



Original Article

# Convolutional Neural Network to Segment Laminae on 3D Ultrasound Spinal Images to Assist Cobb Angle Measurement

JASON WONG,<sup>1</sup> MAREK REFORMAT,<sup>1</sup> ERIC PARENT,<sup>2</sup> and EDMOND LOU<sup>1</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Donadeo Innovation Centre for Engineering, University of Alberta, Donadeo ICE 11-263, 9211-116 Street, Edmonton, AB T6G 1H9, Canada; and <sup>2</sup>Department of Physical Therapy, Faculty of Rehabilitation Medicine, University of Alberta, 2-50 Corbett Hall, Edmonton, AB T6G2G4, Canada

(Received 9 November 2021; accepted 22 January 2022; published online 24 February 2022)

Associate Editor Arash Kheradvar oversaw the review of this article.

**Abstract**—A recent innovation in scoliosis monitoring is the use of ultrasonography, which provides true 3D information in one scan and does not emit ionizing radiation. Measuring the severity of scoliosis on ultrasonographs requires identifying lamina pairs on the most tilted vertebrae, which is difficult and time-consuming. To expedite and automate measurement steps, this paper detailed an automatic convolutional neural network-based algorithm for identifying the laminae on 3D ultrasonographs. The predicted laminae were manually paired to measure the lateral spinal curvature on the coronal view, called the Cobb angle. In total, 130 spinal ultrasonographs of adolescents with idiopathic scoliosis recruited from a scoliosis clinic were selected, with 70 for training and 60 for testing. Data augmentation increased the effective training set size to 140 ultrasonographs. Semi-automatic Cobb measurements were compared to manual measurements on the same ultrasonographs. The semi-automatic measurements demonstrated good inter-method reliability ( $ICC_{3,1} = 0.87$ ) and performed better on thoracic ( $ICC_{3,1} = 0.91$ ) than lumbar curves ( $ICC_{3,1} = 0.81$ ). The mean absolute difference and standard deviation between semi-automatic and manual was  $3.6^\circ \pm 3.0^\circ$ . In conclusion, the semi-automatic method to measure the Cobb angle on ultrasonographs is feasible and accurate. This is the first algorithm that automates steps of Cobb angle measurement on ultrasonographs.

**Keywords**—Machine learning, CNN, Scoliosis, Ultrasonography, Reliability.

## ABBREVIATIONS

AIS	Adolescent idiopathic scoliosis
AOR	Aid of previous radiograph
CI	Confidence interval
COL	Center of lamina
CNN	Convolutional neural network
FOV	Field of view
ICC	Intra-class correlation coefficient
MAD	Mean absolute difference
MT	Main thoracic
MU-Cobb	Manual ultrasound Cobb angle measurements
MX-Cobb	Manual radiographic Cobb angle measurements
ROI	Region of interest
SD	Standard deviation
SEM	Standard error of measurement
SU-Cobb	Semi-automatic ultrasound Cobb angle measurements
TL/L	Thoracolumbar/lumbar

## INTRODUCTION

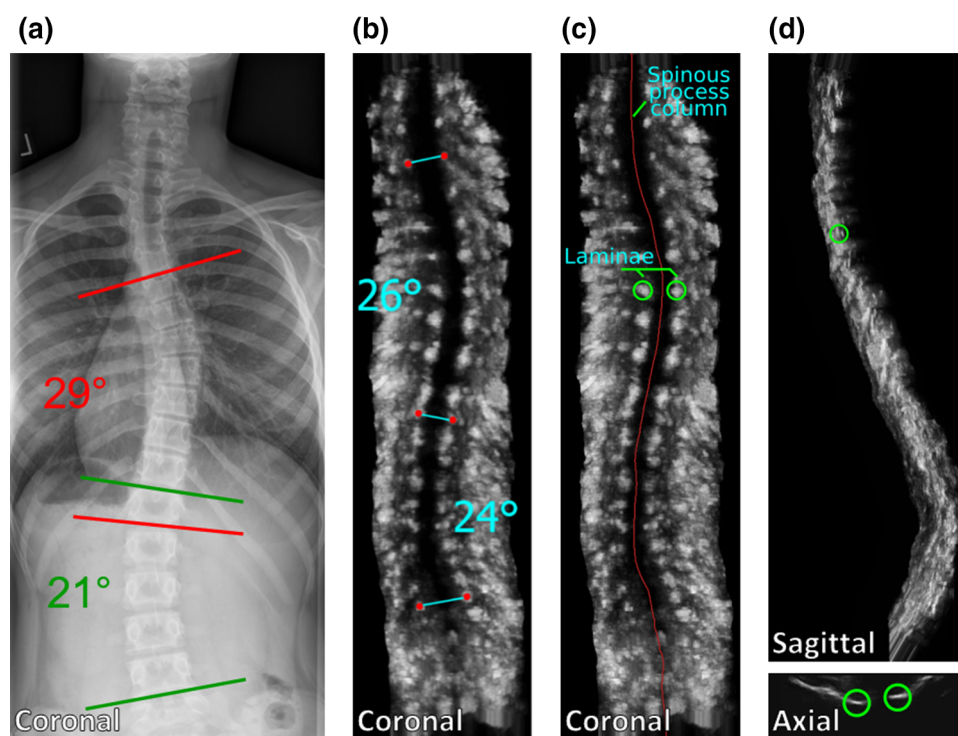
Adolescent idiopathic scoliosis (AIS) is a spinal disorder that develops mostly in adolescent females, aged 10 to 16 years old, where the spine has undergone 3D structural changes. These changes typically take the form of a lateral S-shaped curvature and axial vertebral rotation. AIS affects 1–3% of the adolescent population and can lead to cardiopulmonary problems

Address correspondence to Edmond Lou, Department of Electrical and Computer Engineering, Donadeo Innovation Centre for Engineering, University of Alberta, Donadeo ICE 11-263, 9211-116 Street, Edmonton, AB T6G 1H9, Canada. Electronic mail: elou@ualberta.ca

and back pain if left untreated.<sup>18</sup> Monitoring the spinal condition requires regularly taking posteroanterior radiographs, which exposes adolescents to ionizing radiation, and measuring the severity of the lateral curvature using the Cobb angle.<sup>4</sup> Although the low ionizing radiation dosage X-ray system (EOS Imaging, France) has become more popular, the accumulated radiation on children may still increase the risk of cancer. Children with AIS are five times more likely to develop cancer later in life,<sup>14</sup> and girls with AIS were found to have a 70% excess risk of dying from breast cancer later in life when compared with the general population.<sup>6</sup> Ultrasound, a radiation-free imaging modality, for scoliosis monitoring has recently been investigated and found to be comparable to radiographic measurements in terms of accuracy and reliability.<sup>21</sup> Ultrasound also inherently provides 3D information. This allows clinicians to directly measure parameters, such as vertebral rotation,<sup>2</sup> that would normally require estimation methods using a single posteroanterior radiograph.<sup>10</sup> Clinicians can also obtain a better understanding of the severity of the structural changes by measuring true 3D parameters, such as the Cobb angle on the plane of maximum curvature.<sup>15</sup> Figures 1a and 1b shows the Cobb angles of a child with a major right thoracic and minor left

lumbar curve measured on a posteroanterior radiograph and ultrasonograph, respectively.

The complication with using ultrasonography for AIS is that the scans are less intuitive, and it is therefore more time-consuming to analyze and make measurements. Measuring the Cobb angle on ultrasonographs requires identifying the laminae, which are flat surfaces on the left and right of a vertebra and appear as very bright isolated regions on the scan. Another feature that is useful for Cobb angle measurement is the spinous process column, which is the curve that is formed by the spinous processes of all the vertebrae. The spinous process is a protrusion out of the back of each vertebra that appears as a dark thin region in the middle of the spine on the ultrasound scan. Figure 1c shows an ultrasonograph with the spinous process column in the middle and a pair of laminae on either side of this column. The Cobb angle can be measured using the center of lamina (COL) method. This consists of drawing a line joining each pair of laminae on the same vertebra, taking the pairs with the steepest opposing tilt angles, and calculating the difference between the angles.<sup>3</sup> Identifying which laminae are on the same vertebra involves choosing pairs which form a line that is roughly perpendicular to the spinous process column in the middle of the image. Identifying the lamina locations on ultrasonographs is more challenging, as it requires more



**FIGURE 1.** (a) The measured Cobb angles labelled on a posteroanterior radiograph; (b) the measured Cobb angles labelled on a coronal projection image of an ultrasonograph of the same subject using the center of lamina method; (c) the coronal projection image of the same ultrasonograph with the spinous process column and a pair of laminae labelled; (d) the sagittal view (top) and axial view of one frame (bottom) with the same pair of laminae labelled.

knowledge of vertebral anatomy and uses three different views (coronal, axial, and sagittal). Figure 1d shows the sagittal view of the same subject (top) and the axial view of one B-mode image frame (bottom).

The difficulty in measuring the Cobb angle on ultrasonographs leads to a barrier of entry for clinics to adopt ultrasound monitoring for children with AIS. Consequently, a few research groups have used the spinous process features to automatically measure the spinous process angle (SPA), which has demonstrated a strong correlation with the Cobb angle. However, converting the SPA to Cobb angle requires a large dataset to generate an equation. Wong et al. used phase congruency and a thresholding algorithm on coronal ultrasound images to obtain the SPA.<sup>20</sup> Their automatic SPA measurements were correlated ( $r = 0.82$ ) with the radiographic Cobb angle, but only 60% of their corrected Cobb angles were within the  $5^\circ$  clinically accepted error. Brink et al. used a similar image processing technique to extract the SPA from 33 children with AIS and achieved a mean absolute difference (MAD) and standard deviation (SD) of  $4.9^\circ \pm 3.2^\circ$  between automatic ultrasound SPA and manual radiographic Cobb angle measurements.<sup>1</sup> Ge et al. used a gradient vector flow snake model to automatically extract the SPA. The  $\text{MAD} \pm \text{SD}$  between the automatic ultrasound SPA versus the manual ultrasound and the manual radiographic SPA were  $3.3^\circ \pm 2.4^\circ$  and  $2.7^\circ \pm 2.1^\circ$ , respectively.<sup>7</sup> Zhou et al. expanded on their phase congruency and thresholding algorithm by leveraging symmetry information in the axial views, which increased the correlation with the manual radiographic Cobb angle to 0.87.<sup>23</sup> On the other hand, using the COL method on ultrasonographs without and with the aid of previous radiograph (AOR) produces direct comparison with the radiographic Cobb angle. The MAD between the ultrasound Cobb angle and the radiographic Cobb angle was  $4.6^\circ \pm 3.8^\circ$  (without AOR) and  $2.7^\circ \pm 1.9^\circ$  (with AOR), with an  $R^2$  of 0.58 (without) and 0.87 (with).<sup>22</sup> Since measurement on a posteroanterior radiograph is the gold standard for the Cobb angle,<sup>4</sup> automation using the COL method is sought to produce that direct comparison, while also minimizing measurement error, freeing up clinician workload, and improving measurement accuracy and reliability.

A convolutional neural network (CNN) is a type of machine learning model that is frequently used for image classification and segmentation.<sup>11</sup> CNNs have been used in the medical field for segmentation and have succeeded in both automating and expediting segmentation while maintaining comparable accuracy to clinicians.<sup>8</sup> Therefore, the objectives of this study were to develop a semi-automatic method to extract the Cobb angle from ultrasonographs using a CNN

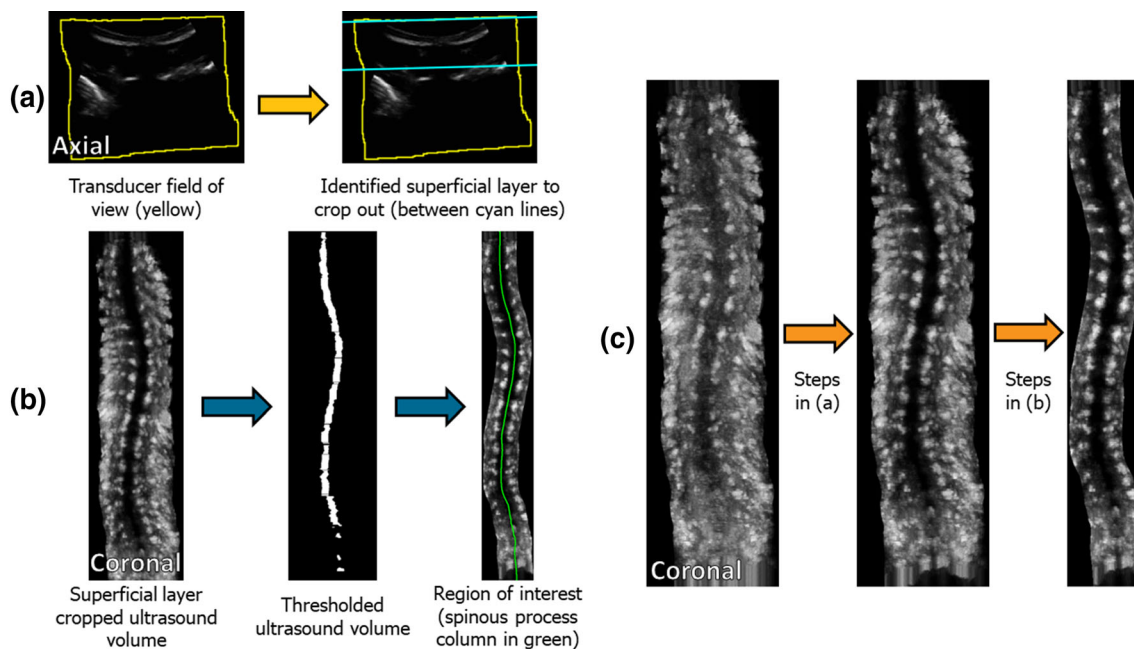
and to determine the reliability and convergent validity of this machine learning method.

## MATERIALS AND METHODS

### *Data Acquisition and Processing*

Spinal ultrasound volume data of children with AIS were acquired at the local scoliosis clinic. Ethics approval was granted by the local research health ethics board and all participants signed written consent forms prior to participating in the study. The inclusion criteria were participants with AIS who had a major Cobb angle less than  $46^\circ$ , no prior surgical treatment, and an out of brace radiograph at that clinic. The ultrasound scans were acquired using a Sonix TABLET ultrasound system coupled with a C5-2/60 GPS curvilinear convex transducer (BK Medical, USA). The scanning parameters were 2.5 MHz scan frequency with a 6cm penetration imaging depth and 10% gain with linear time gain compensation. This system kept track of the position and orientation of the transducer throughout the scan. Participants were instructed to stand in a standard posture similar to the X-ray acquisition while the transducer was moved, following the lateral spinal curvature. Each scan was obtained by an experienced operator, starting at the C7 vertebra and ending at the L5 vertebra. Approximately 700 to 1000 axial B-mode images along with the position and orientation data of the transducer were obtained per spine. Using this series of images, 3D spinal ultrasonographs were reconstructed using an in-house developed software, called Medical Imaging Analysis System. A total of 130 3D spinal ultrasonographs were selected from the local database for this study.

The original ultrasound data were first processed, as they included other features, such as the skin, muscles, and fat, that produce extraneous reflections. This involved taking each axial B-mode image and then identifying the transducer's field of view (FOV) to crop the region of interest. Identifying the FOV was accomplished by locating the nonnegative pixel values, as any pixels outside the transducer's FOV were assigned negative values. A line Hough transform was then employed to obtain both the position and tilted angle of the FOV, and the superficial layer of voxels was cropped a fixed distance below the top of the FOV at the same tilted angle. This fixed distance was 40% of the height of the FOV, the percentage being determined by what produced the clearest results for all ultrasound volumes. Figure 2a illustrates the step of eliminating this superficial area from the volume for one axial image. This cropping is performed for all



**FIGURE 2.** The process for preparing images for input to the CNN: (a) cropping off the ultrasound reflections from the skin, muscles, and fat on an axial image; (b) steps for narrowing volume to the region of interest, shown on coronal projection images; (c) coronal projection images of the ultrasound volume through the different stages of processing.

axial images in the 3D ultrasonograph, resulting in a more pronounced spinous process column in the coronal projection image. Next, the volumes were narrowed to the region of interest (ROI) to reduce the area that the network must search for laminae. This step requires identifying the dark curve created by the spinous process column using thresholding. First, the coronal projection image was divided into horizontal partitions of 50 rows, and each partition was iteratively inverse binary thresholded, with each iteration increasing the brightness threshold, until a certain number of thresholded pixels was reached for each partition (middle image of Fig. 2b). Once each partition underwent this thresholding, the geometric centroids of each row of thresholded pixels was determined to form the basis of the spinous process column. Any rows with a lack of thresholded pixels were filled in with linear interpolation. A moving average with a window of 41 rows was used to smooth out the calculated centroids into a curve. Finally, the volume was cropped to maintain 40% of the coronal projection image's width, centered around the identified spinous process column, to obtain the ROI (right image of Fig. 2b). A flowchart of these steps is depicted in Fig. 2b. All processing steps were performed automatically. Figure 2c illustrates the effects of each pre-processing step from the original ultrasound image (left) to the final pre-CNN processed image (right).

#### *Training Dataset Creation*

Using a custom-built ultrasound volume labelling graphical user interface, the primary author manually labelled the laminae on 70 spinal ultrasonographs based on 3D information. The selected ultrasonographs have a radiographic Cobb angle of  $23.6^\circ \pm 9.1^\circ$  (range  $9^\circ$ – $44^\circ$ ) for the main scoliosis curve. The age of these participants was  $14.5 \pm 1.9$  years old (range 10–18 years) and there were 7 males and 63 females. The procedure for labelling involved using the coronal view to obtain the general lamina locations. Once the laminae on the coronal view were selected with a mouse click, the corresponding axial views of the laminae would be used to adjust and confirm placement. After the 17 pairs of laminae from T1 to L5 were identified, the sagittal view was then used to verify that the laminae placement created a smooth sagittal profile. The primary author has over three years of scoliosis research experience and reviewed over 300 images prior to labelling. Additionally, a senior co-author with over 25 years of experience confirmed the lamina placement on 10 of the ultrasonographs labelled by the primary author before the other 60 were labelled.

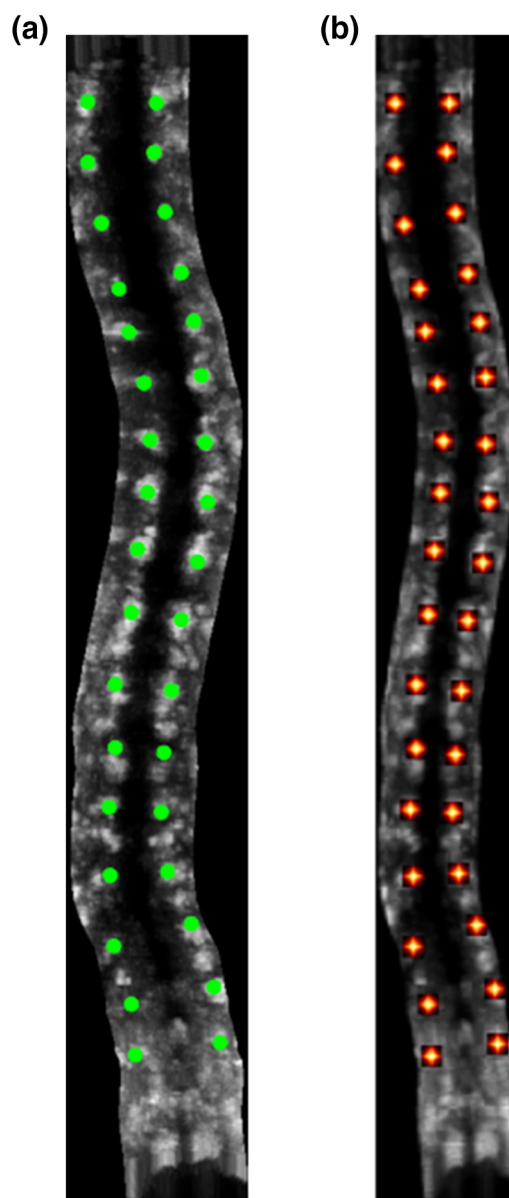
For input into the CNN, the volumes and labels were scaled down to the size  $384 \times 96 \times 48$  pixels, which is about a third of the dimensions of the average spinal ultrasonograph. This size was chosen in consideration of the computational limits for network training (GPU memory size), network architecture (downsampling by a factor of 2 multiple times), and

ability to visually distinguish the COL when down-scaled. Trilinear interpolation was used in the down-scaling of the input volumes. The labels were downscaled by first determining the centroids of the labelled laminae in the unscaled labelled volume, translating the coordinates of these centroids to the downscaled size, and centering a  $7 \times 7 \times 7$  voxel cube around the translated centroid. The voxels in this cube were encoded with a maximum value of 1 in the center of the cube, with values decreasing as you moved further from the center. The degree to which these values decreased was determined using a Gaussian neighborhood function with a standard deviation of 2. Encoding the labels in this ‘fuzzy’ manner forced the network to focus more on the COL and improved network performance over using purely binary labels. Finally, the scaled volume intensities were normalized to zero mean and unit variance. Figure 3 shows an example of an unscaled labelled volume and the corresponding downscaled encoded volume.

#### *Neural Network Architecture and Training*

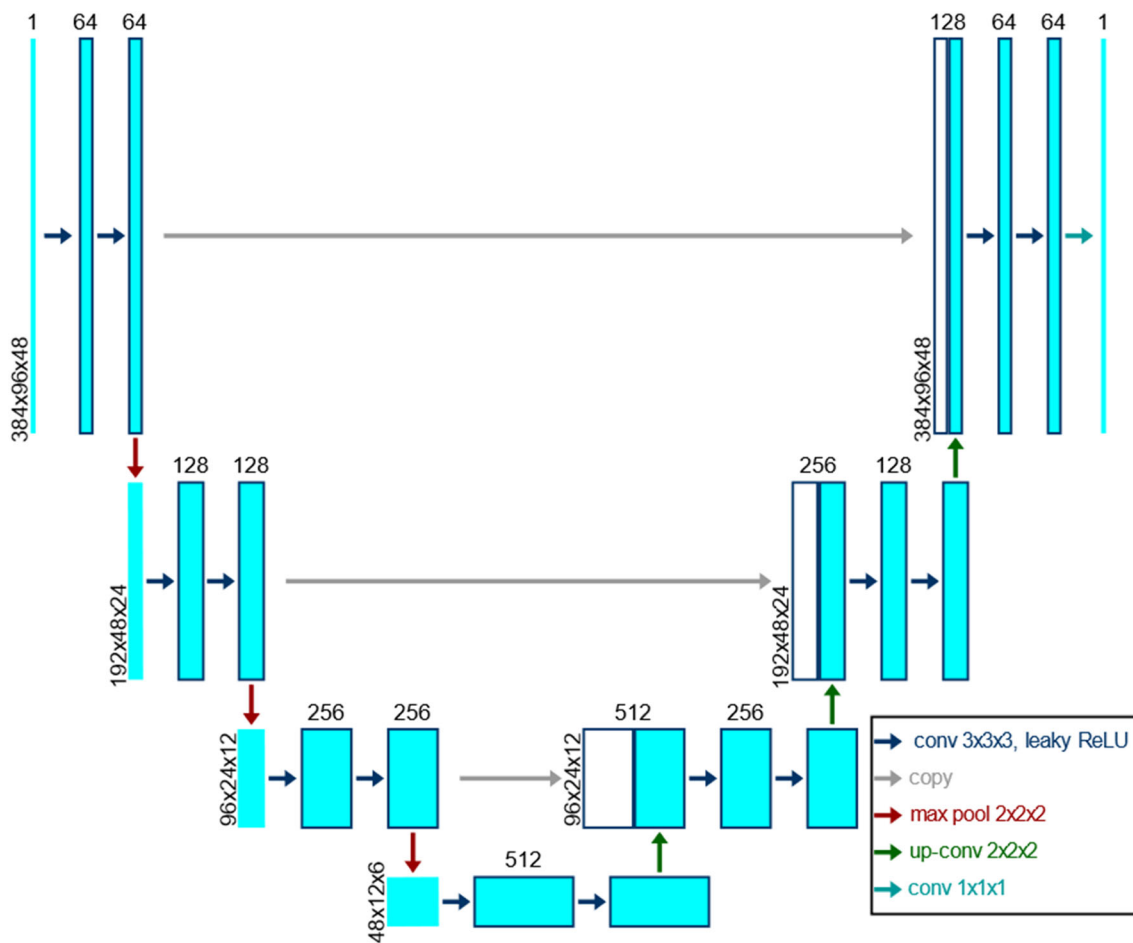
A variant of the U-net CNN architecture<sup>13</sup> was used for segmentation. The U-net is a very common architecture for medical segmentation tasks, as it has been proven to perform well even with very little training data. Three changes were made to this architecture. First, all operations (convolution, pooling, and upsampling) were adapted to take in 3D inputs. Second, to reduce computational complexity, one pooling and one upsampling stage were removed. Third, same padding was employed, which meant that the volumes did not decrease in size after a convolution layer. This design choice was made because in some cases, laminae can be present at the edges of the scans. Figure 4 displays the architecture of the 3D U-net variant. The CNN was trained to identify single COL instead of pairs because for some ultrasonographs, only one lamina of a vertebra is visible. Therefore, encoding the labels as pairs would result in more inaccurate placement for both laminae in these cases. On the other hand, for single lamina encoding, a more accurate placement would be obtained for the visible lamina, which would provide a strong foundation for estimating the position of the other lamina in the pair through post-processing.

Because of the ‘fuzzy’ encoding of the labels, a mean squared error loss function was used instead of the typical binary cross entropy or soft Dice loss. This did come at the expense of longer training time; however, the performance of the network improved significantly by employing this heatmap regression method. The network was trained using an Adam optimizer<sup>9</sup> with a learning rate of  $10^{-4}$ . A linear activation function was



**FIGURE 3.** Coronal projections of the labelled ultrasonographs (a) original labels; (b) ‘fuzzy’ encoded labels with yellow to dark red indicating higher to lower probabilities for the center of lamina locations, respectively.

used for the output layer and a leaky rectified linear unit activation function with an alpha value of 0.01 was used for the hidden layers. The regularization technique of dropout with a probability of 0.5 was used after each pooling and upsampling layer. Because of the intensive memory requirements of the CNN, a batch size of 1 was used. Due to the computational complexity of training a 3D U-net, a hyperparameter optimization strategy, such as a grid search, could not be conducted. Instead, these parameters were tuned through an iterative process of training a network, analyzing the training curve and predictions, and making educated judgments on how to tune the



**FIGURE 4.** 3D U-net architecture for lamina segmentation with cyan boxes indicating feature maps and white boxes representing copied feature maps. The size of the feature maps in each tier is to the left of each convolutional block and the number of feature maps is above each box.

parameters to improve network performance by investigating different aspects of the training curves, such as the smoothness and rate of decrease.

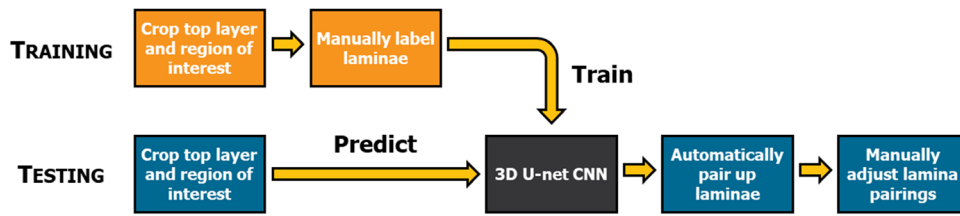
To determine the optimal number of epochs, the 70 volumes were split into a 50-volume training, 10-volume validation, and 10-volume test set. The validation set was used during training to monitor the validation loss. The validation loss was lowest at around 750 epochs, and so this was deemed the optimal number of epochs for the network. The test set was used to evaluate the performance of the CNN's COL placement. The network was then trained from scratch for 750 epochs using all 70 labelled volumes to use for Cobb angle measurement.

A data augmentation method of randomly flipping along the sagittal plane was employed when the volumes were presented to the network for training to increase the diversity. This has the effect of switching right and left on the coronal projection and effectively doubles the size of the training set. Other geometrical

augmentation methods, such as rotation and translation, were not implemented because in some cases, laminae can exist close to the edges of the volume and therefore could be moved outside the boundaries of the image under these augmentation methods.

#### *Cobb Angle Measurement*

The outputs of the 3D U-net can be interpreted as “probability heatmaps”, where each voxel value represents the probability of being a COL. To determine the COL from the probability heatmaps, a more localized and adaptive form of thresholding was implemented. This thresholding is done in an iterative manner since there can be cases where a connected component for a lamina area has more than one peak probability if the probability threshold is high enough. Therefore, the probability threshold was iteratively decremented from 0.5 to 0.1 probability in intervals of 0.05 such that the local peaks could be extracted from



**FIGURE 5.** Flowchart of entire procedure from dataset creation for U-net training to semi-automatic Cobb angle measurement of a test ultrasonography.

the probability heatmap. A distance threshold based on the average anatomical measurements of vertebrae was used to determine whether two predicted COL were too close to each other.<sup>12</sup> If two were deemed too close to each other, the one with the highest probability was chosen as the COL for that area.

A program was developed to facilitate the semi-automatic measurement procedure. Using the spinous process column, a preliminary pairing algorithm was implemented to expedite measurement. A pair of laminae on the same vertebra was estimated such that the two laminae are approximately equidistant from the center of the spinous process column and the line joining the pair forms an angle roughly perpendicular to the spinous process column. These predicted lamina pairs were then displayed. To finalize lamina placement for Cobb angle measurement, a rater needed to confirm and manually adjust any incorrect lamina pairings. No adjustments to the positions of the predicted laminae were made during measurement; only pairing decisions were adjusted. Once the adjustment was finished, the most tilted relevant pairs were manually chosen to calculate the Cobb angle. Figure 5 illustrates a flowchart of the training and testing procedure.

All code was implemented in the Python programming language, using the TensorFlow library for CNN development. The network was trained using a Linux virtual machine on the Industry Sandbox & AI Computing (ISAIC) at the University of Alberta with an Intel Xeon Gold 6138 dual processor, 64GB of RAM, and an NVIDIA Tesla V100 16GB GPU.

#### Validation

To evaluate the performance of the CNN in predicting COL, a distance metric  $d$  describing how far away a ground truth COL (as initially labelled by the primary author) was from any predicted COL was calculated. Let a ground truth and predicted center be denoted as  $\gamma$  and  $\varphi$ , respectively, with the list of predicted centers as  $\Phi$ . The distance metric is then defined

for a ground truth center  $\gamma$  as the minimum distance from  $\gamma$  to any of the centers in  $\Phi$ , or:

$$d_{\gamma} = j = 1 \delta N_{\Phi} \min\{|\gamma - \varphi_j|\}$$

where  $N_{\Phi}$  is the total number of predicted centers. This metric was reported for the 340 COL in the initial 10-volume test set.

To evaluate whether the CNN-based algorithm was positioning the lamina predictions accurately for Cobb angle measurement, the other 60 ultrasonographs which had not been employed for training were used for measurement validation. Currently, the gold standard to measure the Cobb angle is performed on posteroanterior radiographs using the Cobb method.<sup>4</sup> Therefore, semi-automatic ultrasound Cobb angle measurements (SU-Cobb), which were performed by the primary author (rater R1), were compared to manual ultrasound measurements performed with aid of previous radiograph (AOR)<sup>22</sup> (MU-Cobb) and manual X-ray or radiographic Cobb angle measurements (MX-Cobb). The MU-Cobb with AOR were performed by two raters who had over three years of experience and intra-rater reliabilities ( $ICC_{2,1}$ ) of 0.96 and 0.94 with their ultrasound measurements being 2.1° and 2.8° from manual radiographic measurements on average.<sup>21,22</sup> The MX-Cobb were recorded in the clinical records which were measured by clinicians with over 15 years of experience. The rater R1 was blinded to other manual measurements and did not use AOR.

The accuracy of the CNN-based algorithm was evaluated by calculating the mean absolute difference (MAD), standard deviation (SD), and standard error of measurement (SEM) between SU-Cobb vs. MU-Cobb with AOR and SU-Cobb vs. MX-Cobb. The inter-method intraclass correlation coefficient using a two-way mixed model with single measures ( $ICC_{3,1}$ ) with 95% confidence intervals (CI) was calculated. The ICC was qualitatively evaluated using Currier's definitions of poor ( $< 0.70$ ), fair (0.70–0.79), good (0.80–0.89), and excellent (0.90–0.99).<sup>5</sup> Additionally, the percentage of semi-automatic measurements within clinical acceptance of the MU-Cobb with AOR was

calculated. Clinical acceptance was defined as within at most  $5^\circ$  of the manual measurement.<sup>19</sup> Categorical analysis on curve region and severity was performed to identify any systematic differences in performance. The curve region was defined according to the apical vertebral level with main thoracic (MT) as T2-T11 and thoracolumbar/lumbar (TL/L) as T12-L4.<sup>16</sup> Curve severity analysis was split into mild ( $< 25^\circ$ ) and moderate ( $\geq 25^\circ$ ) curves. Statistical analysis was performed using the pingouin and pandas Python libraries.

## RESULTS

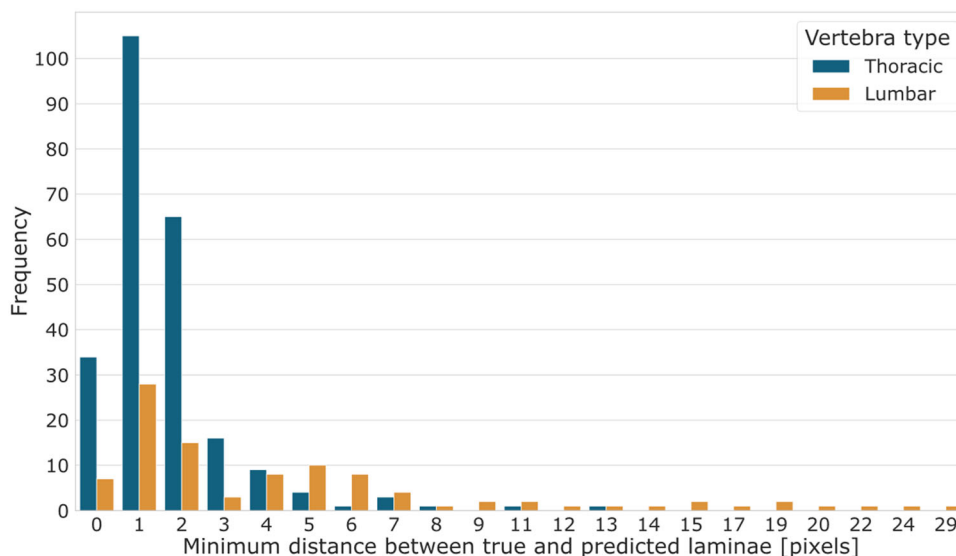
The CNN predicted the 340 COL in the initial test set with a mean  $d$  and standard deviation of  $2.7 \pm 3.7$  pixels. This means that on average, the closest predicted COL was roughly 3 pixels away from a given ground truth COL. The original ultrasound scans have a resolution of 0.2mm per pixel.<sup>17</sup> Since these scans were scaled down roughly one third of its original dimensions, a 3-pixel distance corresponds to roughly 1.8mm. The distribution of  $d$  for the COL, separated by thoracic (T1–T12) and lumbar (L1–L5) vertebrae, is illustrated in Fig. 6. The thoracic vertebrae were more consistently placed closer to their true COL with a  $d$  of  $1.8 \pm 1.6$  pixels, whereas the lumbar vertebrae were more frequently missed by the CNN, resulting in a  $d$  of  $5.0 \pm 5.8$  pixels. The average discrepancies in each anatomical axis between the true and predicted centers were also calculated. The COL were placed more accurately on average in the superficial-deep axis (0.6

pixels) than the lateral-medial (1.0 pixels) or the superior-inferior (2.0 pixels) axes.

Among the 60 spinal ultrasonograph measurement test set (10M, 50F), the mean age of participants was  $14.5 \pm 1.9$  years old (range: 10.8–17.6). A total of 104 MX-Cobb was reported in the clinical records (range:  $9^\circ$ – $45^\circ$ ). There were 118 MU-Cobb with AOR (range:  $8^\circ$ – $45^\circ$ ) and 107 SU-Cobb (range:  $8^\circ$ – $42^\circ$ ) measured. Table 1 shows the 107 paired comparison of SU-Cobb and the MU-Cobb with curve type and severity distribution. The MAD for all categories of measurements were below the clinically acceptable error of  $5^\circ$ . Additionally, the reliability of the semi-automatic method was excellent for MT curves (0.91), and good for all curves (0.87) and TL/L curves (0.81). Figure 7 illustrates the Bland-Altman plot of SU-Cobb vs. MU-Cobb. The SU-Cobb underestimated the MU-Cobb with bias and limits of agreement of  $-1.4^\circ$  ( $-10.1^\circ$ ,  $7.4^\circ$ ). The bias was significant as 0 was not contained within its 95% confidence interval [ $-0.54^\circ$ ,  $-2.25^\circ$ ]. The biases and limits of agreement were  $-1.6^\circ$  ( $-9.2^\circ$ ,  $5.9^\circ$ ) and  $-1.1^\circ$  ( $-11.2^\circ$ ,  $9.0^\circ$ ) for the MT and TL/L categories, respectively.

For MX-Cobb, there were 98 paired measurements with SU-Cobb and 95 paired measurements with MU-Cobb. The MAD  $\pm$  SD of the respective paired measurements was  $5.1^\circ \pm 4.1^\circ$  for MX-Cobb vs. SU-Cobb and  $2.9^\circ \pm 2.7^\circ$  for MX-Cobb vs. MU-Cobb. Table 2 shows the paired comparisons of 98 SU-Cobb vs. MX-Cobb and 95 MU-Cobb vs. MX-Cobb for all curves.

The CNN took approximately 15 h to train on our hardware. The average measurement time when using the CNN's lamina predictions was  $28.9 \text{ s} \pm 13.6 \text{ s}$  with the network taking less than a second on average to



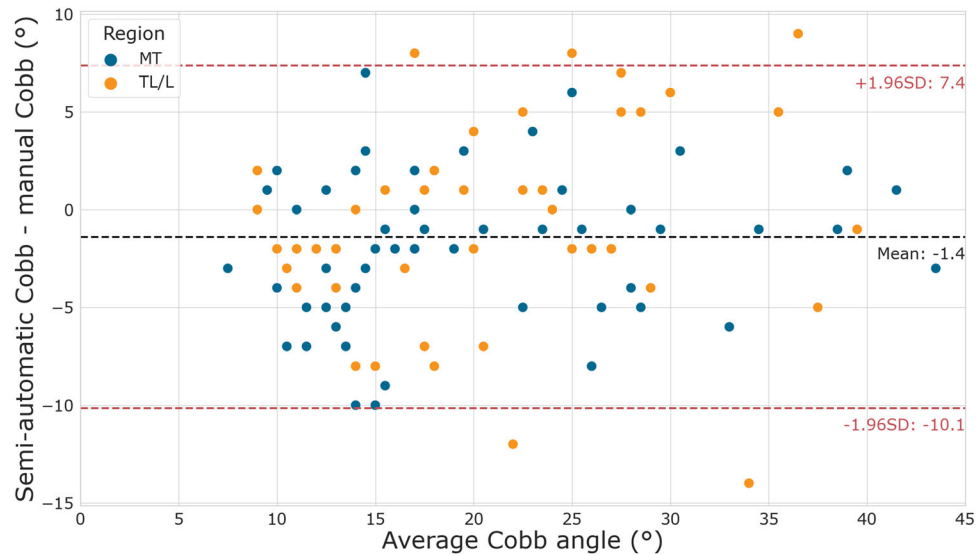
**FIGURE 6.** Histogram of the distance metric  $d$  for the 340 centers of lamina in the initial test set, separated by thoracic and lumbar vertebra type.



**TABLE 1. Results for SU-Cobb vs. MU-Cobb with AOR comparison on the measurement test set.**

Category	# detected	% clinical	MAD $\pm$ SD	SEM ( $^{\circ}$ )	ICC <sub>3,1</sub> [95% CI]	# missed
All	107	76	3.6 $^{\circ}$ $\pm$ 3.0 $^{\circ}$	1.1	0.87 [0.81,0.91]	11
MT	59	80	3.2 $^{\circ}$ $\pm$ 2.6 $^{\circ}$	0.8	0.91 [0.85,0.94]	6
TL/L	48	71	4.1 $^{\circ}$ $\pm$ 3.3 $^{\circ}$	1.4	0.81 [0.69,0.89]	5
Mild	79	75	3.4 $^{\circ}$ $\pm$ 2.8 $^{\circ}$	1.6	0.66 [0.52,0.77]	10
Moderate	28	79	4.2 $^{\circ}$ $\pm$ 3.4 $^{\circ}$	1.9	0.67 [0.41,0.83]	1

MT Main thoracic, TL/L thoracolumbar/lumbar, mild  $< 25^{\circ}$ , moderate  $\geq 25^{\circ}$ .



**FIGURE 7. Bland-Altman plot of SU-Cobb vs. MU-Cobb with AOR, separated by region, with bias (black line) and limits of agreement (red lines).**

**TABLE 2. Results for SU-Cobb vs. MX-Cobb and MU-Cobb vs. MX-Cobb comparison on the measurement test set.**

Comparison	# detected	% clinical	MAD $\pm$ SD	SEM ( $^{\circ}$ )	ICC <sub>3,1</sub> [95% CI]	# missed
SU vs. MR	98	60	5.1 $^{\circ}$ $\pm$ 4.1 $^{\circ}$	1.9	0.78 [0.69,0.85]	6
MU vs. MR	95	88	2.9 $^{\circ}$ $\pm$ 2.7 $^{\circ}$	0.8	0.90 [0.85,0.93]	9

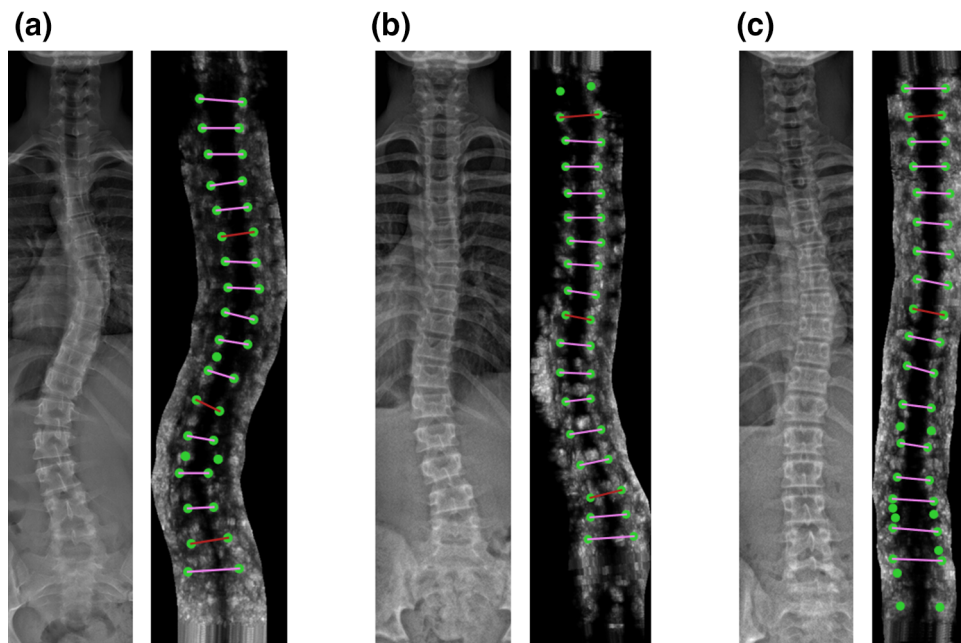
detect the COL, which was much faster than a manual ultrasound measurement (average 4 min per image<sup>22</sup>). Figure 8 shows the semi-automatic Cobb measurements on three test volumes.

## DISCUSSION

To our knowledge, this is the first CNN-based method for measuring coronal curvature severity on 3D spinal ultrasonographs. The other literature on the topic of coronal curvature measurement automation reports on the spinous process angle (SPA) and uses image processing techniques instead of a form of machine learning. No comparison has been performed

with the SPA automated extraction methods because they are a different measure of coronal curvature severity that typically underestimates the Cobb angle.

The  $d$  metric results on the validation set showed that the CNN performed well in terms of positioning COL close to the ground truth. Having a three-pixel positioning difference is little, especially since this is for three dimensions. The distance metric histogram showed that the CNN performed worse at identifying laminae in the lumbar region. This is because the lumbar vertebrae have larger bony flat surface area that can cause extraneous ultrasound reflections and produce brightness oversaturation. It is more difficult to distinguish the true COL in this area. Consequently, only lumbar vertebrae had a distance metric above 14



**FIGURE 8.** Semi-automatic Cobb angle measurement output examples with laminae (green circles) and relevant pairs for Cobb angle (red lines) plotted. Green circles that were left unpaired were deemed not true centers of lamina. Radiographs of the same visit are to the left of each labelled ultrasonograph. The examples have Cobb angles of: (a) 34° MT, 35° TL/L; (b) 15° MT, 25° TL/L; (c) 15° MT.

pixels, meaning that some lumbar laminae were not detected. On the other hand, the thoracic laminae were placed very accurately, with roughly 92% of their distance metrics within 3 pixels. This was reflected in the Cobb angle measurement results as well, as the accuracy and reliability of SU-Cobb was higher on the MT than the TL/L curves. The limits of agreement for TL/L curves (20.2°) also spanned a wider range than MT (15.1°).

There was less of a performance discrepancy in the different curve severity groups. Both ICCs were lower than the other categories of measurements. However, this is expected since grouping by curve severity reduces the variance of the data points. Therefore, the focus should instead be on the difference between the mild and moderate ICCs, which is essentially negligible. It should be noted that out of the 11 undetected curves in SU-Cobb, 10 were mild (6 MT, 4 TL/L) and 1 was moderate (1 TL/L). The mild curves with Cobb angle less than 16° are harder to detect because Cobb angles are defined for pairs of angles with opposing tilts. Therefore, if a manually measured curve involves a pair of laminae with a very shallow opposing tilt, the network may detect that same pair with no opposing tilt. Pairs of laminae without opposite tilt would not meet the required definition to warrant a Cobb angle measurement. There were 4/11 undetected mild curves under this situation. The other primary reason for non-detection was poor image quality. Two scans had

regions with a lack of brightness information and another scan was oversaturated with brightness, both of which made it more difficult to identify the COL automatically. This contributed to 6/11 curves being undetected and included in this group was the moderate curve that was also the only major curve that was missed. Finally, the last undetected curve (1/11) can be attributed to poor network performance.

The main challenge for developing a fully automated CNN-based method to measure the Cobb angle was the false positive laminae that made an automatic pairing algorithm nontrivial. Therefore, further post-processing of the lamina predictions is required to realize complete automation. This post-processing should combine using lamina pairs with the spinous process column to validate that the tilt angle of each lamina pair follows a smooth curve. Implementing these constraints may improve the accuracy of the measurement method. Another method of improving the accuracy is optimizing the CNN performance. This would consist of labelling more data for training and/or optimizing the training parameters of the CNN.

The comparison between SU-Cobb and MU-Cobb with AOR resulted in a MAD of 3.6° with 76% of measurements within clinical acceptance. The MAD between SU-Cobb and MX-Cobb was close to clinical acceptance (5.1°) with 60% of measurements within clinical acceptance. The gold standard for Cobb angle measurement is using a posteroanterior radiograph,

and so accurate and reliable results with radiographic measurements are needed for full clinical validation. Nevertheless, the work in this paper lays promising groundwork for a fully automated method that will meet those requirements. The main strength of the CNN-based algorithm is its ability to display the lamina segmentations. This means that it gives raters a strong starting point for measuring the Cobb angle. Manually adjusting the positions of the lamina predictions were not explored in this study, but this could easily be implemented in the program workflow, meaning that a quicker measurement time could be achieved while still providing raters a means to measure the Cobb angle accurately.

A possible reason for the higher inaccuracy in SU-Cobb vs. MX-Cobb is that these measurements were done without AOR. Having a radiograph from a previous visit to overlay with the current ultrasonograph helps in determining where the laminae should be located, particularly for the lumbar region. Zheng et al. found that using AOR significantly reduced measurement difference from  $4.6^\circ$  to  $2.7^\circ$  when comparing with radiographic measurements<sup>22</sup>. Similarly, the MAD for blinded SU-Cobb vs. MX-Cobb was  $5.1^\circ$ , which was close to the  $4.6^\circ$  from the blinded MU-Cobb vs. MX-Cobb of Zheng et al. Consequently, implementing a method of using AOR in the algorithm may improve accuracy. This would still reduce radiation exposure as only one radiograph at initial visit would be needed for follow-up visits to obtain only ultrasonographs for accurate Cobb angle measurement.

One limitation of this study is that severe curves above  $45^\circ$  Cobb angle were not investigated. The scan quality for severe curves is typically worse since these cases often have higher vertebral rotation, which more frequently results in only one lamina for a vertebra being visible. While these severe cases are rarer, a separate analysis should be conducted to investigate any potential systematic errors for this group. Another limitation is that the hyperparameters for the CNN were not optimized using a robust strategy. Different combinations were explored to improve the validation loss, but a structured strategy could not be conducted because of the significant computational costs it requires.

In conclusion, a novel 3D CNN-based algorithm for automatically detecting laminae on spinal ultrasonographs of children with AIS was developed for Cobb angle measurement. The measurements achieved good reliability when compared with manual ultrasonograph measurements and achieved excellent reliability in MT curves. Further improvements would consist of post-processing the network predictions by using other features of the spine to correct lamina placement and adding more labelled data for network training. Finally, complete automation and further validation with

manual radiographic measurements is planned to truly make ultrasound a more accessible imaging method for diagnosing and monitoring AIS, thereby reducing the risk of cancer in these children.

## ACKNOWLEDGMENTS

We would like to acknowledge the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Women and Children's Health Research Institute (WCHRI). The authors would also like to thank Industry Sandbox & AI Computing (ISAIC) at the University of Alberta for letting us use their supercomputer.

## CONFLICT OF INTEREST

No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

## REFERENCES

- <sup>1</sup>Brink, R., S. Wijdicks, I. Tromp, T. Schlosser, M. Kruyt, F. Beek, and R. Castelein. A reliability and validity study for different coronal angles using ultrasound imaging in adolescent idiopathic scoliosis. *J. North Am. Spine Soc.* 18:979–985, 2018.
- <sup>2</sup>Chen, W., E. Lou, and L. Le. A Reliable Semi-automatic Program to Measure the Vertebral Rotation Using the Center of Lamina for Adolescent Idiopathic Scoliosis. In: 5th Int. Conf. Biomed. Eng. Vietnam, vol. 46, pp. 159–162, 2015.
- <sup>3</sup>Chen, W., E. Lou, P. Zhang, L. Le, and D. Hill. Reliability of assessing the coronal curvature of children with scoliosis by using ultrasound images. *J. Child. Orthop.* 7:521–529, 2013.
- <sup>4</sup>Cobb, J. Outline for the study of scoliosis. *Am. Acad. Orthop. Surg.* 5:261–275, 1948.
- <sup>5</sup>Currier, D. Elements of Research in Physical Therapy. London: Williams & Wilkins, 1990.
- <sup>6</sup>Doody, M., J. Lonstein, M. Stovall, D. Hacker, N. Luckyanov, and C. Land. Breast cancer mortality after diagnostic radiography: findings from the U.S. scoliosis cohort study. *Spine.* 25:2052, 2000.
- <sup>7</sup>Ge, S., H. Zeng, and R. Zheng. Automatic measurement of spinous process angles on ultrasound spine images. In: *EMBC*, Montreal, Canada, 2020, pp. 2101–2104.
- <sup>8</sup>Kamnitsas, K., C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, D. Rueckert, and B. Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* 36:61–78, 2017.
- <sup>9</sup>Kingma, D., and J. Ba. Adam: a method for stochastic optimization. In: *ICLR*, San Diego, USA, 2015.
- <sup>10</sup>Lam, G., D. Hill, L. Le, J. Raso, and E. Lou. Vertebral rotation measurement: a summary and comparison of common radiographic and CT methods. *Scoliosis.* 3:16, 2008.

- <sup>11</sup>LeCun, Y., F. Huang, and L. Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In: *CVPR*, Washington DC, USA, 2004, pp. 97–104.
- <sup>12</sup>Panjabi, M., K. Takata, V. Goel, D. Federico, T. Oxland, J. Duranceau, and M. Krag. Thoracic human vertebrae. Quantitative three-dimensional anatomy. *Spine*. 16:888–901, 1991.
- <sup>13</sup>Ronneberger, O., P. Fischer, and T. Brox. U-Net: convolutional networks for biomedical image segmentation. In: *MICCAI*, Munich, Germany, 2015, pp. 234–241.
- <sup>14</sup>Simony, A., E. Hansen, S. Christensen, L. Carreon, and M. Andersen. Incidence of cancer in adolescent idiopathic scoliosis patients treated 25 years previously. *Eur. Spine J.* 25:3366–3370, 2016.
- <sup>15</sup>Trac, S., R. Zheng, D. Hill, and E. Lou. Intra- and inter-rater reliability of Cobb angle measurements on the plane of maximum curvature using ultrasound imaging method. *Spine Deform.* 7:18–26, 2019.
- <sup>16</sup>Van Goethem, J., A. Van Campenhout, L. van den Hauwe, and P. Parizel. Scoliosis. *Neuroimaging Clin. N. Am.* 17:105–115, 2007.
- <sup>17</sup>Vo, Q. N. Reliability of Three-dimensional Ultrasound Parameters and Their Correlation with the Progression of Adolescent Idiopathic Scoliosis [thesis]. Edmonton, AB: Univ. of Alberta, 2016.
- <sup>18</sup>Weinstein, S., L. Dolan, J. Cheng, A. Danielsson, and J. Morcuende. Adolescent idiopathic scoliosis. *The Lancet.* 371:1527–1537, 2008.
- <sup>19</sup>Wills, B., J. Auerbach, X. Zhu, M. Caird, B. Horn, J. Flynn, D. Drummond, J. Dormans, and M. Ecker. Comparison of Cobb angle measurement of scoliosis radiographs with preselected end vertebrae: traditional versus digital acquisition. *Spine*. 32:98–105, 2007.
- <sup>20</sup>Wong, Y., K. Lai, Y. Zheng, L. Wong, B. Ng, A. Hung, B. Yip, W. Chu, A. Ng, Y. Qiu, J. Cheng, and T. Lam. Is radiation-free ultrasound accurate for quantitative assessment of spinal deformity in idiopathic scoliosis (IS): a detailed analysis with EOS radiography on 952 patients. *Ultrasound Med. Biol.* 45:2866–2877, 2019.
- <sup>21</sup>Zheng, R., D. Hill, D. Hedden, J. Mahood, M. Moreau, S. Southon, and E. Lou. Factors influencing spinal curvature measurements on ultrasound images for children with adolescent idiopathic scoliosis (AIS). *PLoS ONE*. 13:e0198792, 2018.
- <sup>22</sup>Zheng, R., M. Young, D. Hill, L. H. Le, D. Hedden, M. Moreau, J. Mahood, S. Southon, and E. Lou. Improvement on the accuracy and reliability of ultrasound coronal curvature measurement on adolescent idiopathic scoliosis with the aid of previous radiographs. *Spine*. 41:404–411, 2016.
- <sup>23</sup>Zhou, G., D. Li, P. Zhou, W. Jiang, and Y. Zheng. Automating spine curvature measurement in volumetric ultrasound via adaptive phase features. *Ultrasound Med. Biol.* 46:828–841, 2020.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.