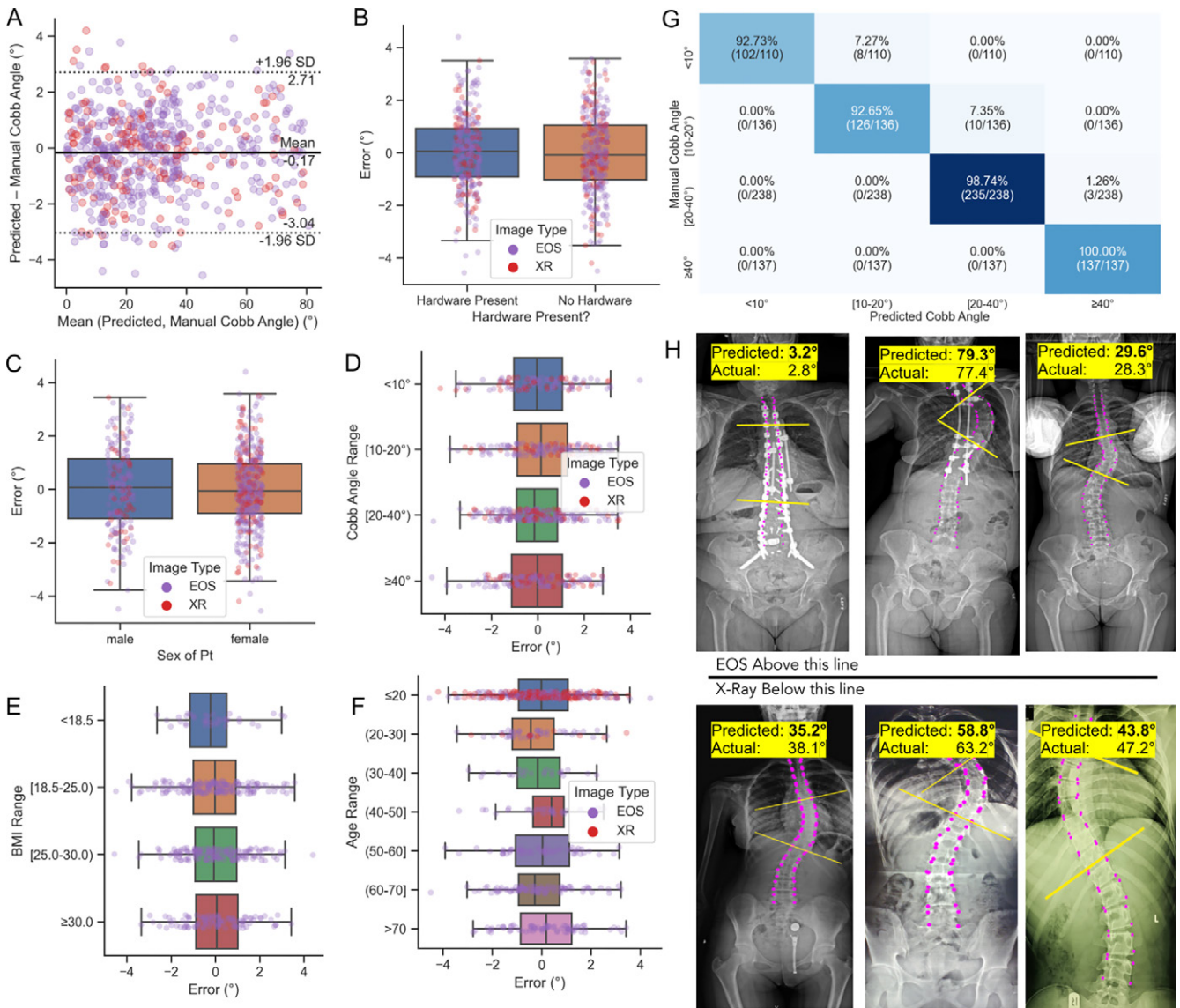
Figure 2: Main Cobb angle measurement results. **(A)** Bland-Altman plot for Cobb angle errors (y-axis shows the difference between the predicted [ie, from the neural network] and manually measured Cobb angle; x-axis shows the mean difference between the predicted and manual Cobb angle). **(B–F)** Box and whiskers plots, where red dots represent radiographs (XR) (external test set) and purple dots represent EOS images (internal test set). Boxes indicate the IQR, and whiskers represent 1.5 × IQR ± first or third quartile, as appropriate. Plots show the measurement error (difference between predicted and manually measured) stratified by **(B)** hardware status (hardware present or not on the image), **(C)** patient sex, **(D)** Cobb angle range (see Table 1 for meaning of cut points), **(E)** patient body mass index (BMI) status (see Table 1 caption for meaning of cut points), and **(F)** patient age (by decade of life). **(G)** Confusion matrix of accuracy of classification (based on cut points for scoliosis severity). Raw numbers indicate number of patients (number of correctly predicted patients in that class of scoliosis severity over true number of patients in that class of scoliosis severity). **(H)** Representative example images of predicted versus actual Cobb angle measurements for cases ranging from minor to major scoliosis with varying degrees of hardware obstruction. pt = patient.

radiographs with 1 SD and greater Cobb angle measurement error, we found that these images also had artifacts from hardware that occluded a majority of the spine (but once again, did not exceed 5° error). These findings indicate that a separate "flag" for radiologist review should be used in practice when only a few vertebral bodies are detected by the network because 5° of error can potentially push patients into a different scoliosis grading category. We also noted that endplates did not match with those chosen by annotators in 93 of 1242 endplates (7.5%) in the test set. While the exact endplate choice may not matter for Cobb angle reporting for screening, it can be a limiting factor in surgical planning, which involves

selecting appropriate vertebrae as end points of the curve. Additionally, we did not have information about the EOS or external radiograph scanning parameters, which can limit replicability in future studies. Per the EOS website (*https://www.eos-imaging.com*), scanning parameters are standardized across devices to minimize radiation exposure, which would indicate that, unless an institution has altered these parameters, the results may still be applicable. Last, while great care was made to ensure that our patient sample represented diverse clinical characteristics, we did not have a proportionally large number of EOS images available to us in younger patients (age ≤ 20 years) for use in this study (80 EOS images in patients age ≤

Table 2: Errors by Category

Category	Mean Error (±1 SD) in the Internal Test Set (degrees)	*P* Value in the Internal Test Set	Mean Error (± 1 SD) in the External Test Set (degrees)	*P* Value in the External Test Set
Hardware		.437		.447
Present	0.03 ± 1.44		–0.03 ± 1.36	
Absent	–0.07 ± 1.48		0.14 ± 1.63	
Sex		.305		.200
Female	–0.06 ± 1.40		0.16 ± 1.49	
Male	0.08 ± 1.58		–0.18 ± 1.57	
Body mass index*		.629		
<18.5	–0.10 ± 1.37			
18.5 up to 25	–0.06 ± 1.55			
25 up to 30	0.08 ± 1.44			
≥30	0.03 ± 1.40			
Cobb angle severity		.275		.435
<10°	0.11 ± 1.55		–0.16 ± 1.66	
10° up to 20°	0.15 ± 1.49		0.25 ± 1.66	
20° up to 40°	–0.06 ± 1.39		–0.06 ± 1.39	
≥40°	–0.22 ± 1.48		0.33 ± 1.41	
Age		.436		.683
≤20 years	–0.02 ± 1.53		–0.10 ± 1.61	
21–30 years	–0.31 ± 1.27		–0.19 ± 1.21	
31–40 years	–0.23 ± 1.48			
41–50 years	0.27 ± 1.48			
51–60 years	–0.09 ± 1.46			
61–70 years	–0.14 ± 1.47			
>70 years	0.20 ± 1.44			

Note.—Category-by-category errors in Cobb angle output by the neural network. Data are presented as means ± SDs. Values under the *P* value column come from the analyses described in the main text (either *t* test or analysis of variance). $P < .05$ is considered indicative of statistically significant difference.
* Body mass index was calculated as patient weight in kilograms divided by patient height in meters squared.

20 years population). This could lead our algorithm to be biased toward individuals with lower bone mineral density and could have caused the network to learn on density-dependent features that are more prevalent in older individuals. The effect of bone density could have also led our system to learn on features such as fractured or deformed vertebral bodies. However, we used radiographs from a public dataset as an external test set to increase the representation of younger patients. The results indicated that the network was able to measure Cobb angles accurately on these external data even though the network did not have many training images in this age range. Additionally, complete racial demographics for our patient sample were not available for this study, limiting our ability to demonstrate robust performance across racial groups.

Overall, we have demonstrated that the present neural network architecture can produce highly accurate Cobb angle measurements in a rapid manner. Furthermore, outputs from this network have been shown to produce consistent measurements regardless of different clinical characteristics, including the presence of hardware, which, to our knowledge, has not been evaluated in prior studies. The network architecture used in this study has been demonstrated to be highly adaptable and can be applied to a variety of object detection– and key point detection–based tasks in musculoskeletal radiology. With regard to the specific task of Cobb angle measurement automation, this network offers the opportunity to quickly analyze images accurately for Cobb angle measurements, providing the opportunity for trending of Cobb angles over multiple scans without additional work. Future studies should further validate the network's performance for clinical implementation by including additional adolescent individuals, evaluating performance across individuals of different races and bone density statuses, and determining the network's use in a live clinical setting.

References

1. Kebaish KM, Neubauer PR, Voros GD, Khoshnevisan MA, Skolasky RL. Scoliosis in adults aged forty years and older: prevalence and relationship to age, race, and gender. Spine 2011;36(9):731–736.

2. Labelle H, Richards SB, De Kleuver M, et al. Screening for adolescent idiopathic scoliosis: an information statement by the Scoliosis Research Society international task force. Scoliosis 2013;8(1):17.

3. Weiss HR. Rehabilitation of adolescent patients with scoliosis—what do we know? A review of the literature. Pediatr Rehabil 2003;6(3-4):183–194.

4. Warren MP, Brooks-Gunn J, Hamilton LH, Warren LF, Hamilton WG. Scoliosis and fractures in young ballet dancers. Relation to delayed menarche and secondary amenorrhea. N Engl J Med 1986;314(21):1348–1353.

5. Li XF, Li H, Liu ZD, Dai LY. Low bone mineral status in adolescent idiopathic scoliosis. Eur Spine J 2008;17(11):1431–1440.

6. Asher MA, Burton DC. Adolescent idiopathic scoliosis: natural history and long term treatment effects. Scoliosis 2006;1(1):2.

7. Diebo BG, Shah NV, Boachie-Adjei O, et al. Adult spinal deformity. Lancet 2019;394(10193):160–172.

8. Janicki JA, Alman B. Scoliosis: review of diagnosis and treatment. Paediatr Child Health 2007;12(9):771–776.

9. Kim H, Kim HS, Moon ES, et al. Scoliosis imaging: what radiologists should know. RadioGraphics 2010;30(7):1823–1842.

10. Fujishiro T, Boissière L, Cawley DT, et al. Adult spinal deformity surgical decision-making score. Part 2: development and validation of a scoring system to guide the selection of treatment modalities for patients above 40 years with adult spinal deformity. Eur Spine J 2020;29(1):45–53.

11. Mac-Thiong JM, Remondino R, Joncas J, Parent S, Labelle H. Long-term follow-up after surgical treatment of adolescent idiopathic scoliosis using high-density pedicle screw constructs: is 5-year routine visit required? Eur Spine J 2019;28(6):1296–1300.

12. Loder RT, Urquhart A, Steen H, et al. Variability in Cobb angle measurements in children with congenital scoliosis. J Bone Joint Surg Br 1995;77(5):768–770.

13. Wu W, Liang J, Du Y, et al. Reliability and reproducibility analysis of the Cobb angle and assessing sagittal plane by computer-assisted and manual measurement tools. BMC Musculoskelet Disord 2014;15(1):33.

14. Horng MH, Kuok CP, Fu MJ, Lin CJ, Sun YN. Cobb angle measurement of spine from x-ray images using convolutional neural network. Comput Math Methods Med 2019;2019:6357171.

15. Caesarendra W, Rahmaniar W, Mathew J, Thien A. Automated Cobb angle measurement for adolescent idiopathic scoliosis using convolutional neural network. Diagnostics 2022;12(2):396.

16. Wu C, Meng G, Lian J, et al. A multi-stage ensemble network system to diagnose adolescent idiopathic scoliosis. Eur Radiol 2022;32(9):5880–5889.

17. Tu Y, Wang N, Tong F, Chen H. Automatic measurement algorithm of scoliosis Cobb angle based on deep learning. J Phys Conf Ser 2019;1187(4):042100.

18. Pan Y, Chen Q, Chen T, et al. Evaluation of a computer-aided method for measuring the Cobb angle on chest X-rays. Eur Spine J 2019;28(12):3035–3043.

19. Yao Y, Yu W, Gao Y, et al. W-Transformer: accurate Cobb angles estimation by using a transformer-based hybrid structure. Med Phys 2022;49(5):3246–3262.

20. Liu J, Yuan C, Sun X, Sun L, Dong H, Peng Y. The measurement of Cobb angle based on spine X-ray images using multi-scale convolutional neural network. Phys Eng Sci Med 2021;44(3):809–821.

21. Ha AY, Do BH, Bartret AL, et al. Automating scoliosis measurements in radiographic studies with machine learning: comparing artificial intelligence and clinical reports. J Digit Imaging 2022;35(3):524–533.

22. Alharbi RH, Alshaye MB, Alkanhal MM, Alharbi NM, Alzahrani MA, Alrehaili OA. Deep learning based algorithm for automatic scoliosis angle measurement. In: 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, March 19–21, 2020. IEEE, 2020; 1–5.

23. Fu X, Yang G, Zhang K, Xu N, Wu J. An automated estimator for Cobb angle measurement using multi-task networks. Neural Comput Appl 2021;33(10):4755–4761.

24. Sardjono TA, Wilkinson MH, Veldhuizen AG, van Ooijen PM, Purnama KE, Verkerke GJ. Automatic Cobb angle determination from radiographic images. Spine 2013;38(20):E1256–E1262.

25. Suri A, Jones BC, Ng G, et al. Vertebral deformity measurements at MRI, CT, and radiography using deep learning. Radiol Artif Intell 2021;4(1):e210015.

26. Suri A, Jones BC, Ng G, et al. A deep learning system for automated, multi-modality 2D segmentation of vertebral bodies and intervertebral discs. Bone 2021;149:115972.

27. Fraiwan M, Audat Z, Manasreh T. A dataset of scoliosis, spondylolisthesis, and normal vertebrae X-ray images. Published January 17, 2022. Accessed January 20, 2023.

28. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15(2):155–163. [Published correction appears in J Chiropr Med 2017;16(4):346.]

29. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. In: Proceedings of the 9th Python in Science Conference, 2010; 92–96.

30. Chung N, Cheng YH, Po HL, et al. Spinal phantom comparability study of Cobb angle measurement of scoliosis using digital radiographic imaging. J Orthop Translat 2018;15:81–90.