



# Radiographic scoliosis angle estimation: spline-based measurement reveals superior reliability compared to traditional COBB method

Peter Bernstein<sup>1</sup> · Johannes Metzler<sup>2</sup> · Marlene Weinzierl<sup>1</sup> · Carl Seifert<sup>1</sup> · Wadim Kisel<sup>1</sup> · Markus Wacker<sup>2</sup>

Received: 15 May 2020 / Revised: 25 July 2020 / Accepted: 17 August 2020 / Published online: 27 August 2020  
© The Author(s) 2020

## Abstract

**Introduction and objective** Although being standard for scoliosis curve size estimation, COBB angle measurement is well known to be inaccurate, due to a high interobserver variance in end vertebra selection and end plate contour delineation. We propose a stepwise improvement by using a spline constructed from vertebra centroids to resemble spinal curve characteristics more closely. To enhance precision even further, a neural net was trained to detect the centroids automatically.

**Materials & Methods** Vertebra centroids in AP spinal X-ray images of varying quality from 551 scoliosis patients were manually labeled by 4 investigators. With these inputs, splines were generated and the computed curve sizes were compared to the manually measured COBB angles and to the curve estimation obtained from the neural net.

**Results** Splines achieved a higher interobserver correlation of 0.92–0.95 compared to manual COBB measurements (0.83–0.92) and showed 1.5–2 times less variance, depending on the anatomic region. This translates into an average of 1° of interobserver measurement deviation for spline-based curve estimation compared to 3°–8° for COBB measurements. The neural net was even more precise and achieved mean deviations below 0.5°.

**Conclusion** In conclusion, our data suggest an advantage of spline-based automated measuring systems, so further investigations are warranted to abandon manual COBB measurements.

**Keywords** Radiographic · Scoliosis curve · Automatic measurement · COBB angle · Low image quality · Deep learning

## Introduction

### Drawbacks of traditional COBB method

COBB angle measurement on coronal whole-spine standing X-rays for the quantification of scoliosis curve severity has been implemented as gold standard in 1966 by the Scoliosis Research Society (SRS). The method itself is flawed by high inter- and intraobserver errors, ranging from 3° to 10° (95% CI), which can mainly be attributed to the manual nature of end vertebra selection and delineation [1]. In addition to this inaccuracy, the COBB method neglects important parts of the curve characteristics, e.g., apical vertebra translation.

Whole-spine X-ray images are currently produced at a varying expense in radiation exposure (ranging from conventional X-ray to modern low-dose imaging systems) with variable results. Thus, treatment decisions are being based on a high image acquisition variability and only moderate measurement reliability. However, it has been shown that the error originating from imaging modality ranges from none to below 2° [2, 3].

### Approaches to mitigate COBB method-associated inaccuracy

To level out observer-based errors, a certain amount of automation (e.g., for the selection of end vertebrae) was proposed. However, it has been realized that the original COBB method is not suitable for automation tasks. To overcome this problem, it would be necessary to focus on vertebra centroids that—following the original idea of Ferguson—offer the advantage of generating spinal curves that resemble global scoliotic spinal deformity more closely. In addition, they are less prone to measurement errors by smoothing curves through cubic

✉ Peter Bernstein  
bernspe@gmail.com

<sup>1</sup> Department for Orthopaedics and Traumatology, University Comprehensive Spine Center, University Hospital Dresden, Fetscherstrasse 74, 01307 Dresden, Germany

<sup>2</sup> Faculty of Informatics/Mathematics, HTW Dresden, Friedrich-List-Platz 1, 01069 Dresden, Germany

interpolation [1]. Thus far, first semiautomated computerized measurement projects have reported mixed results in terms of reliability. However, they all could show a good correlation between COBB angle and the investigated method [4–6].

In the advent of modern neural network (NN)-driven image processing, automatic segmentation, curvature estimation and even 3D-reconstruction from 2D data have come at reach. Wu et al. [7] were able to reduce the standard deviation of COBB angle estimation to  $1^\circ$  (circular mean absolute error of  $4^\circ$ ) by iterative training of a convolutional NN incorporating AP and lateral views from 154 patients. However, as their method requires a rather standardized approach with biplanar imaging, it might not be of use for the above-mentioned scenario with varying image quality and only AP imaging information. However, we used Wu's results as reference as it resembles the most precise measurement modality known to us.

The neglectable quality aspect of X-ray images for a proper scoliosis curve estimation was discussed in depth by Bonanni. He proposed an angle function over a structural curve drawn on either side of the spine to quantify the scoliotic deformation from AP X-ray or MRI images. He thus dropped non-necessary radiographic details (like endplates) and rather focused on the extraction of global spinal information (e.g., curve shape) [8, 9]. The group could show that the standard deviation of measurements could be brought below  $5^\circ$ —even in very low-quality images (54 patients). However, they had to rely on a fixed scheme with preselected landmarks carried out by trained scoliosis surgeons to do the measurements [9].

### Automatic vertebrae detection

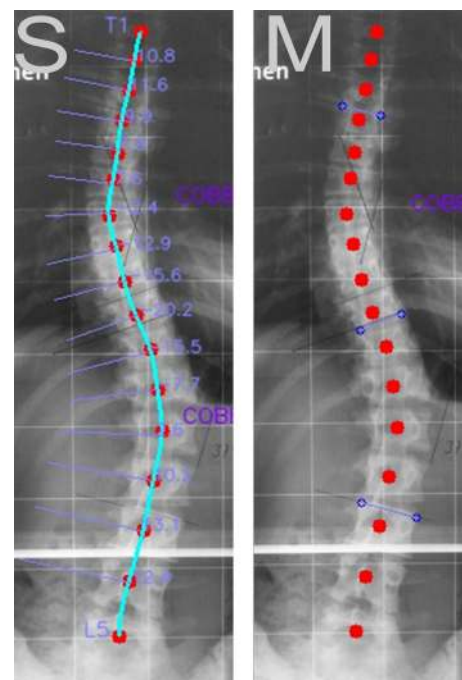
In contrast to CT and MRI images, the automatic detection of vertebrae in X-ray images is a very challenging computer vision task, due to multiple overlapping shadows of the ribs and pelvis, relatively weak contrast, and the need to identify all thoracic and lumbar vertebrae individually [10, 11]. Lecron, Xu and coworkers were the first to show the possibility of vertebrae detection in X-ray images using classical machine learning techniques [12, 13]. With the advance of convolutional NN (CNN) architectures, the opportunities have improved dramatically. De Carvalho, Al Arif and coworkers showed groundbreaking results with extraordinary accuracy in the prediction of vertebrae in X-ray images [2, 14]. However, these methods were limited to a small part of the spine (lumbar or cervical) and were trained on high-quality images.

### Aiming at higher precision at no additional cost

The aim of this project was to analyze, if there could be a solution for quantitative radiographic scoliosis assessment that:

- offers high precision measurement,
- is easy to use in an everyday clinical setting,
- does not depend on special hard- or software,
- even works in difficult situations (e.g., low-quality X-rays)—thus reducing the need of repeated radiation exposure,
- incorporates modern machine learning techniques (e.g., deep learning),
- had no additional cost.

We did not want to challenge alternative scoliosis imaging/screening modalities as ultrasound or photogrammetry [15–17].



**Fig. 1** S: Manually labeled vertebrae (red dots) were used to algorithmically construct a spline and perpendiculars on each level. Maximum tilted perpendiculars (=turning points of the curve) were defined as end vertebrae if the tilt exceeded  $10^\circ$ . M: COBB lines were drawn on the respective vertebral structures using a magnifying glass (not shown here) to facilitate precision (hollow blue circles, connected by blue line). Note that this image was actually photographed from a film-based spine radiograph with already existing lines drawn on the image

**Fig. 2** End vertebra detection by manual COBB measurement and spline-based computation. All investigators (C–W) showed the same pattern, manual COBB and auto-spline were identical for the NN (N) as those values were retrieved algorithmically. Mean standard deviation between all observers was 0.45 vertebrae for spline-based and 0.56 vertebrae for manual COBB-based measurements

## Material and methods

### Image material and labeling

This study has been approved by our institutional review board (#EK339092018). Standing coronal X-rays displaying the whole vertebral column or at least the most scoliotic part of the thoracic or lumbar column originated from different individuals with any form of idiopathic scoliosis subjected to bracing and without prior treatment. Image quality ranged from digitally acquired (most) to film or even paper scans. Images were saved with a random 8-digit identifier alias.

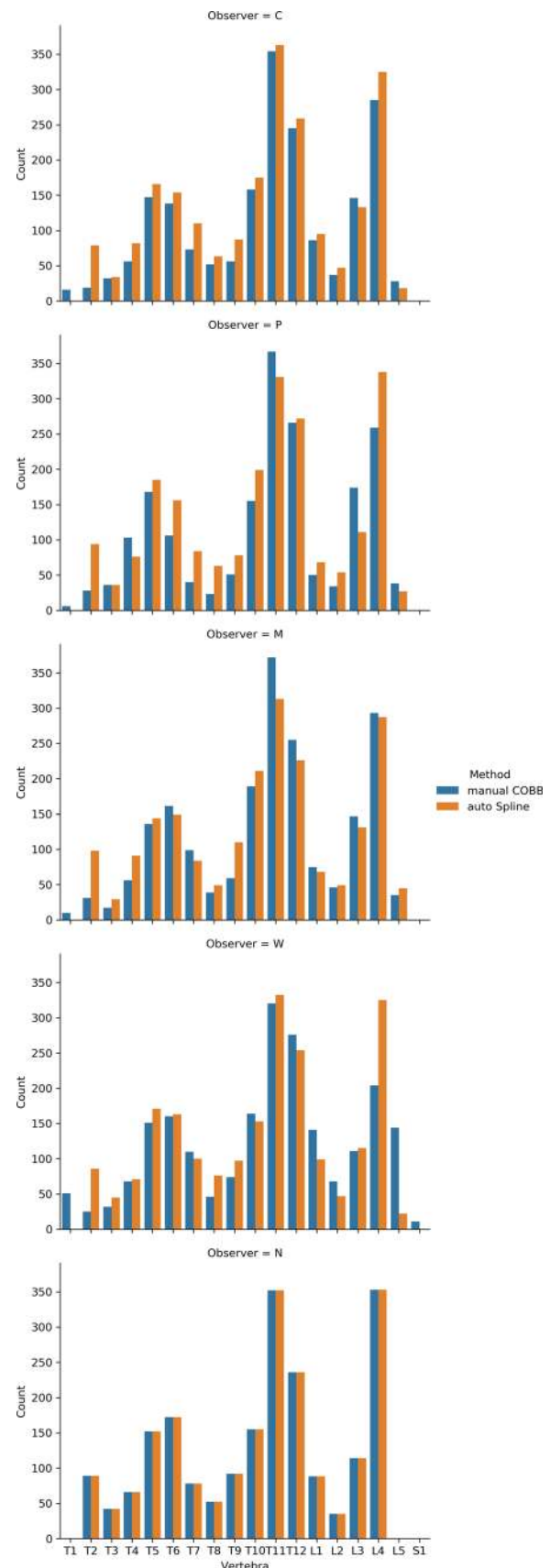
The images were split into 30-image-sized batches (551 images in total) that were successively labeled with a specially developed label tool by a group of 4 analyzers after appropriate training but with different educational levels, ranging from medical student (M), orthopedic residents in training (C and W) to experienced scoliosis surgeon (P).

Labeling included centroid identification of each vertebra, start and end vertebra description and angle line drawing following COBBs method (Fig. 1 and supplemental material for the workflow). Precision of the labeling task was supported by a magnifying glass in a second window. The selected centroids were used to generate a spline using cubic interpolation automatically. Perpendicular lines on this spline were generated at each vertebral level. The most tilted lines (with respect to the horizon) were used to calculate the corresponding COBB angle. Only values  $> 10^\circ$  were processed further.

Allocation of measurements to three distinct spinal regions was labeled as “high thoracic,” “thoracic” and “lumbar” if the curve was located above Th5, L2 or below, respectively [18].

### Neural net training

Manually labeled vertebrae of all 4 analyzers were used to train a NN via deep learning. Splines were generated from vertebral centroids that were detected by the NN. The architecture of the NN is based on UNet, that is a standard approach in medical image segmentation [19]. It comprises an encoder/decoder structure, incorporating the special feature of skip connections between the encoder and decoder in order to retain important information over several layers [20]. To obtain better results as the original UNet, we implemented masked loss functions that learn to penalize areas

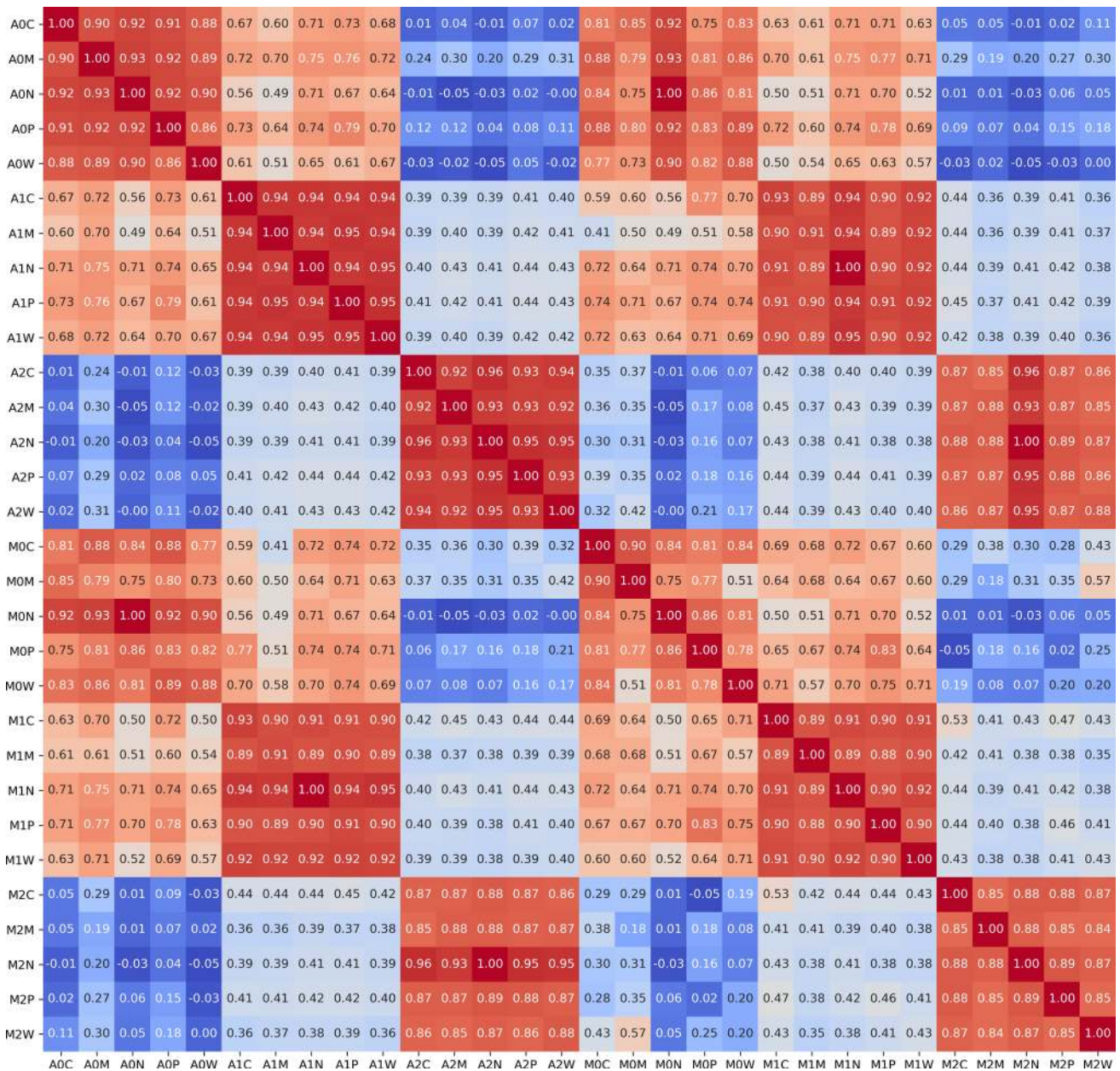




outside the vertebrae. The whole dataset consisted of 571 X-ray images. In training process, we randomly split the data into chunks of 457 and 114 images for training and validation, respectively. The size of X-ray images was resized to 200×500 pixels in width x height. We also normalized the labeled vertebrae centroids as a percentage of the scaled image.

### Statistics

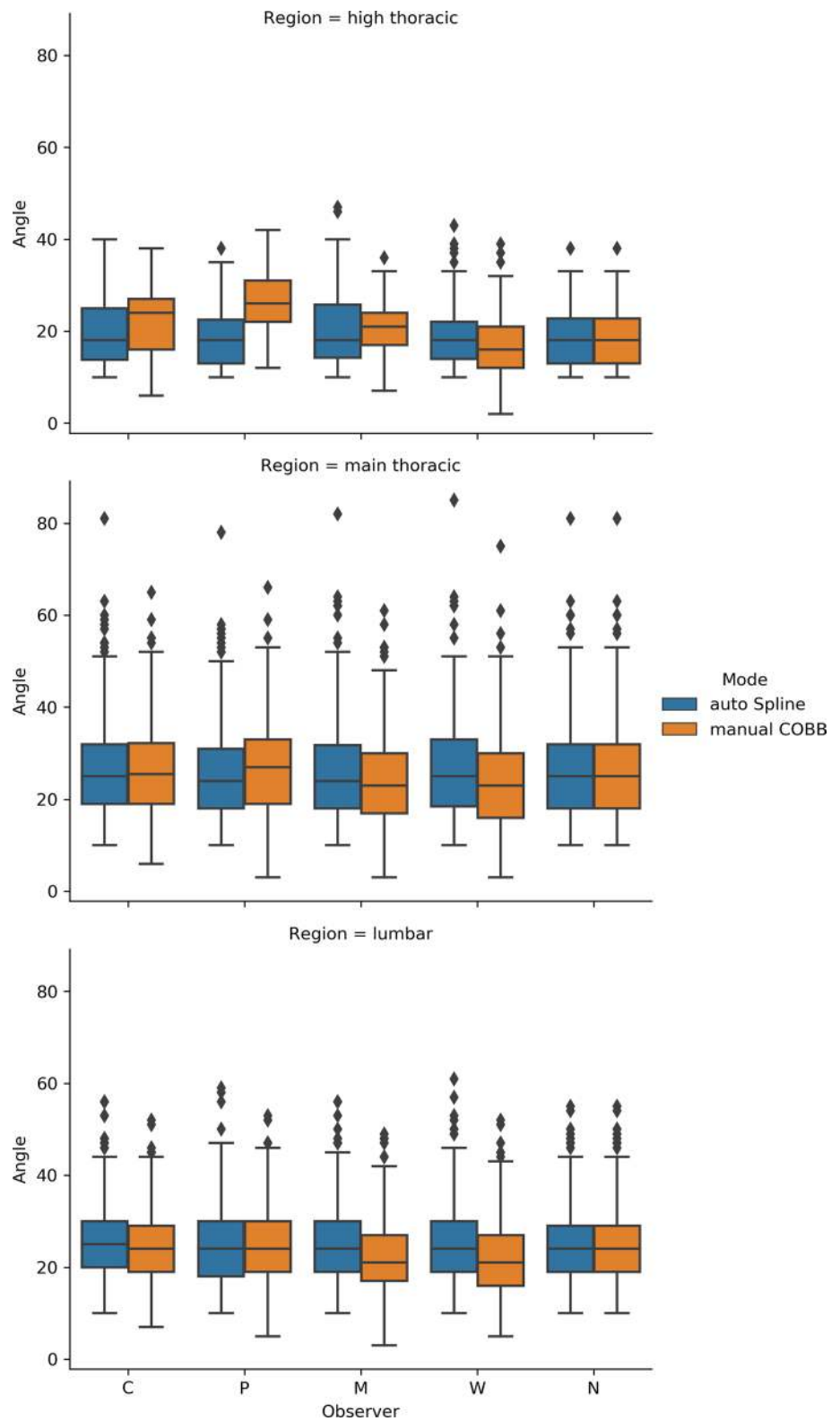
For the statistical analysis, we used standard methods such as Pearson’s correlation coefficient and student’s *t* test. In case of missing values, the respective row was deleted from the observer’s dataset.



**Fig. 3** Correlation heatmap showing Pearson correlation coefficients between different methods (A=automatic spline, M=Manual COBB) at different spinal regions (O=high thoracic, I=thoracic, 2=lumbar) for different observers (C, M, P, W=human investigators, N=NN). A higher correlation could be observed for interobserver

automatic spline computations (upper left diagonal) compared to interobserver manual COBB measurements (lower right diagonal) or between automated spline and manual COBB measurements (upper right diagonal)

**Fig. 4** Box plots depicting the median COBB angles of the complete dataset at three distinct anatomic regions. Blue boxes show spline-based COBB angle computations which have similar values for every investigator, whereas orange boxes depict manually measured COBB angles that display greater variability between investigators



## Results and discussion

### End vertebra selection

Selection of end vertebrae revealed the typical double curve scoliotic pattern with peaks at Th5/6, Th11 and L4 for any investigator as well as the NN (Fig. 2).

Interobserver differences were 0.56 vertebrae over all measurements for manual end vertebra selection. When the selection was left to the spline algorithm, it was only a little more uniform (0.45 vertebrae deviation). Interestingly, the deviation of the centroids was not zero, which shows that even the simple task of setting centroid points onto vertebrae yields observer-dependent results.

Another source of inconsistency is that observers were required to indicate the range of vertebrae visible on each X-ray (e.g., from T2 to L5). As vertebral recognition is not always easy, due to variable formation of the 12<sup>th</sup> rib, it might explain the observed variability. It seems to confirm Vrtovec's view being end vertebra selection as one main source of error in COBB angle determination [1].

To this end, we investigated more closely if fuzzy end vertebra determination has an impact on measurement variance for spline and NN approaches that estimate the global curve characteristics.

### Interobserver correlations

In our analysis, we have found out that spline-based COBB measurement showed higher correlations than manual COBB measurements (Fig. 3). This phenomenon was especially pronounced in the border regions (high thoracic and lumbar). Mean spline-based measurements intraclass interobserver correlation coefficients ranged from 0.92 (high thoracic) to 0.95 (main thoracic and lumbar), whereas traditional COBB-based measurements yielded coefficients ranging from 0.83 (high thoracic) to 0.89 (lumbar) and 0.92 (main thoracic). De Cavalho et al. achieved an interobserver correlation of 0.88–0.9, using high-quality images with predefined end vertebrae [2]. However, they did not convey if they observed the same region dependence nor did any other author. We hypothesize that a certain amount of variability, especially in the case of manual COBB measurement, might result from kyphosis-dependent obfuscation of vertebral structures. This uncertainty translates directly into bad correlation. Consequently, interclass (spline versus COBB) correlation was rather poor, ranging from 0.84 (high thoracic) to 0.92 (main thoracic).

Interestingly, any observer (manual or spline)—NN correlations were always above 0.9, which is another view on

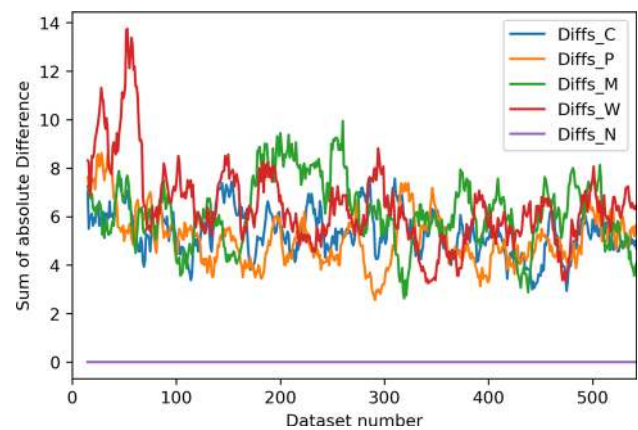
the nn's possibilities to mitigate interobserver variance, even in difficult anatomic regions. Further analysis will reveal how correlation can be translated into angle measurement error.

### Distribution of differences between investigators

Our dataset was derived from X-ray images of patients subjected to brace treatment, typically in the range from 20° to 40°. Especially in this range, the detection of minor angle changes is most important to optimize conservative treatment results. Manual COBB angle measurement is still considered gold standard for scoliosis curve estimation, despite of a known inaccuracy of 3°–10° [1]. This high deviation can obviously lead to erroneous decisions, even surgically.

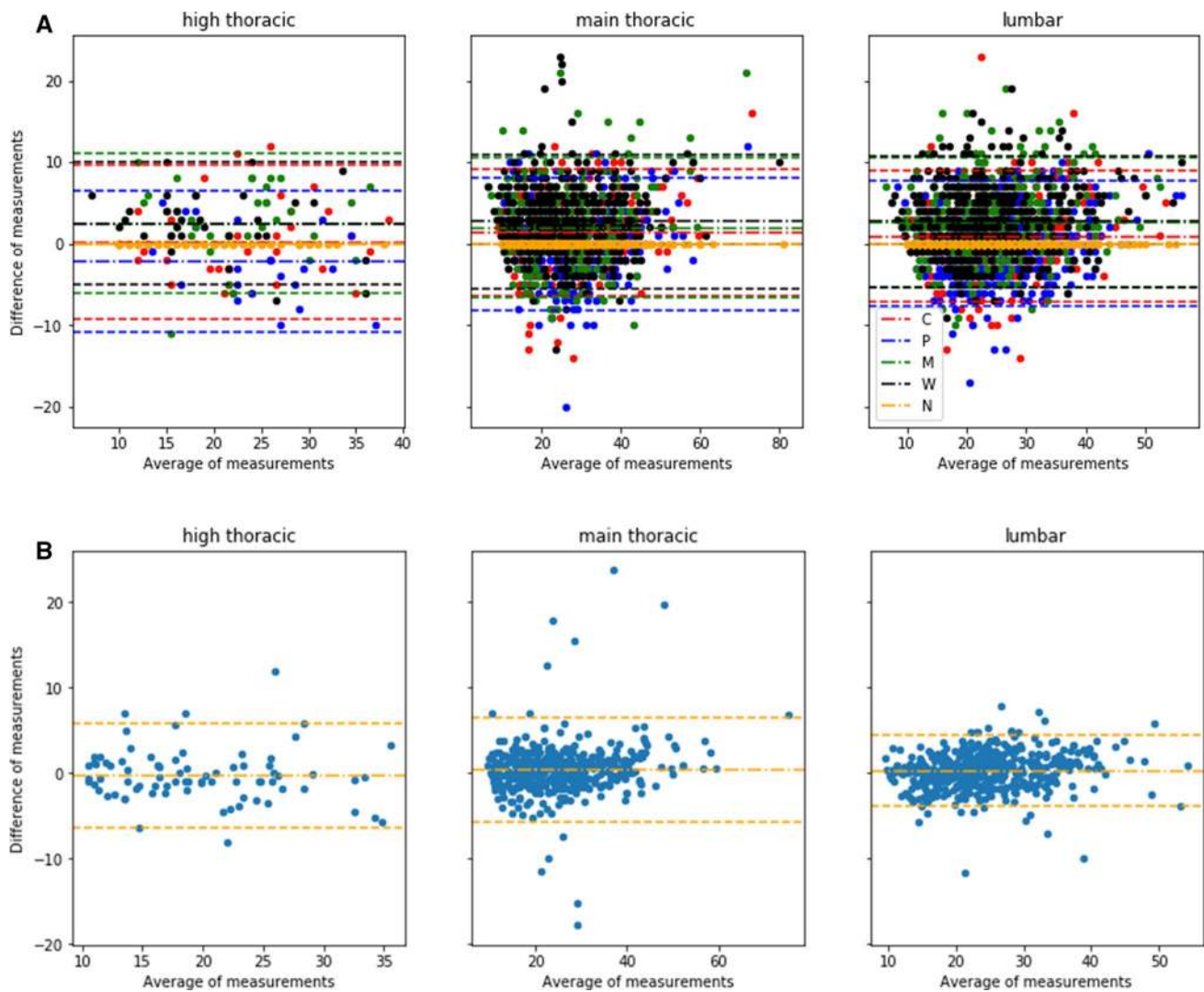
As expected, we had to confirm the inaccurate nature of manual COBB angle measurement. We saw region-dependent interobserver medians spanning 8° (high thoracic), 4° (thoracic) and 3° (lumbar) of deviation; see Fig. 4. Using standardized measurement setups, other authors were able to decrease the average angle difference to 2°–3° [2, 21].

To answer the question if there was a training effect in measurement precision, a rolling mean evaluation of the above-mentioned deviations with a window size of 16 datasets had been performed (Fig. 5). A trend toward deviation reduction over the dataset could be observed only for one observer (W). All investigators retained an error range, summing up to 4°–8° for all spinal regions. We therefore conclude that manual COBB measurement cannot be ameliorated by training. This disadvantage is caused by the method!



**Fig. 5** Rolling window calculation of mean absolute measurement difference between spline-based and manual COBB method, summed for the three spinal regions and plotted against the time course of dataset evaluations. Learning patterns (red) as well as unsteady precision in the beginning (green) or relative stability (blue and orange) can be observed. However, there remains a constant deviation of about 4°–8° that cannot be trained away





**Fig. 6** Bland–Altman plot of spline-based measurements and manual COBB measurements (a) and NN derived curve estimations and all human observations (b). Means are plotted against their differences.

Positive deviation = spline algorithm (a) or NN (b) overestimates the curve magnitude. Dashed lines = 1.96-fold standard deviation, dash-dotted line = mean deviation

Spline-based curve size estimation showed an overall constant median angle, deviating not more than  $1^\circ$  between observers in any spinal region (Fig. 4).

### Stability of measurements

In order to demonstrate the stability of measurements over the angle range of the dataset, Bland–Altman plots were generated (Fig. 6a and additional Table 1). We see that rather unexperienced analyzers significantly tended to manually

underestimate (C, W, M, positive values), whereas the experienced analyzer (P) only overestimated the high thoracic curve (negative value) and stayed close to the spline values in the remaining spine regions, despite some considerable high outliers (blue dots). However, no systematic bias for lower or higher angles could be found, demonstrating the rigidity of the spline method.

The above-mentioned data have demonstrated, that the NN achieves high correlations with all observers. The stability of measurements is further displayed in the

**Table 1** Mean deviations of spline-based curve computation and manual COBB measurement for each investigator and respective *P* value. Positive values=spline algorithm overestimates curve magnitude, negative values=manual COBB measurement overestimates curve magnitude

	Mean deviation	<i>P</i> value
<i>High thoracic</i>		
C	0.20	0.807
P	−2.17	0.024
M	2.47	0.001
W	2.49	0.001
N	0.00	0.000
<i>Main thoracic</i>		
C	1.42	0.000
P	−0.07	0.738
M	1.96	0.000
W	2.72	0.000
N	0.00	0.000
<i>Lumbar</i>		
C	0.89	0.000
P	−0.02	0.907
M	2.58	0.000
W	2.70	0.000
N	0.00	0.000

Bland–Altman plot in Fig. 6b. There, the nn’s measurements have been plotted against all human obtained values (manual measurements and spline measurements). It can be seen, that the confidence interval for any region is narrower, comprising only one half to two-thirds of the confidence intervals of spline versus manual COBB measurements. The reasons for this stability are explained in the next section.

### The strength of nn-based measurements

Our NN achieves a sensitivity of 0.87 and a relative pixel error of 0.011 for detecting the correct vertebrae in an AP X-ray of the spine (Fig. 7, Table 3). This is three times more precise than the NN of Wu et al. who reported an error of

0.0398 [7]. This translates to an angle deviation of  $< 1^\circ$  for any region with high correlations in the main thoracic and lumbar spine (Table 2), which is also better than Wu’s error of  $1^\circ$ . Again, we need to point out that Wu used biplanar standardized imaging—we have used only AP images of heterogeneous quality (Table 3).

### Conclusion

In conclusion, we were able to show, that even in the case of poor image quality, the reliability of COBB angle measurement can be substantially raised by

- investigating the spinal curve in a more global fashion by using a vertebral centroid generated spline
- automating vertebra detection by neural network which was trained by several observers

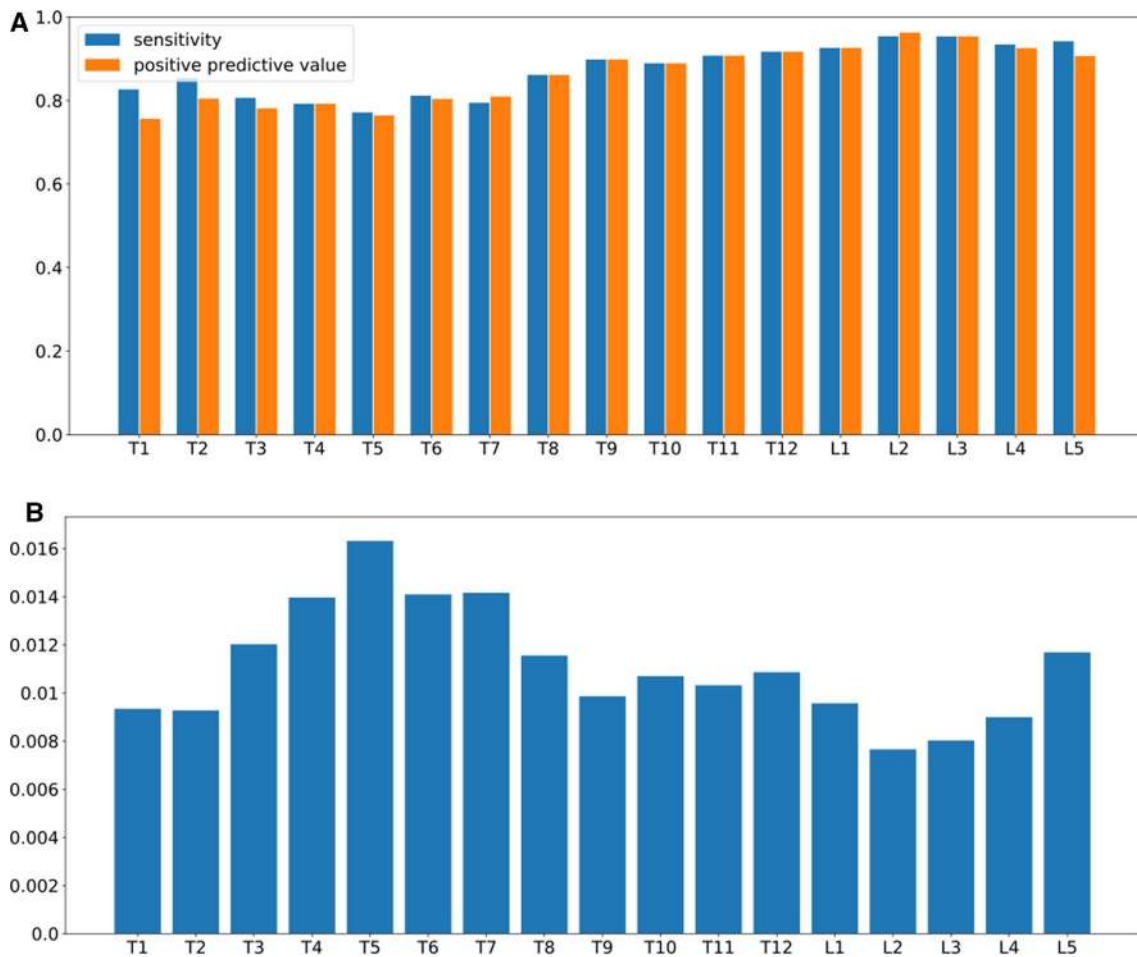
Measurement error can be brought down below  $0.5^\circ$ , which is less than one-tenth of traditional COBB measurement error.

Limitations of our current approach include the uncertainty in the higher thoracic spine which will be subject to optimization in our further investigations. We will also need to include various other types of spinal X-ray imaging to enhance vertebra recognition even further. Also, the angle range as well as the scoliosis spectrum (e.g., malformations) has to be exploited further, as severe rotated curves might exhibit different geometric properties.

From the data of this work and by taking into account all the criticism already mentioned over the years, manual COBB measurement cannot be recommended for scoliosis curve estimation any more. It is flawed by considerable inter-observer deviations that can be eliminated completely by machine-based diagnosis, even if the image material is of poor quality.

The implementation of this knowledge into clinical practice is now mandatory. However, there are still further steps to accomplish. The current version of NN-based spline generation worked well with a limited set of images.





**Fig. 7** A: Sensitivity and positive predictive values (PPV) of the estimated vertebrae locations from the NN. B: relative pixel error of the estimated vertebrae locations from the NN

**Table 2** Mean deviation of NN versus gathered human COBB angle measurements. Positive values=NN overestimates COBB angle. Correlation=Pearson correlation of NN measurements with all human manual and spline-based values

	Mean deviation	<i>P</i> value	Correlation
High thoracic	−0.23	0.514	0.89
Main thoracic	0.41	0.006	0.96
Lumbar	0.30	0.002	0.97

**Table 3** Sensitivity and positive predictive values (PPV) of the estimated vertebrae locations from the NN

Sensitivity	Positive predictive value	Relative pixel error	Absolute pixel error
0.87	0.86	0.011 ± 0.017	4.28 ± 7.98

Future model training has to include a broader image database, e.g., including operated spines, early-onset and adult

scoliosis—to mention just a few. We plan to provide open access of the model to the scientific and medical community.

**Acknowledgements** The authors thank Thomas Neumann who shared his experience in graphical neural network construction.

**Authors' contributions** PB contributed to conceptualization, methodology, software, investigation, validation, formal analysis, resources, data curation, writing, visualization, supervision and project administration; JM contributed to methodology, software, formal analysis, data curation, writing—draft and visualization; MW contributed to investigation and data curation; CS contributed to investigation and writing—review; WK contributed to investigation and writing—review; MW contributed to resources, writing—review and supervision.

**Funding** This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors. Open Access funding provided by Projekt DEAL.

**Availability of data and material** Raw data are published at Mendeley data repository: <https://doi.org/10.17632/ssfmrw2fg9.2>.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Ethics approval** This study has been approved by the institutional review board of the University Hospital (#EK339092018).

**Consent for publication** Only X-ray images were used in a blinded fashion. In the study, a trace toward the individual participant could not be made, so that the obtainment of an informed consent was not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Vrtovec T, Pernuš F, Likar B (2009) A review of methods for quantitative evaluation of spinal curvature. *Eur Spine J* 18:593–607
- De Carvalho A, Vialle R, Thomsen L et al (2007) Reliability analysis for manual measurement of coronal plane deformity in adolescent scoliosis. Are 30 × 90 cm plain films better than digitized small films? *Eur Spine J*. <https://doi.org/10.1007/s00586-007-0437-4>
- Wills BPD, Auerbach JD, Zhu X et al (2007) Comparison of Cobb angle measurement of scoliosis radiographs with preselected end vertebrae: Traditional versus digital acquisition. *Spine (Phila Pa 1976)*. <https://doi.org/10.1097/01.brs.0000251086.84420.d1>
- Dutton KE, Jones TJ, Slinger BS et al (1989) Reliability of the Cobb angle index derived from traditional and computer-assisted methods. *Australas Phys Eng Sci Med* 12:16
- Tanure MC, Pinheiro AP, Oliveira AS (2010) Reliability assessment of Cobb angle measurements using manual and digital methods. *Spine J*. <https://doi.org/10.1016/j.spinee.2010.02.020>
- Shea KG, Stevens PM, Nelson M et al (1998) A comparison of manual versus computer-assisted radiographic measurement: Intraobserver measurement variability for Cobb angles. *Spine (Phila Pa 1976)*. <https://doi.org/10.1097/00007632-199803010-00007>
- Wu H, Bailey C, Rasoulinejad P, Li S (2018) Automated comprehensive Adolescent Idiopathic Scoliosis assessment using MVC-Net. *Med Image Anal*. <https://doi.org/10.1016/j.media.2018.05.005>
- Bonanni PG (2017) Contour and angle-function based scoliosis monitoring: relaxing the requirement on image quality in the measurement of spinal curvature. *Int J Spine Surg*. <https://doi.org/10.14444/4022>
- Papalioidis DN, Bonanni PG, Roberts TT et al (2017) Computer assisted Cobb angle measurements: a novel algorithm. *Int J Spine Surg*. <https://doi.org/10.14444/4021>
- Huang SH, Chu YH, Lai SH, Novak CL (2009) Learning-based vertebra detection and iterative normalized-cut segmentation for spinal MRI. *IEEE Trans Med Imaging*. <https://doi.org/10.1109/TMI.2009.2023362>
- Glocker B, Zikic D, Konukoglu E, et al (2013) Vertebrae localization in pathological spine CT via dense classification from sparse annotations. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)
- Lecron F, Benjelloun M, Mahmoudi S (2012) Fully automatic vertebra detection in x-ray images based on multi-class SVM. In: *Medical imaging 2012: image processing*
- Xu X, Hao HW, Yin XC, et al (2012) Automatic segmentation of cervical vertebrae in X-ray images. In: *Proceedings of the international joint conference on neural networks*
- Al Arif SMMR, Knapp K, Slabaugh G (2018) Fully automatic cervical vertebrae segmentation framework for X-ray images. *Comput Methods Programs Biomed*. <https://doi.org/10.1016/j.cmpb.2018.01.006>
- Brink RC, Wijdicks SPJ, Tromp IN et al (2018) A reliability and validity study for different coronal angles using ultrasound imaging in adolescent idiopathic scoliosis. *Spine J*. <https://doi.org/10.1016/j.spinee.2017.10.012>
- Takács M, Orlovits Z, Jáger B, Kiss RM (2018) Comparison of spinal curvature parameters as determined by the ZEBRIS spine examination method and the Cobb method in children with scoliosis. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0200245>
- Knott P, Pappo E, Cameron M, et al (2014) SOSORT 2012 consensus paper: reducing x-ray exposure in pediatric patients with scoliosis. *Scoliosis*
- Lenke LG, Edwards CC 2nd, Bridwell KH (2003) The Lenke classification of adolescent idiopathic scoliosis: how it organizes curve patterns as a template to perform selective fusions of the spine. *Spine (Phila Pa 1976)* 28:S199–S207. <https://doi.org/10.1097/01.BRS.0000092216.16155.33>
- Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*
- Stokes IAF, Aronsson DD (2006) Computer-assisted algorithms improve reliability of king classification and Cobb angle measurement of scoliosis. *Spine (Phila Pa 1976)*. <https://doi.org/10.1097/01.brs.0000203708.49972.ab>

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