IEEE Journal of Translational Engineering in Health and Medicine

Received 10 July 2023; revised 21 September 2023; accepted 7 November 2023.
Date of publication 14 November 2023; date of current version 24 November 2023.

Digital Object Identifier 10.1109/JTEHM.2023.3332618

Applying Machine Learning and Point-Set Registration to Automatically Measure the Severity of Spinal Curvature on Radiographs

JASON WONG, MAREK REFORMAT, (Member, IEEE), AND EDMOND LOU, (Member, IEEE)

Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada

CORRESPONDING AUTHOR: J. WONG (jcwong2@ualberta.ca)

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and in part by the Stollery Children's Hospital Foundation through the Women and Children's Health Research Institute.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Alberta Health Research Ethics Board under Application No. Pro00102044.

ABSTRACT *Objective:* Measuring the severity of the lateral spinal curvature, or Cobb angle, is critical for monitoring and making treatment decisions for children with adolescent idiopathic scoliosis (AIS). However, manual measurement is time-consuming and subject to human error. Therefore, clinicians seek an automated measurement method to streamline workflow and improve accuracy. This paper reports on a novel machine learning algorithm of cascaded convolutional neural networks (CNN) to measure the Cobb angle on spinal radiographs automatically. *Methods:*The developed method consisted of spinal column segmentation using a CNN, vertebra localization and segmentation using iterative vertebra body location coupled with another CNN, point-set registration to correct vertebra segmentations, and Cobb angle measurement using the final segmentations. Measurement performance was evaluated with the circular mean absolute error (CMAE) and percentage within clinical acceptance ($\leq 5°$) between automatic and manual measurements. Analysis was separated by curve severity to identify any potential systematic biases using independent samples Student's t-tests. *Results:* The method detected 346 of the 352 manually measured Cobb angles (98%), with a CMAE of 2.8° and 91% of measurements within the 5° clinical acceptance. No statistically significant differences were found between the CMAEs of mild ($<25°$), moderate (25°-45°), and severe ($\geq 45°$) groups. The average measurement time per radiograph was 17.7 ± 10.2s, improving upon the estimated average of 30s it takes an experienced rater to measure. Additionally, the algorithm outputs segmentations with the measurement, allowing clinicians to interpret measurement results. *Discussion/Conclusion:*The developed method measured Cobb angles on radiographs automatically with high accuracy, quick measurement time, and interpretability, suggesting clinical feasibility.

INDEX TERMS Convolutional neural network, point-set registration, machine learning, radiograph, scoliosis.

***Clinical and Translational Impact Statement*—**Implementing the developed method could allow for quick and robust Cobb angle measurements on radiographs to expedite clinical workflow and advise adolescent idiopathic scoliosis (AIS) treatment diagnosis.

I. INTRODUCTION

SCOLIOSIS is a three-dimensional spinal disorder, where the spine is characterized by lateral curvature and axial vertebral rotation. Adolescent idiopathic scoliosis (AIS) is the most common type of scoliosis and occurs in approximately 3% of adolescents [1]. Girls have a higher chance of developing more severe curves. If left untreated, AIS can result in visible deformity, cardiopulmonary compromise, and back pain [1]. Determining the appropriate treatment for AIS involves routine imaging of the spine with a posteroanterior (PA) radiograph and measuring the severity of the lateral curvature using the Cobb angle. The Cobb angle is

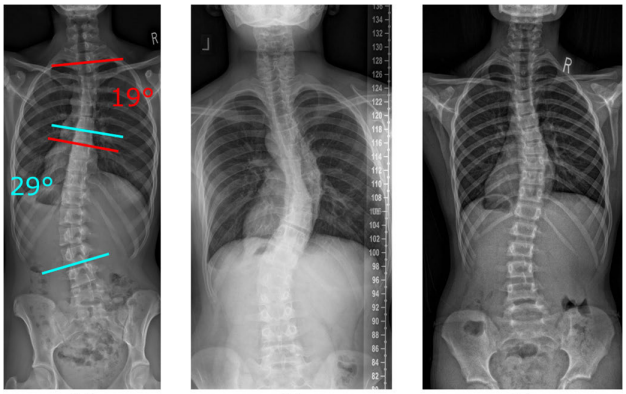


FIGURE 1. **(a) Cobb angle measurement of 19° main thoracic (red) and 29° thoracolumbar (cyan) curves; (b) A conventional x-ray PA radiograph; (c) A low dose radiation (EOS system) PA radiograph.**

the gold standard for quantifying the severity of the spinal curvature, which is measured by first identifying the pairs of most slanted vertebrae with opposing tilt angles [2]. The difference between the angles of the upper endplate of the superior most tilted and the lower endplate of the inferior most tilted vertebrae is then calculated. The Cobb angles measured on a child with AIS are depicted in **Fig. 1a**, in which the child has a double curve. A minimum Cobb angle of 10° is required for a child to be diagnosed with AIS [1]. **Fig. 1b** and **1c** show the PA radiographs obtained from a conventional digital and low dose radiation x-ray system (EOS Imaging Inc, Paris, France), respectively.

Treatment options for children with AIS are typically based on whether they meet certain Cobb angle thresholds [1]. If the adolescent exhibits curve progression, an increase of more than 6° in the Cobb angle between consecutive visits (typically six months apart), treatment may change from observation to bracing and/or exercise or from bracing and/or exercise to surgery [3]. Obtaining accurate and reliable measurements is therefore crucial to AIS treatment prescription and evaluation of treatment outcomes. Additionally, some clinics have a large number of AIS patient visits per day combined with regular clinical visits (2-3 times per year), and so measuring the Cobb angle quickly and accurately is highly desired by clinicians. An automatic measurement method has been widely sought to minimize measurement error, reduce clinician workload, and improve measurement reliability [4]. However, some of the major concerns that clinicians have with using machine learning algorithms for automation are medical liability due to algorithmic error and skepticism in a 'black box' diagnosis [5]. Consequently, the automatic measurement algorithm must also fulfill a high interpretability criterion to give clinicians confidence in the final automated diagnosis and the option to easily correct the output if they see something wrong with it.

Because of the outlined benefits, some groups have tackled the problem of automating Cobb angle measurement on PA radiographs. Many of them use a convolutional neural network (CNN) based approach and achieved accurate results [6], [7], [8]. However, these methods suffer from either testing on only a small subset in the wide range of curve severities present in the AIS population, or offering little interpretability on how the method measured the Cobb angles. Many other groups have outlined their own automatic measurement methods, but they do not obtain comparable measurement accuracies [9], [10], [11], [12], [13].

Our group [14] previously reported on a method that achieved a circular mean absolute error (CMAE) and standard deviation of circular absolute errors (SD) of 2.8°±2.8° and 88% of its Cobb angle measurements within clinical acceptance (≤5°) when compared with manual measurements. However, this was only tested on a 100-image test set, and there was no reported analysis of performance by curve severity. Additionally, the method took 90±41 seconds on average to measure per radiograph, which is roughly the time it takes for a less experienced rater to manually measure. More experienced clinicians can measure Cobb angles on a radiograph in approximately 30 seconds, making the previous method's runtime a barrier of entry for adoption in real scoliosis clinics. This manuscript reports the development of an improved fully automated algorithm to quickly measure the Cobb angle on PA radiographs and to display highly interpretable output images. The measurement results on an expanded test set of 200 PA spinal radiographs with a wide range of curve severities, along with an analysis on the accuracy performance, are also reported.

II. METHODS AND PROCEDURES

A. DATA

The PA radiographs of children with AIS used in this study were provided by a local scoliosis clinic. These images were taken by either a conventional digital x-ray system or the EOS system. Ethics approval for this study was granted by the University of Alberta research health ethics board (Pro00102044 – chart review). A total of 330 PA radiographs were used in this study and split into three separate groups of 110, 20, and 200 radiographs. The first group of 110 radiographs was used for training and validating a spinal column segmentation CNN. The second group of 20 radiographs was used to create 340 vertebral body images (17 relevant vertebrae per spine) for training and validating a vertebral body CNN. Finally, the last 200-radiograph group was used for Cobb angle measurement testing. Further details on the data sets used for each task are provided in their respective section below. All ground truth labels for CNN segmentation were annotated by two raters. Prior to creating the training sets, both raters labelled 10 practice spines with verification by a researcher who had over 20 years of experience in Cobb angle measurement. The 200 radiographs in the Cobb angle test set were all measured by that experienced researcher.

IEEE Journal of Translational Engineering in Health and Medicine

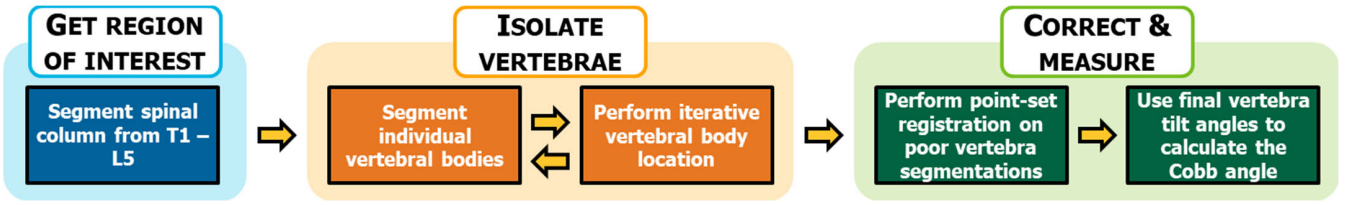


FIGURE 2. Flowchart of the overall Cobb angle measurement automation algorithm.

B. AUTOMATION ALGORITHM

A flowchart of the overall automatic Cobb angle measurement algorithm is illustrated in **Fig. 2**. The general steps consist of identifying the spinal column from the top thoracic vertebra (T1) to the bottom lumbar vertebra (L5) as a region of interest, isolating the individual vertebral bodies for segmentation, performing point-set registration to correct poor vertebral body segmentations, and finally measuring the Cobb angle using the corrected segmentations. The overall structure of the algorithm is like our previous work [14], except that the point-set registration step was added to minimize Cobb angle measurement errors from poor vertebral body segmentations. Segmentation of the spinal column and vertebral body was accomplished with CNNs trained on a Linux virtual machine on the Industry Sandbox & Artificial Intelligence Computing (ISAIC) supercomputer, using an NVIDIA Tesla V100 16GB GPU, an Intel Xeon Gold 6138 dual processor, and 64GB of RAM. All code was implemented in Python, using the TensorFlow library for CNN development as well as the *pandas* and *pingouin* libraries for statistical analysis. Any images used for CNN training were manually labelled using a user interface from the Image Processing Toolbox in MATLAB [15].

1) SPINAL COLUMN SEGMENTATION

The spinal column was segmented to narrow the PA radiograph to a region of interest for vertebral body segmentation. This step applied the same procedure of pre-processing, segmentation, and post-processing as described in [14], except that the spinal column segmentation CNN is trained differently due to further network optimizations.

A CNN architecture based on the U-net was used to segment the spinal column from the processed images [15]. The architecture of the U-net based CNN is illustrated in **Fig. 3**. In total, 110 PA radiographs were manually labelled with a continuous spinal column segment from T1, the first thoracic vertebra, to L5, the last lumbar vertebra. These images were split into a 96-image training set and 24-image validation set. Each set consisted of half conventional and half EOS radiographs. To increase the effective size of the training set, the data augmentation methods of random horizontal flipping and rotations of up to 10° were employed. The average of the manually measured Cobb angles including all curves from the 110 images was 24.6°±12.4° (range: 6° - 97°).

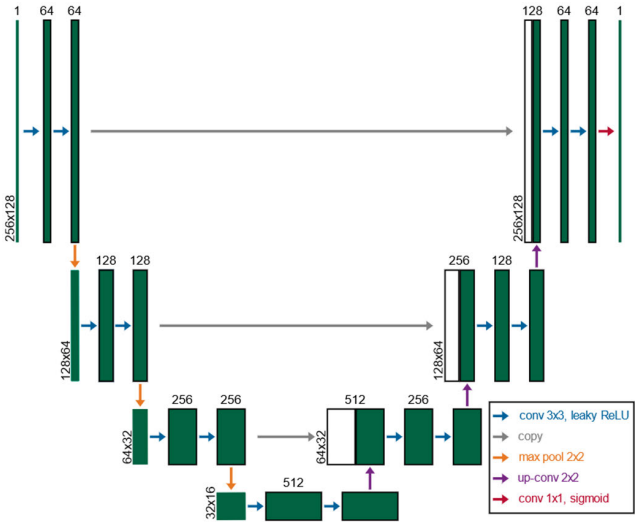


FIGURE 3. Architecture of the spinal column segmentation CNN, based on the U-net. Dark green boxes represent the feature maps and white boxes represent copied feature maps with the size of them indicated on the left of each convolutional block and the number of them above each box. The leaky rectified linear unit (ReLU) activation function after each convolutional layer had an alpha value of 0.01.

The CNN was trained using the Adam optimizer [16] and a Dice loss function [17]. Optimizing the hyperparameters of the CNN was accomplished with a grid search, which involved fitting multiple models with different hyperparameter combinations and using the hyperparameters of the highest performing network, according to which one produced the lowest Dice loss during training. The ranges of hyperparameters explored are listed in **Table 1**. The optimized spinal column CNN was trained for 1,000 epochs. A learning rate of 10^{-3} and a batch size of 2 was employed. To improve the CNN's ability to generalize, dropout [18] of 0.5 probability was performed after each pooling and upsampling layer, and batch normalization was performed before each pooling and upsampling layer. The network with the lowest validation loss during training was taken as the final optimized CNN.

Similar to the work in [14], any small stray segmentations were removed and only the largest connected component was treated as the spinal column. Then, the spinal column curve (SCC) was determined by fitting a ten-degree polynomial to the spinal column. These two steps were executed to improve performance in the subsequent vertebra isolation step. **Fig. 4**

illustrates the spinal column segmentation pipeline of an initial PA radiograph to fully segmented with the SCC labelled.

2) VERTEBRAL BODY SEGMENTATION

The vertebral body was segmented individually so that the most opposite tilted vertebra angles could be derived for Cobb angle measurement. The details of the processing steps were reported in [14], but the vertebral body segmentation CNN training differed in this study.

Segmentation of the vertebral body was achieved using the same CNN architecture as the spinal column CNN, except the initial image size was modified to be 128 × 128. A total of 340 vertebral body images from 20 subjects with AIS were manually labelled and split into 272-image training and 68-image validation sets, with each set comprising of half conventional and half EOS radiographs. Data augmentation methods of horizontal flipping, rotations up to 45°, horizontal and vertical shifts up to 10%, and zooming from 80% to 120% were employed. The average of the manually measured Cobb angles of the 20 subjects was 24.9°±10.8° (range: 7° - 54°).

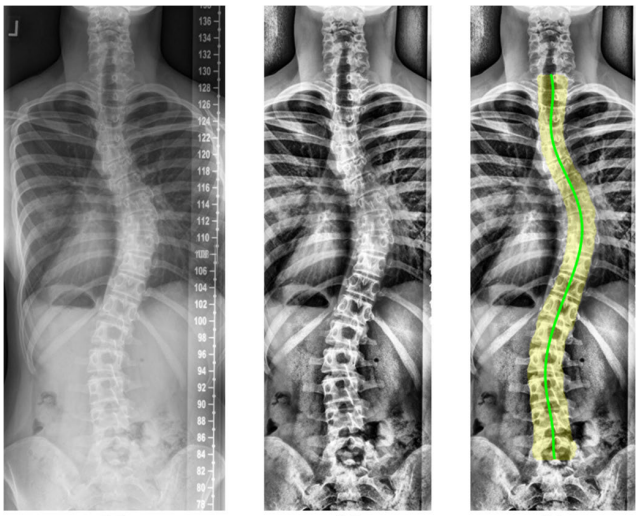
A grid search was similarly employed to optimize the hyperparameters of the vertebral body segmentation CNN with the hyperparameter ranges explored listed in **Table 1**. The optimized CNN used an Adam optimizer with a learning rate of 10^{-4} and a Dice loss function. It was trained for 1,000 epochs with a batch size of 4. A dropout of 0.125 was used after each pooling and upsampling layer. Batch normalization was not used.

Similar to the spinal column segmentation, small stray segmentations were removed from the final vertebral body segmentation and only the largest connected component was kept. **Fig. 5** illustrates the vertebral body segmentation pipeline of a cropped vertebral body image.

3) ITERATIVE VERTEBRAL BODY LOCATION

Because the vertebral body segmentation CNN was trained on square images centered on the vertebral body, a method of identifying potential vertebral body images within the PA radiograph is required for accurate vertebral body segmentation. Localizing vertebral body images to crop out for segmentation was accomplished using an iterative algorithm which was different from the procedure described in [14]. Iterative vertebral body location started by estimating a position of T12, the last thoracic vertebra, to crop out and segment. This vertebra was chosen because it typically is the clearest in a radiograph and therefore results in the clearest segmentations. First, the vertical position of T12 was estimated by calculating a ratio, r, describing its relative vertical position within the T1 to L5 spinal column. This is calculated using vertebral body heights, h, given in [19], [20], and vertebral index, j, with T1 as $j = 1$ to L5 as $j = 17$:

$$r = \frac{\sum_{j=1}^{12} h_j}{\sum_{j=1}^{17} h_j} \tag{1}$$



(a) (b) (c)

FIGURE 4. (a) An initial PA radiograph; (b) The pre-processed version of the radiograph; (c) Automatically segmented spinal column (yellow) with spinal column curve (green) overlaid on the pre-processed radiograph.

This ratio is then multiplied by the height of the segmented spinal column to obtain T12's estimated vertical position, v. To maximize the chances of obtaining a cropped image that was centred on T12, five vertical positions were chosen for cropping, with each being separated by a quarter of the spinal column width (SCW) at v. The horizontal positions were the centroids of the spinal column segmentation at each respective vertical position. The size of each cropped image was determined using the SCW at each vertical position. These cropped images were input into the vertebral body CNN and the best quality segmentation was selected.

To determine the quality of a segmentation automatically, thirteen vertebral body masks from the training set were treated as standard masks and compared with the predicted segmentation. A similarity loss was calculated, which involved extracting the contours of both the predicted segmentation and each standard mask, performing rotation and scaling on the contour of the predicted segmentation so that it was consistent with each standard mask, and then calculating the minimum distances from each point on the predicted contour to each standard contour and vice versa. A lower similarity loss corresponded to a higher quality. Let the distributions of points in the predicted and standard contour be Φ and Ω with a single point being φ and ω in each contour, respectively. The difference loss, ℓ, is calculated as shown below with N_x being the number of points in a given distribution x and M being the set of all standard masks:

$$\ell = \min_{M} \left[\sqrt{\frac{1}{N_\Phi} \sum_{i=1}^{N_\Phi} \left(\min_{\omega \in \Omega} \|\varphi_i - \omega\| \right)^2} + \sqrt{\frac{1}{N_\Omega} \sum_{i=1}^{N_\Omega} \left(\min_{\varphi \in \Phi} \|\omega_i - \varphi\| \right)^2} \right] \tag{2}$$

IEEE Journal of Translational Engineering in Health and Medicine

TABLE 1. Ranges of hyperparameters explored in the grid searches for both segmentation CNNs.

(Learning rate, batch normalization)	Batch size	Dropout	Leaky ReLU alpha
{(10^{-3}, True), (10^{-4}, False)}	{1, 2, 4, 8, 16}	{0, 0.125, 0.5}	{0, 0.01}

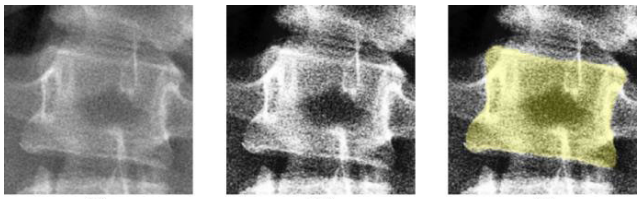


FIGURE 5. (a) An initial cropped vertebral body image; (b) The processed version of the vertebral body; (c) Automatically segmented vertebral body (yellow) overlaid on the vertebral body image.

Once T12 was set, the position of the vertebra above (T11) was estimated for segmentation. The SCW of T12 was multiplied by vertebral body height-to-width ratios to obtain the distance to travel to reach the next vertebra [19], [20]. The direction in which to travel was determined using the tangent of the SCC at T12. Crop width was determined using the SCW at the new estimated position. This cropped image was then passed to the CNN for segmentation. Checks were employed to ensure that a reasonable segmentation was achieved, such as testing that its ℓ was low and that the two segmentations were reasonably far apart. If any of these checks were failed, the cropping parameters were modified in scaling and/or translation to achieve a higher quality segmentation. To avoid cases of infinite looping where the vertebra in question can never be segmented satisfactorily, only seven segmentation attempts were allowed before the algorithm proceeded. Once the vertebral body segmentation is finalized, the procedure of cropping location estimation, vertebral body segmentation, and quality verification was repeated for the next vertebra above in an iterative procedure until the top of the spinal column segmentation was reached. If the number of vertebral bodies that was segmented to reach the top of the spinal column was different than expected, then the algorithm shifted the levels of the vertebrae appropriately. The T11 or L1 vertebra was sometimes segmented at the start instead of the T12 vertebra due to spinal structural differences in children with AIS. The algorithm then moved back to the vertebra above the initially segmented vertebra to iterate downwards until the bottom of the spinal column segmentation was reached. This vertebra was chosen as the starting point instead of the initially segmented one since there was now more information to leverage for its cropping location estimation and a higher quality segmentation for it could be achieved. This overall algorithm is illustrated in **Fig. 6**.

4) POINT-SET REGISTRATION CORRECTION

To measure the Cobb angle, the tilt angle of the minimum bounding boxes of each vertebral body segmentation is used. One limitation in the algorithm from previous work [14] was that the tilt angle of a vertebral body segmentation was sensitive to small protrusions due to the use of a bounding box to determine vertebral tilt. Consequently, a step of point-set registration correction was added to minimize the number of extraneous protrusions and ensure high quality bounding boxes that accurately reflect the tilt of the vertebral body.

Identifying whether a vertebral body required point-set registration to correct its segmentation involved checking that the tilt angle of the current bounding box, the angle formed from the top corners of the segmentation, and the angle formed from the bottom corners of the segmentation were all within a 3.5° difference from each other. The algorithm also checked if the tilt angle of the current bounding box was within 10° of the angle perpendicular to the SCC at the centroid of the vertebral body segmentation. If any of these checks failed, the vertebral body went through point-set registration to correct its segmentation.

Point-set registration was accomplished using the scaling iterative closest points algorithm (SICP) [21]. This algorithm found the affine transformation that minimized the sum of squared minimum distances between the points of the contours from the vertebral body segmentation and standard mask that produced the lowest ℓ during the iterative vertebral body location algorithm. An inlier ratio of 0.8 was used, and the SICP algorithm was repeated for a maximum number of 100 iterations unless an improvement of less than 10^{-3} was achieved between consecutive iterations. Once the SICP algorithm was complete, the registered standard mask was used for measurement instead of the segmentation.

5) COBB ANGLE MEASUREMENT

The Cobb angle was measured using the angles of the bounding boxes of the vertebral body segmentations or registered standard masks. Extracting the Cobb angles involved identifying the apices of the spinal column, the vertebrae that are most laterally shifted from the centerline of the body. The apical vertebrae therefore correspond to the points on the SCC where the angles perpendicular to the SCC are 0°. A Cobb angle was then calculated for each apex, taking the difference of the steepest opposing vertebral body tilt angles (calculated from their minimum bounding boxes). If no opposing vertebral body tilt angles were found, the apex was skipped. For each curve, the algorithm determines its Cobb angle, direction, upper end vertebra, apical vertebra, and lower end vertebra.

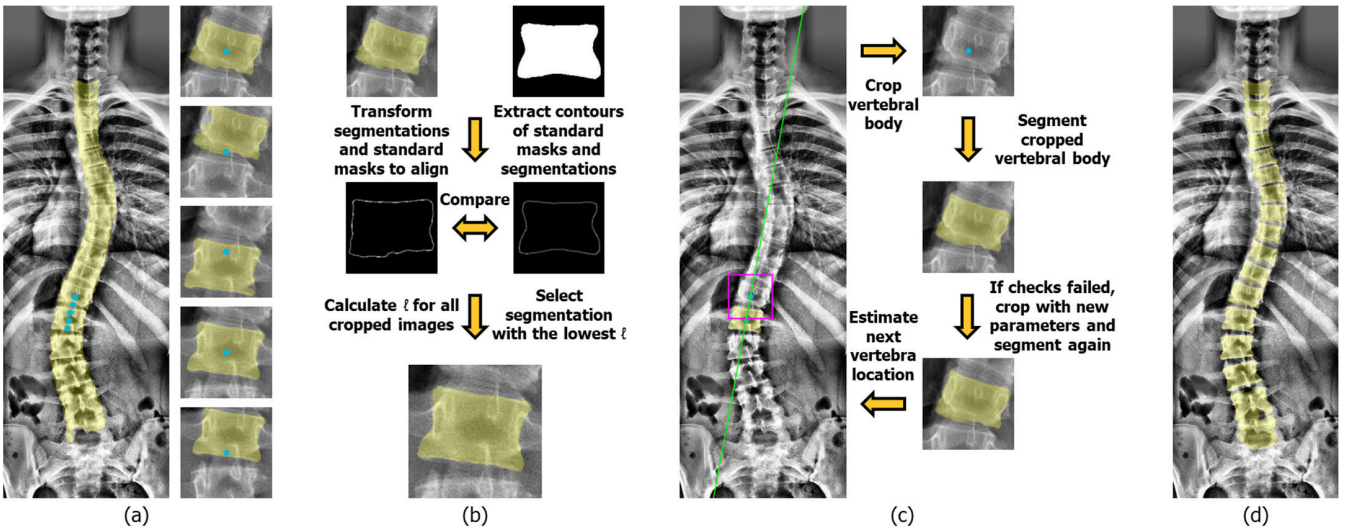


FIGURE 6. (a) Search for the starting vertebra by segmenting five cropped images at different vertical positions; (b) Quality coefficient calculation to determine the initial segmented vertebral body to use as the starting point for the iterative vertebral body location algorithm; (c) Iterative vertebral body location algorithm where the most recently segmented vertebral body is used to estimate the position of the vertebral body above and below to segment in an iterative procedure until all vertebral bodies are segmented; (d) Final result of the iterative vertebral body location algorithm where the vertebral bodies are now separated from each other.

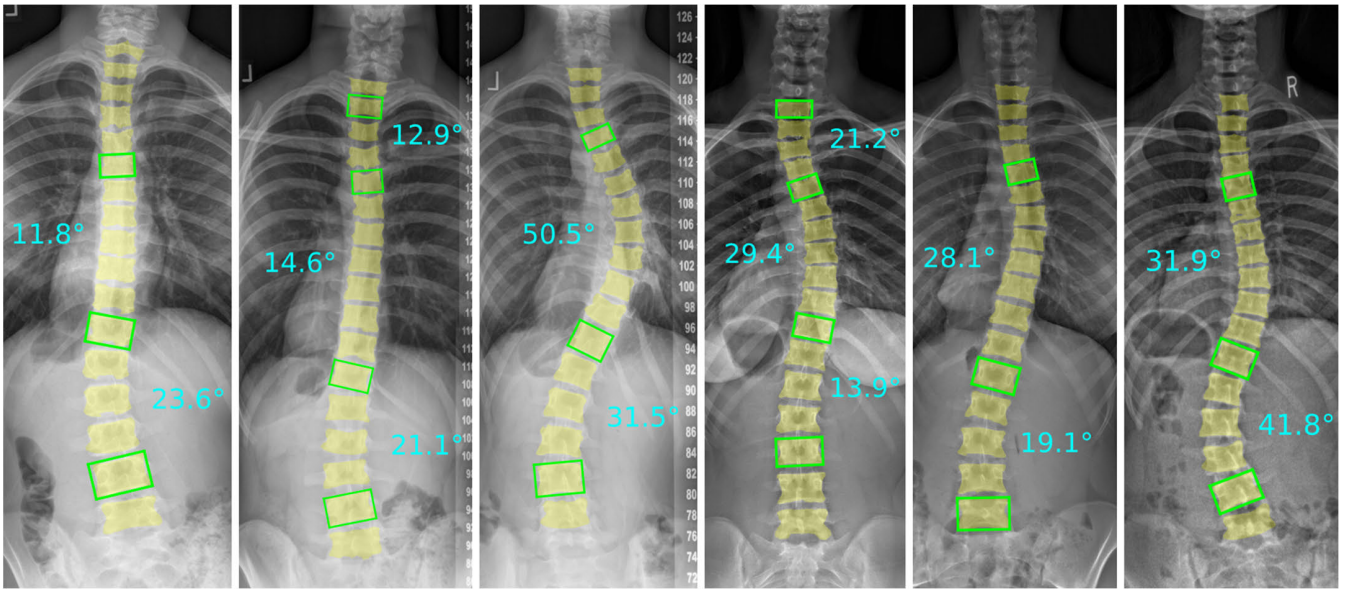


FIGURE 7. Automatic measurement algorithm outputs with relevant vertebrae used in Cobb angle measurement outlined (green boxes) and final Cobb angles listed (cyan text). The three leftmost images are from the conventional x-ray system, and the three rightmost images are from the EOS x-ray system.

C. VALIDATION

1) SPINAL COLUMN AND VERTEBRAL BODY SEGMENTATION

The performance of the optimized segmentation networks was evaluated using 5-fold cross validation. The same hyper-parameters and design choices for the final optimized CNNs were used in cross validation, except for the number of images. All labelled images were used in each 5-fold cross validation, with 110 spinal column images and 340 vertebral body images. The mean and standard deviation of the Dice coefficient, precision, and recall over all folds were reported as the segmentation performance metrics.

2) COBB ANGLE MEASUREMENT

A total of 200 spinal PA radiographs were randomly selected for automatic Cobb angle measurement to validate the developed method. The average of the manually measured Cobb angles of these 200 subjects was 24.6°±9.7° (range: 8° - 52°). None of the images in this set were involved in network training or algorithm tuning. The automatic Cobb angle (A-Cobb)

IEEE Journal of Translational Engineering in Health and Medicine

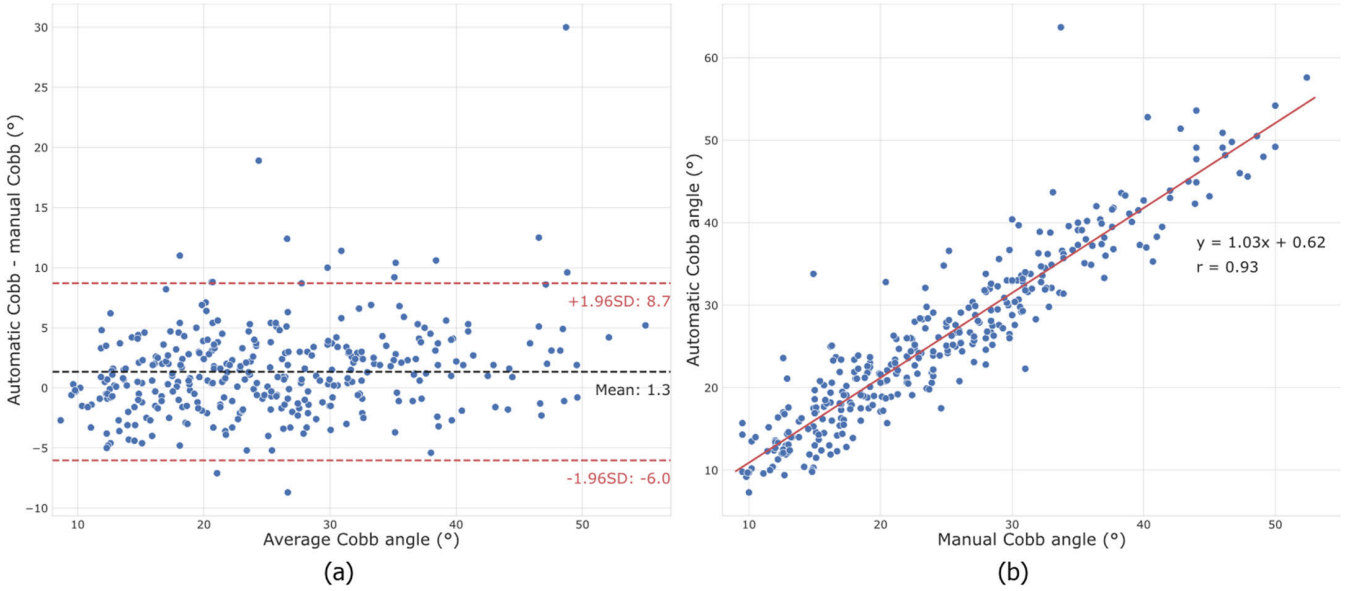


FIGURE 8. (a) Bland-Altman plot of A-Cobb and M-Cobb measurements with a bias of 1.3° (black line) and (−6.0°, 8.7°) limits of agreement (red lines); (b) Scatter plot of A-Cobb vs. M-Cobb measurements with line of best fit in red (y = 1.03x + 0.62).

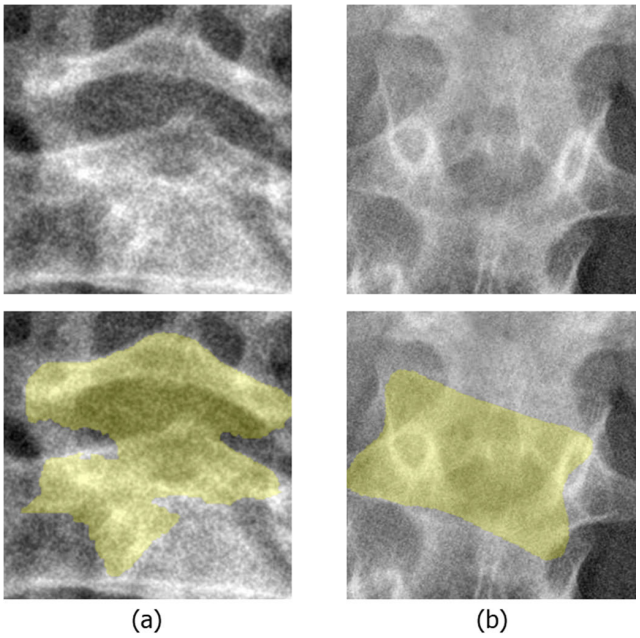


FIGURE 9. Poor segmentations of the end vertebrae that resulted in high circular absolute errors of: (a) 30° and (b): 19°.

measurements were compared to manual Cobb angle (M-Cobb) measurements performed on the same radiograph. All M-Cobb measurements were performed by a rater who was blinded to the A-Cobb measurements and had over 20 years of experience measuring Cobb angles manually.

To evaluate the measurement accuracy, the circular mean absolute error (CMAE) and standard deviation of circular absolute errors (SD) were calculated. These metrics were

TABLE 2. Cross validation (5-FOLD) results for CNNs.

Network	Dice	Precision	Recall
Spinal column	0.957±0.003	0.959±0.004	0.955±0.006
Vertebral body	0.915±0.006	0.911±0.011	0.929±0.006

calculated as follows:

$$\text{CMAE} = \frac{1}{N}\sum_{i=1}^{N} \arctan\left(\frac{\sin\left(\left|\theta_i^a - \theta_i^m\right|\right)}{\cos\left(\left|\theta_i^a - \theta_i^m\right|\right)}\right) \tag{3}$$

$$\text{SD} = \sqrt{\frac{\sum_{i=1}^{N}\left[\arctan\left(\frac{\sin\left(\left|\theta_i^a - \theta_i^m\right|\right)}{\cos\left(\left|\theta_i^a - \theta_i^m\right|\right)}\right) - \text{CMAE}\right]^2}{N-1}} \tag{4}$$

where N is the number of paired measurements and θ^a and θ^m are the A-Cobb and M-Cobb angle, respectively. Additionally, percentage of measurements within clinical acceptance between A-Cobb and M-Cobb measurements was reported. Clinical acceptance was defined as an A-Cobb measurement being within at most 5° of the respective M-Cobb measurement. The inter-method intraclass correlation coefficients ($\text{ICC}_{2,1}$) were calculated to evaluate the reliability. The vertebral level agreement was evaluated using the error index (EI), calculated as follows:

$$\text{EI} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\left(u_i^a - u_i^m\right)^2 + \left(l_i^a - l_i^m\right)^2} \tag{5}$$

where u^a and u^m are the automatic and manual upper end vertebral levels and l^a and l^m are the automatic and manual lower end vertebral levels [22].

Analysis was further split by curve severity into a mild (<25°), moderate (25°-45°), and severe (≥45°) group. These angle thresholds were chosen because they correspond with

TABLE 3. Comparison results for M-cobb vs. A-cobb measurements on the measurement test set for different categories.

Grouping	Method	# of curves detected	# of curves missed	% within clinical	CMAE±SD (°)	$ICC_{2,1}$	EI
All	Sigurdson [9]	341	11	86% (292/341)	3.2°±3.9°	0.87	1.3
	Current	346	6	91% (316/346)	2.8°±2.8°	0.92	1.1
Mild (<25°)	Sigurdson [9]	185	7	86% (159/185)	3.3°±4.6°	--	1.4
	Current	186	6	93% (173/186)	2.7°±2.4°		1.3
Moderate (25°-45°)	Sigurdson [9]	144	4	84% (121/144)	3.1°±3.0°	--	1.2
	Current	148	0	89% (131/148)	3.0°±3.3°		0.9
Severe (≥45°)	Sigurdson [9]	12	0	100% (12/12)	2.3°±1.4°	--	0.8
	Current	12	0	100% (12/12)	2.6°±1.5°		0.6

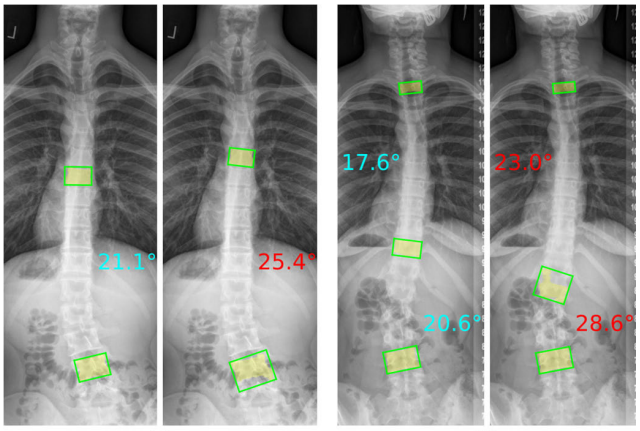


FIGURE 10. Examples of improvements between our proposed method (left) and the method of Sigurdson et al. (right) for PA radiographs in the Cobb angle test set. For both cases, all measured curves were pushed within clinical acceptance with the improved method. The manual measurements for the curves are: (a) 18°; (b) 13° upper and 16° lower.

different treatment options for AIS [1]. The $ICC_{2,1}$ values were not reported for the curve severity groups because the restriction of population variance attenuates the coefficients. Independent samples Student's t-tests were conducted between the CMAEs of curve severity groups to identify any potential systematic biases in algorithm performance. A threshold of $p < 0.05$ indicated statistical significance. To assess the improvements from the network optimizations and point-set registration over the method developed by Sigurdson et al. [14], the 200-image test set in this study was also measured using the method in [14] for comparison.

Bland-Altman analysis was conducted to evaluate the level of agreement between the two measurement methods [23]. Linear regression analysis was performed to evaluate the relationship and linear correlation between the M-Cobb and A-Cobb measurements.

III. RESULTS

A. SPINAL COLUMN AND VERTEBRAL BODY SEGMENTATION

The 5-fold cross validation results for each segmentation network are listed in **Table 2**. Both networks achieved a Dice coefficient greater than or equal to 0.9 for all folds. The optimized spinal column and vertebral body segmentation networks converged in 398 and 195 epochs, respectively.

B. COBB ANGLE MEASUREMENT

Among the 200 test images, the experienced rater measured 352 Cobb angles (M-Cobb). The developed automatic algorithm successfully identified 346 of these curves, missing 6 curves. Overall, 91% of the A-Cobb measurements were within the clinically accepted error of 5°. Additionally, the CMAE for all categories of measurements were below the 5° clinical acceptance threshold. There were no statistically significant differences between mild, moderate, and severe groups, indicating no systematic biases in measurement performance. The method of Sigurdson et al. successfully identified 341 curves, missing 11 in total. It achieved a lower percentage within clinical acceptance of 86% and a higher CMAE of 3.2°. **Table 3** outlines the results of the M-Cobb vs. A-Cobb paired measurements comparison for all curves and the different curve severity categories.

The current automatic measurement algorithm took on average 17.7±10.2 seconds to measure the Cobb angles per radiograph. The algorithm measured much more quickly on radiographs taken by the EOS system (9.8±3.1 seconds) than the conventional system (25.7±8.3 seconds). Examples of segmented and measured radiograph outputs from the current method are illustrated in **Fig. 7**. The method of Sigurdson et al. took 76.2±34.2 seconds on average to measure the Cobb angles per radiograph.

A Bland-Altman plot and scatter plot with line of best fit for A-Cobb vs. M-Cobb measurements are illustrated in **Fig. 8**. The Bland-Altman plot resulted in a bias and limits of agreement of 1.3° (-6.0°, 8.7°). The 95% confidence interval of the mean difference was [0.94°, 1.74°], meaning that the bias was significant as this interval excludes 0°. The equation for the line of best fit in the scatter plot was $y = 1.03x + 0.62$ with a correlation (r) of 0.93.

IV. DISCUSSION

A. RESULTS ANALYSIS

The current method achieved 91% of A-Cobb measurements within clinical acceptance of the M-Cobb measurements. The percentage of measurements within clinical acceptance is a

IEEE Journal of Translational Engineering in Health and Medicine

useful metric in determining the feasibility of the algorithm from a clinical perspective. This high accuracy coupled with the high interpretability that our method offers – namely, providing overlays of the individually segmented vertebral bodies along with the angles of each vertebral body – can quell clinicians' doubts in using a machine learning algorithm to automate Cobb angle measurement. With the added benefit of measuring a subject's radiograph within 18 seconds on average, our proposed method has achieved strong clinical feasibility. The estimated average time it takes an experienced rater to manually measure the Cobb angle is 30 seconds per radiograph, so our completely automatic algorithm can significantly expedite clinical workflow. Moreover, our algorithm measures radiographs even more quickly on EOS-based radiographs, with an average measurement time of 10 seconds per radiograph. The EOS radiographs provide better image quality as its x-ray source moves vertically which generates more even penetration energy when compared with the conventional x-ray system in which the x-ray source is in a fixed position (usually at the middle). Image quality at the upper and lower parts of the body from the conventional x-ray system is relatively poor. Therefore, lower quality vertebral body segmentations are more frequently obtained during iterative vertebral body location, resulting in more attempts at re-segmentation and longer measurement times. Nowadays, the EOS system is a more common x-ray system in many scoliosis centers among developed countries due to its low-dose radiation capability.

Moderate curves were automatically measured slightly less accurately than mild or severe curves, but based on the Student's t-tests between pairwise comparisons of curve severity groups, we conclude that there were no substantial discrepancies in measurement accuracy among the different curve severities. The Bland-Altman and scatter plots also support this conclusion, as there are no visual trends of larger error as the value of the Cobb angle increases. The level agreement in mild curves was less accurate (EI: 1.3) than moderate (EI: 0.9) and severe (EI: 0.6) curves. This is expected, however, as mild curves have end vertebrae with shallower tilts. This means that the neighbors of the end vertebrae typically have tilt angles closer to the true end vertebra, making identification of the true end vertebra difficult from both an automatic and manual perspective.

A total of 6 curves that were in the M-Cobb measurements were not detected by the automatic algorithm. The primary reason for these undetected curves was the inclusion of a vertebra with a very shallow tilt angle in the curve. It is easier for the algorithm to miss a curve with at least one end vertebra that has a small tilt angle because the Cobb angle is defined for vertebrae with opposing tilt angles. In 4/6 of the missed curves, the algorithm segmented the relevant end vertebra, but with no opposing tilt angle, even though the actual value of its angle was very close to the manual measurement. One of the other curves was missed due to a poor vertebral body segmentation of the bottom lumbar vertebra (L5). The shape of the L5 vertebra is typically different than the other vertebrae due to its 3D orientation in the spine, meaning that mis-segmentations of L5 are more common. Finally, the last curve was not detected due to a slight mis-segmentation in the spinal column. Because the spinal column segmentation was off, the SCC for this subject did not detect an apex in the upper region of the spine, resulting in no Cobb angle measured at all for that region. It should be noted that none of the undetected curves were the major curve for the subject. The major curve is more important to detect since it is what determines the appropriate treatment option for AIS. Instead, all undetected curves were mild and minor curves, and so would be less important to the diagnosis of the child with AIS.

There were two Cobb angle measurements with notably larger circular absolute errors of 30° and 19°. In both cases, the automatic measurement algorithm overestimated the severity of the curvature due to poor vertebral body segmentation. The end vertebrae that negatively influenced these measurements the most are illustrated in Fig. 9. The segmentation was either so poor that image registration failed at correcting the tilt angle sufficiently, or the poor segmentation was not detected as unreasonable and therefore did not undergo image registration for correction. Fixing these errors could potentially be accomplished by training the vertebral body segmentation network with more data. However, these errors, especially the 30° error, are at least apparent and would likely have been caught by a clinician with a quick inspection of the segmented images.

B. COMPARISON WITH OTHER METHODS

Two other relevant papers reported on an approach that involved CNN segmentation of spinal features to derive automatic Cobb angle measurement [6], [8]. However, the segmentation targets for our method differ from the other papers. The methods in the related literature segment the separated vertebral bodies from the full spinal radiograph, whereas our method splits the task into segmentation of the spinal column as one continuous segment on the full spinal radiograph and segmentation of each individual vertebral body from a square cropped image with the vertebral body roughly centered. Therefore, we unfortunately cannot draw a meaningful one-to-one comparison with these papers.

Table 4 outlines the comparison between our method and other similar methods reported in the literature. The only two algorithms that outperformed our CMAE performance either had a very limited curve severity test set distribution [6] or provided no curve severity test set distribution information at all [8]. Accurate performance on curves above 25° is crucial because this is the approximate Cobb angle threshold where treatment options are seriously considered. Brace treatment is considered as an option at around the 25° mark, and surgery is considered for curves at around 45° [1]. Therefore, for an automatic measurement algorithm to be clinically feasible, accurate and reliable measurements must be achieved on more severe curves to avoid misdiagnosis and ensure optimal treatment outcomes. The other method [7] performed worse in terms of CMAE, but had a potentially broader curve

TABLE 4. Comparison of test set information and accuracy performance of various automatic cobb angle measurement algorithms.

Method	CMAE (°)	Curve distribution	# of images
Horng [6]	2.5°±1.7°	<20°	35
Fu [7]	3.2°±3.1°	2°-92°	240
Zhao [8]	2.5°	--	75
Sigurdson [9]	3.2°±3.9°	24.5°±9.5° (10°-52°)	200
Current	2.8°±2.8°	24.5°±9.5° (10°-52°)	200

severity distribution. However, this method suffers from not providing interpretability, meaning that one cannot visually confirm the validity of the measurements.

Based on the performance of the method of Sigurdson et al. on our expanded test set, we conclude that the network optimizations and point-set registration step positively impacted the Cobb angle measurement automation algorithm. The CMAE and SD improved by 0.4° and 1.1°, respectively. More importantly, the percentage of measurements within clinical acceptance improved by 5% from 86% to 91%, which is a metric that is of great relevance to clinicians when determining a measurement algorithm's clinical feasibility.

Fig. 10 depicts a comparison of performance between the two methods for two PA radiographs in the Cobb angle test set. The lower end vertebra for the curves in **Fig. 10a** particularly illustrates the impact that an extraneous protrusion can have on a Cobb angle measurement. For this case, the added point-set registration step removed the protrusions, thereby pushing the automatic measurement within clinical acceptance.

C. LIMITATIONS

A limitation of this study is that there were not as many severe curves included in the Cobb angle measurement test set. Our method reports the curve severity distribution and measures accurately on curves >20°, unlike Horng et al. and Zhao et al. [6], [8]. However, while the algorithm did achieve 100% of A-Cobb measurements within clinical acceptance for severe curves, 12 datapoints is not enough to confidently conclude that the algorithm performs well on curves ≥45°. Severe curves make up a small proportion of AIS cases, since bracing is typically prescribed to prevent moderate curves from progressing to these high Cobb angles. Therefore, with a randomly selected measurement set, severe curves will naturally comprise a smaller proportion of the set's population. Consequently, a further study with a test set of only cases with severe major Cobb angles needs to be conducted to confidently validate the proposed method.

Another limitation is that the study population was limited only to subjects with AIS. There are different forms of scoliosis, such as congenital or neuromuscular, that would benefit from an automatic Cobb angle measurement method as well. While we have not tested our algorithm on congenital or neuromuscular scoliosis patients, we strongly suspect that the CNNs would need to be re-trained with labelled images of other types of scoliosis to be accurate on those forms of scoliosis.

Finally, there were no non-scoliotic cases in the validation set population. We suspect that the algorithm would not misclassify non-scoliotic cases as scoliotic, since non-scoliotic radiographs were included in the training sets of the CNNs for this reason. However, future work would consist of testing the algorithm on a set of non-scoliotic radiographs to confidently validate this suspicion.

V. CONCLUSION

A fully automatic Cobb angle measurement method on PA radiographs was developed using a cascaded design of two CNNs, where one CNN segmented the spinal column and the other segmented individual vertebral bodies. A point-set registration step was added to improve vertebral body segmentations and was found in many cases to push measurements within the clinically accepted error. The developed method yielded 91% of measurements with clinical acceptance with a 2.8° CMAE, indicating high accuracy. Measurements were obtained within 18s on average per radiograph, which is quicker than what an experienced rater manually measures. Additionally, measurements were output in an interpretable fashion, with the segmentations being output along with the measurement so that clinicians can quickly confirm the validity of the measurement. These three characteristics of the method are key in realizing actual implementation in clinical practice. Future work consists of validating on a set of more severe curves to cement this method as truly clinically feasible. When this is completed, the algorithm could offer robust measurements for informed treatment diagnosis of AIS and streamline clinical workflow.

ACKNOWLEDGMENT

The authors would like to thank the Edmonton Scoliosis Interdisciplinary Research Group (ESIRG) for the radiographic data used in this study, also would like to thank the Industry Sandbox & AI Computing (ISAIC) at the University of Alberta for the use of their supercomputer that made this research possible, and also would like to thank Solvin Sigurdson for his work on the initial algorithm that served as a foundation for the current algorithm reported in this study.

REFERENCES

[1] S. L. Weinstein, L. A. Dolan, J. C. Y. Cheng, A. Danielsson, and J. A. Morcuende, "Adolescent idiopathic scoliosis," *Lancet*, vol. 371, no. 9623, pp. 1527–1537, May 2008, doi: 10.1016/S0140-6736(08)60658-3.

[2] J. Cobb, "Outline for the study of scoliosis," in *American Academy of Orthopaedic Surgeons Instructional Course Lectures*i, vol. 5. Ann Arbor, MI, USA: Edwards, 1948, pp. 261–275.

[3] S. Negrini et al., "2016 SOSORT guidelines: Orthopaedic and rehabilitation treatment of idiopathic scoliosis during growth," *Scoliosis Spinal Disorders*, vol. 13, no. 1, p. 3, Jan. 2018, doi: 10.1186/s13013-017-0145-8.

IEEE Journal of Translational Engineering in Health and Medicine

[4] T. Vrtovec, F. Pernuš, and B. Likar, "A review of methods for quantitative evaluation of spinal curvature," *Eur. Spine J.*, vol. 18, no. 5, pp. 593–607, May 2009, doi: 10.1007/s00586-009-0913-0.

[5] J. Scheetz et al., "A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology," *Sci. Rep.*, vol. 11, no. 1, Mar. 2021, Art. no. 1, doi: 10.1038/s41598-021-84698-5.

[6] M.-H. Horng, C.-P. Kuok, M.-J. Fu, C.-J. Lin, and Y.-N. Sun, "Cobb angle measurement of spine from X-ray images using convolutional neural network," *Comput. Math. Methods Med.*, vol. 2019, Feb. 2019, Art. no. 6357171, doi: 10.1155/2019/6357171.

[7] X. Fu, G. Yang, K. Zhang, N. Xu, and J. Wu, "An automated estimator for cobb angle measurement using multi-task networks," *Neural Comput. Appl.*, vol. 33, no. 10, pp. 4755–4761, Nov. 2020, doi: 10.1007/s00521-020-05533-y.

[8] Y. Zhao, J. Zhang, H. Li, X. Gu, Z. Li, and S. Zhang, "Automatic cobb angle measurement method based on vertebra segmentation by deep learning," *Med. Biol. Eng. Comput.*, vol. 60, no. 8, pp. 2257–2269, Jun. 2022, doi: 10.1007/s11517-022-02563-7.

[9] F. Galbusera et al., "Fully automated radiological analysis of spinal disorders and deformities: A deep learning approach," *Eur. Spine J.*, vol. 28, no. 5, pp. 951–960, May 2019, doi: 10.1007/s00586-019-05944-z.

[10] K. C. Kim, H. S. Yun, S. Kim, and J. K. Seo, "Automation of spine curve assessment in frontal radiographs using deep learning of vertebral-tilt vector," *IEEE Access*, vol. 8, pp. 84618–84630, 2020, doi: 10.1109/ACCESS.2020.2992081.

[11] Y. Pan et al., "Evaluation of a computer-aided method for measuring the cobb angle on chest X-rays," *Eur. Spine J.*, vol. 28, no. 12, pp. 3035–3043, Dec. 2019, doi: 10.1007/s00586-019-06115-w.

[12] T. A. Sardjono, M. H. F. Wilkinson, A. G. Veldhuizen, P. M. A. van Ooijen, K. E. Purnama, and G. J. Verkerke, "Automatic cobb angle determination from radiographic images," *Spine*, vol. 38, no. 20, pp. E1256–E1262, Sep. 2013, doi: 10.1097/brs.0b013e3182a0c7c3.

[13] H. Wu, C. Bailey, P. Rasoulinejad, and S. Li, "Automated comprehensive adolescent idiopathic scoliosis assessment using MVC-net," *Med. Image Anal.*, vol. 48, pp. 1–11, Aug. 2018, doi: 10.1016/j.media.2018.05.005.

[14] S. Sigurdson, J. Wong, M. Reformat, and E. Lou, "Applying a convolutional neural network based iterative algorithm to automatically measure spinal curvature on radiographs for children with scoliosis," *J. Med. Biol. Eng.*, vol. 42, no. 3, pp. 388–396, May 2022, doi: 10.1007/s40846-022-00712-9.

[15] *MATLAB (R2022b), Image Processing Toolbox*, MathWorks, Natwick, MA, USA, 2022.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2015* (Lecture Notes in Computer Science). Munich, Germany: Springer, 2015, pp. 234–241.

[17] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, 2015.

[18] S. Jadon, "A survey of loss functions for semantic segmentation," in *Proc. IEEE Conf. Comput. Intell. Bioinf. Comput. Biol. (CIBCB)*, Oct. 2020, pp. 1–7, doi: 10.1109/CIBCB48159.2020.9277638.

[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.

[20] M. M. Panjabi et al., "Thoracic human vertebrae quantitative three-dimensional anatomy," *Spine*, vol. 16, no. 8, pp. 888–901, Aug. 1991, doi: 10.1097/00007632-199108000-00006.

[21] M. M. Panjabi et al., "Human lumbar vertebrae: Quantitative three-dimensional anatomy," *Spine*, vol. 17, no. 3, pp. 299–306, Mar. 1992, doi: 10.1097/00007632-199203000-00010.

[22] S. Du, N. Zheng, S. Ying, Q. You, and Y. Wu, "AN extension of the ICP algorithm considering scale factor," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2007, pp. 193–196, doi: 10.1109/icip.2007.4379798.

[23] W. Chen, E. H. M. Lou, P. Q. Zhang, L. H. Le, and D. Hill, "Reliability of assessing the coronal curvature of children with scoliosis by using ultrasound images," *J. Children's Orthopaedics*, vol. 7, no. 6, pp. 521–529, Dec. 2013, doi: 10.1007/s11832-013-0539-y.

[24] D. G. Altman and J. M. Bland, "Measurement in medicine: The analysis of method comparison studies," *Statistician*, vol. 32, no. 3, pp. 307–317, 1983, doi: 10.2307/2987937.