Original Article

J Korean Soc Radiol 2024;85(5):926-936
https://doi.org/10.3348/jksr.2023.0111
eISSN 2951-0805

JOURNAL of THE KOREAN SOCIETY of RADIOLOGY

Radiographic Analysis of Scoliosis Using Convolutional Neural Network in Clinical Practice

컨볼루션 신경망을 이용한 척추측만증의 방사선학적 분석의 임상 적용

Ha Yun Oh, MD[1], Tae Kun Kim, MD[1,2], Yun Sun Choi, MD[1]*, Mira Park, MD[3], Ra Gyoung Yoon, MD[1], Jin Kyung An, MD[1]

[1]Department of Radiology, Nowon Eulji Medical Center, Eulji University, Seoul, Korea
[2]Department of Radiology, Armed Forces Seoul District Hospital, Seoul, Korea
[3]Department of Preventive Medicine, School of Medicine, Eulji University, Daejeon, Korea

Received September 4, 2023
Revised November 9, 2023
Accepted January 25, 2024
Published Online April 29, 2024

***Corresponding author**
Yun Sun Choi, MD
Department of Radiology, Nowon Eulji Medical Center, Eulji University, 68 Hangeulbiseok-ro, Nowon-gu, Seoul 01830, Korea.

Tel 82-2-970-8290
Fax 82-2-970-8346
E-mail cys0128@eulji.ac.kr

Purpose To assess the reliability and accuracy of an automated Cobb angle measurement (ACAM) using a convolutional neural network (CNN) for scoliosis evaluation and to compare measurement times.

Materials and Methods ACAM was applied to spine radiographs in 411 patients suspected of scoliosis. Observer 1 (consensus of two musculoskeletal radiologists) and observer 2 (a radiology resident) measured Cobb angle (CA). CA measurements were categorized using observer 1's measurements as the reference standard. Inter-observer reliability and correlation were assessed using intraclass correlation coefficient (ICC) and Spearman's rank correlation coefficient, respectively. Accuracy and measurement time of ACAM and observers were evaluated.

Results ACAM demonstrated excellent reliability and very high correlation with observer 1 (ICC = 0.976, Spearman's rank correlation = 0.948), with a mean CA difference of 1.1. Overall accuracy was high (88.2%), particularly in mild (92.2%) and moderate (96%) scoliosis. Accuracy was lower in spinal asymmetry (77.1%) and higher in severe scoliosis (95%), although the CA was lower compared to the observers. ACAM significantly reduced measurement time by nearly half compared to the observers ($p < 0.001$).

Conclusion ACAM using CNN enhances CA measurement for assessing mild or moderate scoliosis, despite limitations in spinal asymmetry or severe scoliosis. Nonetheless, it substantially decreases measurement time.

Index terms Scoliosis; Cobb Angle; Convolutional Neural Network; Radiography

JOURNAL of THE KOREAN SOCIETY of RADIOLOGY

INTRODUCTION

Scoliosis is defined as a lateral spinal curvature with a Cobb angle (CA) ≥ 10° (1). CA is an objective measurement used to determine the severity of scoliosis. It is the angle formed by the intersection of two lines, one parallel to the endplate of the most tilted superior end vertebra and the other parallel to the endplate of the most tilted inferior end vertebra (Fig. 1) (1). Furthermore, it is the main standard for the diagnosis, monitoring, therapeutic planning, and epidemiologic analysis of scoliosis (1). Therefore, CA measurement should be precise and reproducible (2).

However, manual measurement of CA is time-consuming and unreliable, with high inter- and intra-observer variability, and different selections of end vertebrae are a major source of errors (3-5). Therefore, it is difficult for inexperienced observers to accurately measure CA. In previous studies, the variability of CA measurements was reported to range from 3° to 10° (6, 7). With the development of computer technology, the use of artificial intelligence (AI) in measuring CA has gained popularity in clinical practice, and it has shown improved measurement precision and good correlation with manual measurements (8-12). The use of AI-based diagnostic supporting software is expected to improve the reduction of CA measurement time and enhance reproducibility in scoliosis evaluation (3, 6, 10, 11, 13-16). However, to our knowledge, there has been no research conducted utilizing this kind of software in Korea. In the present study, we developed a convolutional neural network (CNN), compared the reliability and accuracy of an automated CA measurement (ACAM) with the CA measurement performed by radiologists for scoliosis assessment, and investigated the usefulness of a CNN in clinical practice.

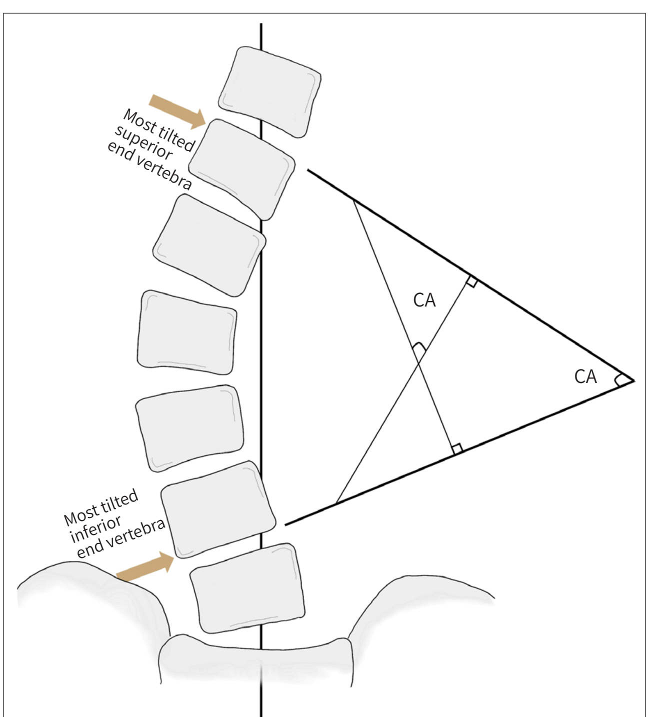


Fig. 1. Diagram of CA measurement. CA is determined by the intersection of the two lines. One line is parallel to the superior endplate of the most tilted superior end vertebra and the other line is parallel to the inferior endplate of the most tilted inferior end vertebra. CA is defined as the angle between the tangential lines or between two lines drawn perpendicular to the tangents.
CA = Cobb angle

JOURNAL of THE KOREAN SOCIETY of RADIOLOGY

MATERIALS AND METHODS

This retrospective study was approved by our Institutional Review Board, and the requirement for consent was waived owing to its retrospective design (IRB No. 2022-10-019).

STUDY POPULATION

Patients with suspected scoliosis who underwent spine radiographs at our hospital between January 1, 2016, and June 30, 2022, were included in the present study. A total of 493 patients were included during this period. Patients with a CA indicating cervical spine (n = 35), severe kyphosis (n = 4), or severe degenerative spondylosis (n = 35) and those who underwent spinal surgery or cement augmentation (n = 4), compression fracture (n = 3), or bony anomaly (n = 1) were excluded. During the study period, there were no spine radiographs found of a patient with hemivertebra. Finally, 411 patients were included in the study.

IMAGING MODALITIES

All spine radiographs included whole-spine (C-T-L-S spine) images taken either in an anterior–posterior or posterior–anterior view of the coronal plane. Spine radiographs were obtained using Discovery XR656G2 (GE Healthcare, Milwaukee, WI, USA) and Ysio (Siemens Healthcare, Erlangen, Germany) digital radiography. Each radiograph was taken under the following conditions: 60–70 KVP, 20–35 mAs for the Discovery XR656G2, and 80–99 KVP, 21–30 mAs for the Ysio. The stored Digital Imaging and Communications in Medicine (DICOM) images comply with the DICOM 3.0 standard, with resolutions of 2022 × 2022 pixels and 2860 × 2874 pixels, respectively.

IMAGING ANALYSIS

Three radiologists (radiologist 1, with 20 years of musculoskeletal radiology experience; radiologist 2, with 7 years of experience [1 year of musculoskeletal radiology]; and radiologist 3, a trained radiology resident with 3 years of experience) retrospectively reviewed the spine radiographs. We measured CA using the picture archiving and communication system tool in the thoracic, lumbar, and thoracolumbar vertebrae. In cases of double thoracolumbar scoliosis, the larger of the two measured values was utilized for analysis. The measured CAs were divided into four groups based on the measurements taken by observer 1 (consensus of radiologist 1 and 2) as a reference standard: spinal asymmetry (CA < 10°), mild scoliosis (CA 10–25°), moderate scoliosis (CA 25–40°), and severe scoliosis (CA > 40°) (7, 17). All measurements were repeated once by observers 1 and 2 at least one month after the initial measurements were taken.

To measure the CA using a CNN, each spine radiograph was extracted after DICOM images were sent and measured using the DeepSPINE-SC-01 software program (version 1.3.6, DEEPNOID, Seoul, Korea). This software performed preprocessing by utilizing the original DICOM images to execute normalization, contrast enhancement, and resizing. Radiographs underwent a normalization process to ensure consistent intensity levels across all images. The Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm was applied to enhance image quality and to improve the visibility of important features. Additionally, images were resized to a standardized dimension of 512 × 512 pixels, while maintaining the original image's aspect ratio through zero-padding., An i5-13600 (14 Core, 2.7 GHz) CPU, GTX1060 GPU for computing, 32 GB RAM, and Ubuntu 20.04 operating system were utilized for processing. This model analyzed the most tilted superior and inferior end vertebrae, and it measured the CA (Fig. 2). We also evaluated the respective measurement times of ACAM, observer 1, and observer 2 in measuring CA.

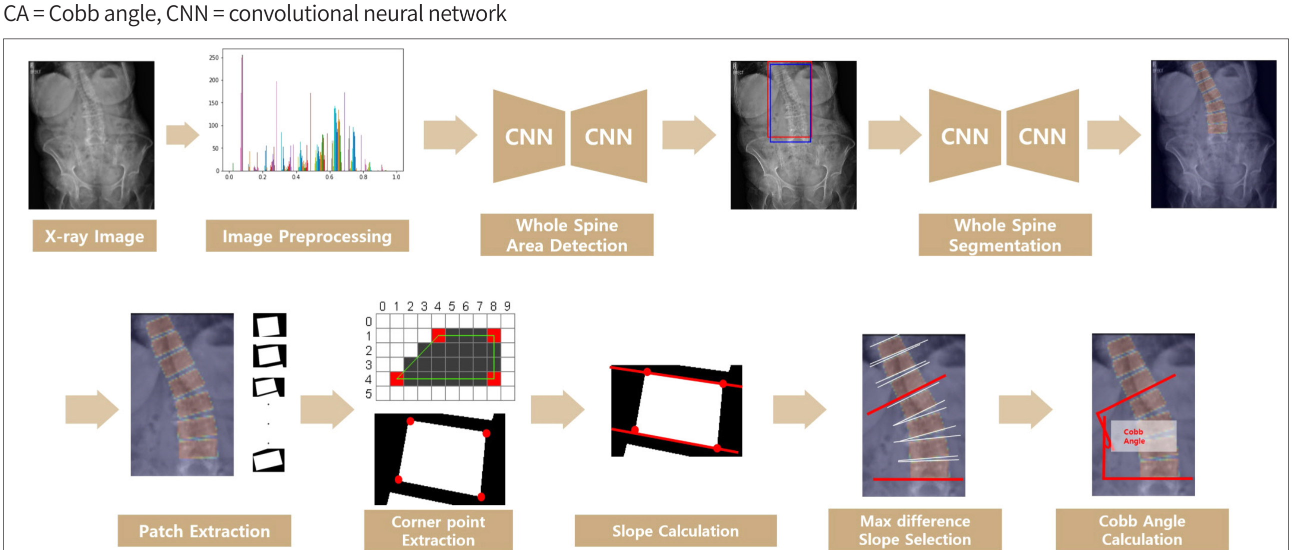
STATISTICAL ANALYSIS

SPSS (version 21.0; IBM Corp., Armonk, NY, USA) was used for the statistical analyses. The intraclass correlation coefficient (ICC) was used to evaluate the inter- and intra-observer reliabilities of CA measurements. ICC was classified into the following categories: poor (< 0.5), fair (0.5–0.75), good (> 0.75–0.9), and excellent (> 0.9). The Spearman's rank correlation coefficient was used to evaluate correlation and it was categorized as follows: little, if any relationship (± < 0.3), low (± 0.3–< 0.5), moderate (± 0.5–< 0.7), high (± 0.7–< 0.9), or very high (> ± 0.9). The Bland–Altman analysis was employed to illustrate the variability of the measurements and the difference in CA measurements between ACAM and observer 1, as well as between observer 2 and observer 1. Using the CA measured by observer 1 as the reference standard, the Wilcoxon signed-rank test was used to evaluate the differences between the ACAM and observer 1 measurements, and measurements of observers 2 and 1. Additionally, the median percentage accuracy (interquartile range [IQR]), median CA measurement differences (IQR), and proportion of CA measurements within ± 5° were calculated. The accuracy was calculated using the following formula:

$$\text{Accuracy (\%)} = 100 - 100\left(\left|\frac{(C - C')}{C}\right|\right)$$

where C was the CA measurement taken by observer 1 and C' was the CA measurement by

Fig. 2. Automated CA measurement using a CNN.
To measure CA, spine radiographs are sent to a software server. After processing the images, a CNN is used to detect the entire spine area and to segment each spine. The CNN extracts a corner point and calculates the angle of each spine. Finally, the CNN selects the slope showing the maximum difference, and the final CA is thus determined.
CA = Cobb angle, CNN = convolutional neural network



ACAM or observer 2. Statistical significance was set at $p < 0.05$. Furthermore, the Wilcoxon signed-rank test was employed to evaluate the differences in CA measurement time between ACAM, observer 1, and observer 2.

RESULTS

PATIENT CHARACTERISTICS AND CLASSIFICATION OF CA MEASUREMENTS

Out of a total of 411 patients, the most common group had spinal asymmetry (CA < 10°), while the second group had mild scoliosis (CA between 10° and 25°). Table 1 summarizes the demographic data of the patients and the classification of CA measurements.

INTER- AND INTRA-OBSERVER RELIABILITY OF THE CA MEASUREMENT

The inter-observer reliabilities of the ACAM, observer 1 (reference), and observer 2 are shown in Table 2. The inter-observer reliability was excellent between ACAM and observer 1 (ICC = 0.976; 95% confidence interval [CI], 0.971–0.980), and the ACAM and observer 2 (ICC = 0.976; 95% CI, 0.971–0.980). Spearman's rank correlation showed a very high correlation between ACAM and observer 1 (0.948, $p < 0.001$) and observer 2 (0.951, $p < 0.001$). The inter-observer reliability between the two observers was excellent (ICC = 0.999; 95% CI, 0.999–0.999), and Spearman's rank correlation showed a very high correlation (0.997, $p < 0.001$). The intra-observer reliability was excellent for observer 1 (ICC = 1.000; 95% CI = 1.000–1.000) and observer 2 (ICC = 0.999; 95% CI, 0.999–0.999).

Bland–Altman analysis revealed a mean difference in CA measurement between ACAM and observer 1 of 1.1, and the mean difference in CA measurement between observer 1 and observer 2 was 0.05 (Fig. 3).

Table 1. Demographic Data of the Patients

	Total ($n = 411$)	Male ($n = 150$)	Female ($n = 261$)
Age (yrs)	33.6 ± 21.9	35.5 ± 22.9	32.5 ± 21.1
Cobb angle (°)			
< 10	208 (50.6)	76 (50.7)	132 (50.6)
≥ 10 and ≤ 25	148 (36.0)	60 (40.0)	88 (33.7)
> 25 and ≤ 40	35 (8.5)	7 (4.6)	28 (10.7)
> 40	20 (4.9)	7 (4.6)	13 (5.0)

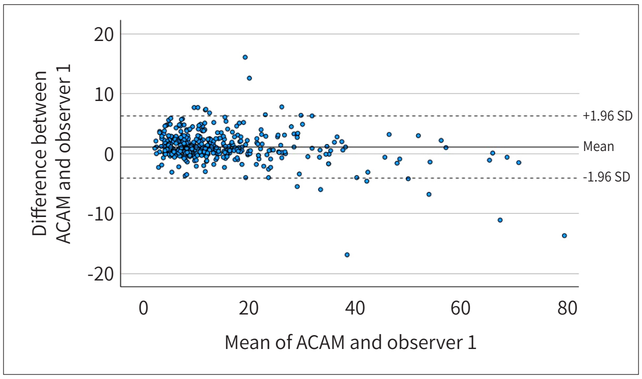
Data are reported as mean ± standard deviation or numbers (%). Cobb angle measurements were classified using observer 1 measurement as a reference standard.

Table 2. The Reliability of the Proposed ACAM

	ICC (95% CI)	Spearman's Correlation
Observer 1 vs. ACAM	0.976 (0.971, 0.980)	0.948 ($p < 0.001$)
Observer 2 vs. ACAM	0.976 (0.971, 0.980)	0.951 ($p < 0.001$)
Observer 1 vs. Observer 2	0.999 (0.999, 0.999)	0.997 ($p < 0.001$)

Observer 1 measurement was used as a reference standard.
ACAM = automated Cobb angle measurement, CI = confidence interval, ICC = intraclass correlation coefficient

Fig. 3. Bland–Altman plots of ACAM, observer 1, and observer 2.
The mean difference in CA measurement between ACAM and observer 1 is 1.1, while the mean difference in CA measurement between observer 1 and observer 2 is 0.05.
ACAM = automated Cobb angle measurement, CA = Cobb angle, SD = standard deviation



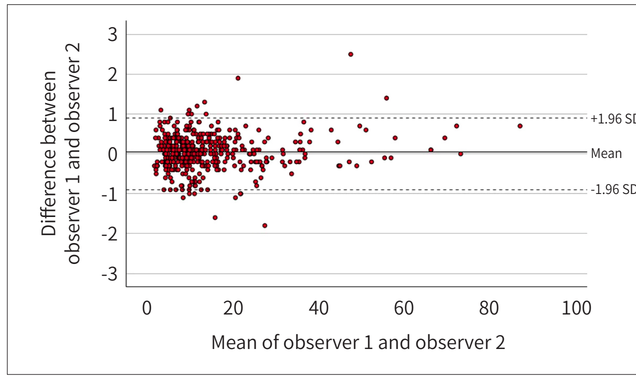


Table 3. The Accuracy of the ACAM and Observer 2 Measurement

	ACAM vs. Observer 1	Observer 2 vs. Observer 1	*p*-Value
Overall			
Median accuracy (%)	88.2 (69.4–96.3)	97.7 (94.1–99.0)	< 0.001
Median CA difference (°)	0.9 (0–2.4)	0.1 (-0.2–0.3)	< 0.001
CA difference within ± 5°	384/411 (93.4)	411/411 (100)	
CA < 10° (spinal asymmetry)			
Median accuracy (%)	77.1 (47.5–91.9)	95.6 (91.5–97.8)	< 0.001
Median CA difference (°)	1.3 (0.3–2.7)	0.1 (-0.2–0.3)	< 0.001
CA difference within ± 5°	198/208 (95.2)	208/208 (100)	
CA 10° to 25° (mild scoliosis)			
Median accuracy (%)	92.2 (84.1–97.1)	98.1 (96.6–99.2)	< 0.001
Median CA difference (°)	0.7 (0–2.1)	0.1 (-0.2–0.3)	< 0.001
CA difference within ± 5°	139/148 (93.9)	148/148 (100)	
CA > 25° to 40° (moderate scoliosis)			
Median accuracy (%)	96.0 (87.6–98.6)	99.2 (98.5–99.6)	< 0.001
Median CA difference (°)	0.5 (-0.5–2.7)	-0.1 (-0.3–0.3)	0.031
CA difference within ± 5°	31/35 (88.6)	35/35 (100)	
CA > 40° (severe scoliosis)			
Median accuracy (%)	95 (90.0–98.2)	99.3 (98.9–99.8)	< 0.001
Median CA difference (°)	-1.5 (-4.4–-0.1)	0.2 (-0.2–0.7)	0.008
CA difference within ± 5°	16/20 (80)	20/20 (100)	

Data are reported as number (%) or median (interquartile range). The *p*-value was obtained using the Wilcoxon signed-rank test.
ACAM = automated Cobb angle measurement, CA = Cobb angle

ACCURACY OF THE ACAM AND OBSERVER 2

Using the CA measurement of observer 1 as a reference standard, Table 3 shows a comparison of the differences between the ACAM and observer 1 measurements and between the measurements of observers 2 and 1. The overall accuracy of ACAM was high (88.2%), particularly in the mild and moderate scoliosis groups (92.2% and 96%, respectively) (Fig. 4). The

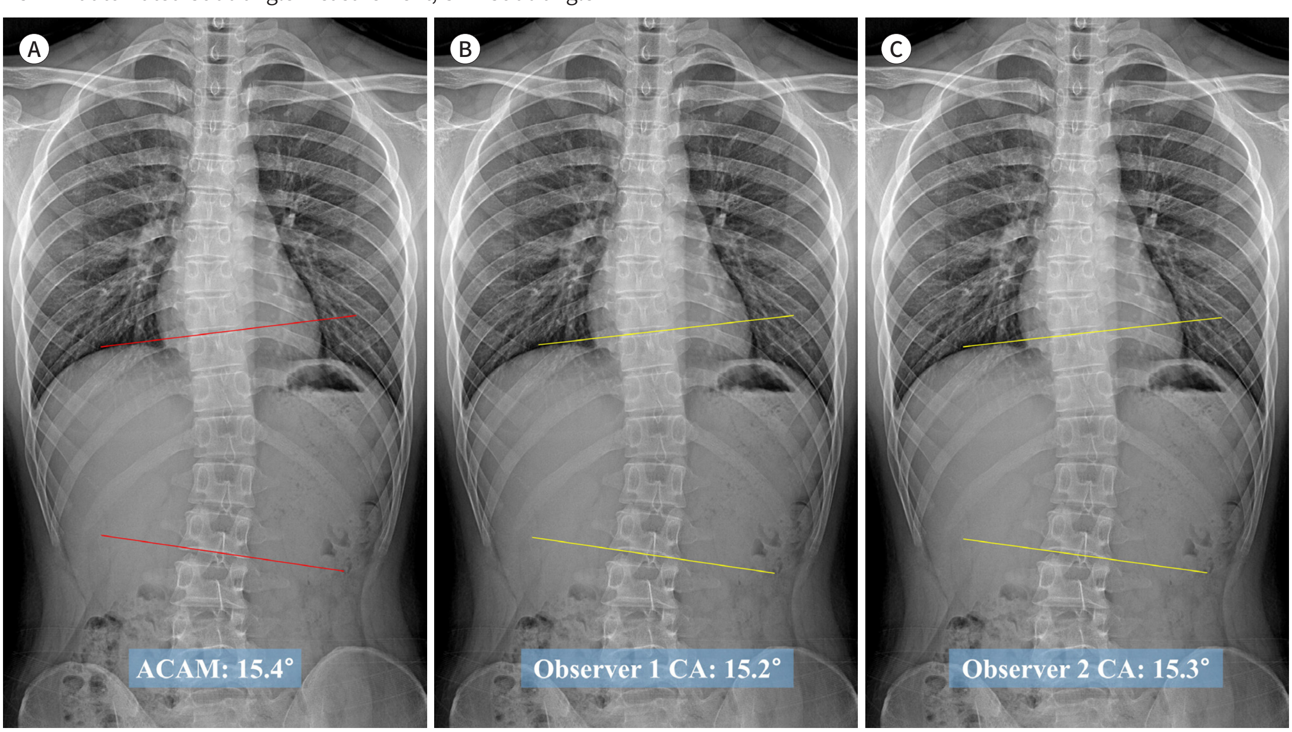
ACAM showed the lowest median accuracy (77.1%) in the spinal asymmetry group (CA < 10°). For the severe scoliosis group (CA > 40°), the ACAM showed higher accuracy (95%), but a lower CA than the observers and it showed the lowest percentage (80%) of CA differences within ± 5°. The Wilcoxon signed-rank test showed significant differences in median accuracy and CA between ACAM and observer 2 (all $p < 0.05$). The ACAM showed lower median accuracy (88.2% vs. 97.7%, $p < 0.001$), larger median CA differences (0.9° vs. 0.1°, $p < 0.001$), and a lower percentage of CA measurements within ± 5° (93.4% vs. 100%) than observer 2 (Fig. 5).

Regarding CA measurement time, it took approximately 18 seconds (standard deviation [SD] = 1.7) for the ACAM to measure the CA (excluding DICOM transfer), whereas it took an average of 30 seconds (SD = 7.8) for observer 1 and 35 seconds (SD = 7.6) for observer 2. The ACAM showed a significant difference in measurement time compared to observer 1 ($p < 0.001$) and observer 2 ($p < 0.001$), respectively.

DISCUSSION

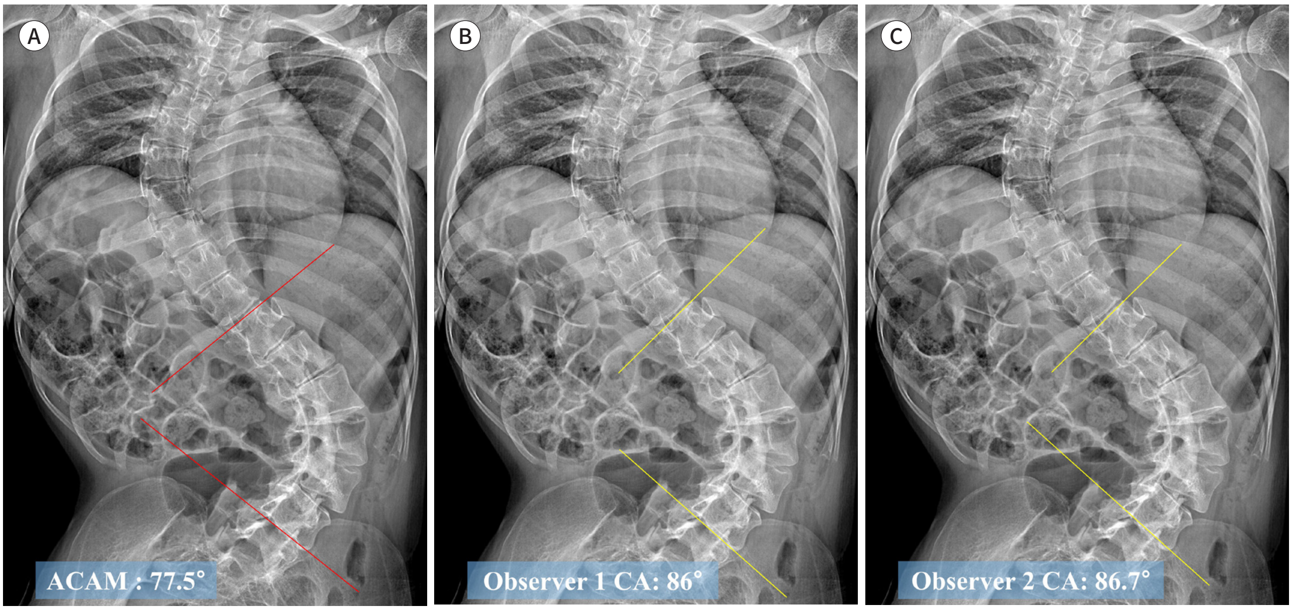
Many studies have developed deep or machine learning models for ACAM that are gaining increasing popularity in clinical practice and showing improved measurement precision and good correlation with manual measurements (3, 7, 9-12). The use of AI-based diagnostic supporting software is expected to improve measurement time and reproducibility (3, 6, 10, 11, 14, 15).

Fig. 4. Agreement between the ACAM and observers' CA measurements.
A. The ACAM using a convolutional neural network yields 15.4°. The most tilted superior end vertebra is T10, and the most tilted inferior end vertebra is L2.
B. The CA measured by observer 1 is 15.2°. The most tilted superior or inferior end vertebrae are the same as those for the ACAM.
C. The CA measured by observer 2 is 15.3°. Thus, the ACAM is almost identical to measurements by the observers.
ACAM = automated Cobb angle measurement, CA = Cobb angle



JOURNAL of THE KOREAN SOCIETY of RADIOLOGY

Fig. 5. Discrepancy between the ACAM and CA measurements by observers.
A. The ACAM using a convolutional neural network yields 77.5° for the major curve. The most tilted superior end vertebra is T11, and the most tilted inferior end vertebra is L4.
B. The major CA measured by observer 1 is 86°. The most tilted superior or inferior end vertebrae are the same as those for the ACAM.
C. The major CA measured by observer 2 is 86.7°. The ACAM is unable to draw a line parallel to the inferior endplate of the most tilted inferior vertebra.
ACAM = automated Cobb angle measurement, CA = Cobb angle



In our study, both observers demonstrated excellent intra-observer reliability in the measurement of CA. The consistency of observer 1's measurements served as a crucial standard for CA measurement. In terms of inter-observer reliability, the ACAM showed excellent reliability (ICC = 0.976) and a very high correlation (Spearman's correlation = 0.948) with the measurements taken by observer 1. When using observer 1 as the reference standard, ACAM also displayed excellent accuracy (88.2%). However, observer 2 (97.7%) outperformed ACAM in CA measurement. We hypothesized that observer 2, who had undergone 3 years of training in CA measurement, demonstrated higher performance compared to ACAM. In a previous study that used an expert neurosurgeon specializing in scoliosis as the reference, ACAM showed higher accuracy (93.6%) compared to a non-specialized neurosurgeon (85.9%) (7).

In our study, when it came to severe scoliosis cases (CA > 40°), the ACAM showed a negative median difference in CA measurement compared to observer 1. The current model tended to measure a lower CA than both observers. This discrepancy can be attributed to the severe tilting and rotation of the vertebrae in severe scoliosis, along with the overlapping of ribs with the vertebrae, which poses challenges in accurately measuring the CA. The ACAM was unable to measure CA parallel to the vertebral bodies in such cases, resulting in lower CA measurements than those of the observers. However, our study had the smallest sample size for severe scoliosis cases ($n = 20$).

In the spinal asymmetry group (CA < 10°), the ACAM exhibited the lowest median accuracy (77.1%) and the highest CA difference within 5° (95.2%). This suggests that subtle differences in labeling the end vertebrae may decrease the accuracy of the ACAM in CA measurements.

We speculate that severe scoliosis or spinal asymmetry may decrease the performance of the AI model in CA measurement. Therefore, we suggest that the application of the model may be more suitable for mild or moderate scoliosis groups.

In this study, patients with severe kyphosis or degenerative spondylosis were excluded. In such cases, the ACAM was unable to draw a line parallel to the most tilted endplate due to overlap with adjacent vertebrae in the anterior-posterior view or the presence of bony spurs (7, 10). Consequently, the ACAM used in this study might have limited utility for severe kyphosis or degenerative spondylosis, since it demonstrated variations in the selection of end vertebrae or the identification of the cervical spine as the end vertebra. Moreover, ACAM might produce inaccurate results when there are more than three spinal curves. When implementing ACAM in clinical practice, it is crucial to ensure the proper selection of end vertebrae. Therefore, the use of ACAM in actual clinical practice should be supervised by a radiologist. To the best of our knowledge, this is the first study to compare ACAM with CA measurement performed by a trained radiology resident.

The use of ACAM is expected to reduce measurement time. Previous studies reported that an automated method took an average 4.45 to 180 seconds to measure CA on each radiograph but it still necessitated manual selection and labeling of the end vertebrae (6, 9, 10, 13, 18). In our study, ACAM significantly reduced measurement time by nearly half. These results suggest that ACAM holds promise for practical use in clinical settings, effectively reducing the time required for angle measurements.

This study had several limitations. Firstly, it was performed retrospectively and had a relatively small sample size since it was a single-center study. Therefore, further prospective trials with larger populations should be conducted to validate the role and performance of ACAM. Secondly, a significant proportion of the patients (208 out of 411, 50.6%) exhibited spinal asymmetry (CA < 10°), which might have impacted the overall reliability and correlation of the CA measurements. To enhance the efficiency and accuracy of ACAM in diagnosing and treating scoliosis, an upgraded model utilizing CNN will be necessary in the future.

In conclusion, the use of ACAM with CNN can enhance CA measurement on spine radiographs, proving to be effective in assessing mild or moderate scoliosis, although it may have limitations in cases of spinal asymmetry or severe scoliosis. Nonetheless, its notable advantage lies in significantly reducing angle measurement time, making it a valuable tool for efficient use in clinical practice.

Availability of Data and Material

The datasets generated or analyzed during the study are available from the corresponding author on reasonable request.

Author Contributions

Conceptualization, O.H.Y., C.Y.S.; data curation, O.H.Y., C.Y.S.; formal analysis, O.H.Y., K.T.K., C.Y.S.; investigation, O.H.Y., C.Y.S.; methodology, O.H.Y., K.T.K., C.Y.S., P.M.; project administration, C.Y.S.; resources, O.H.Y., K.T.K., C.Y.S.; P.M.; software, C.Y.S.; supervision, C.Y.S.; validation, O.H.Y., C.Y.S.; visualization, O.H.Y., C.Y.S.; writing—original draft, O.H.Y., C.Y.S.; and writing—review & editing, all authors.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

ORCID iDs

Ha Yun Oh https://orcid.org/0000-0002-0874-7286
Tae Kun Kim https://orcid.org/0000-0002-5130-2194
Yun Sun Choi https://orcid.org/0000-0001-7082-3991
Mira Park https://orcid.org/0000-0003-3827-9089
Ra Gyoung Yoon https://orcid.org/0000-0003-3203-921X
Jin Kyung An https://orcid.org/0000-0003-1086-426X

Funding

None

REFERENCES

1. Kim H, Kim HS, Moon ES, Yoon CS, Chung TS, Song HT, et al. Scoliosis imaging: what radiologists should know. *Radiographics* 2010;30:1823-1842
2. Malfair D, Flemming AK, Dvorak MF, Munk PL, Vertinsky AT, Heran MK, et al. Radiographic evaluation of scoliosis: review. *AJR Am J Roentgenol* 2010;194(3 Suppl):S8-S22
3. Ha AY, Do BH, Bartret AL, Fang CX, Hsiao A, Lutz AM, et al. Automating scoliosis measurements in radiographic studies with machine learning: comparing artificial intelligence and clinical reports. *J Digit Imaging* 2022;35:524-533
4. Carman DL, Browne RH, Birch JG. Measurement of scoliosis and kyphosis radiographs. Intraobserver and interobserver variation. *J Bone Joint Surg Am* 1990;72:328-333
5. De Carvalho A, Vialle R, Thomsen L, Amzallag J, Cluzel G, le Pointe HD, et al. Reliability analysis for manual measurement of coronal plane deformity in adolescent scoliosis. Are 30 x 90 cm plain films better than digitized small films? *Eur Spine J* 2007;16:1615-1620
6. Zhang J, Lou E, Shi X, Wang Y, Hill DL, Raso JV, et al. A computer-aided Cobb angle measurement method and its reliability. *J Spinal Disord Tech* 2010;23:383-387
7. Caesarendra W, Rahmaniar W, Mathew J, Thien A. Automated Cobb angle measurement for adolescent idiopathic scoliosis using convolutional neural network. *Diagnostics (Basel)* 2022;12:396
8. Chen K, Zhai X, Sun K, Wang H, Yang C, Li M. A narrative review of machine learning as promising revolution in clinical practice of scoliosis. *Ann Transl Med* 2021;9:67
9. Sun Y, Xing Y, Zhao Z, Meng X, Xu G, Hai Y. Comparison of manual versus automated measurement of Cobb angle in idiopathic scoliosis based on a deep learning keypoint detection technology. *Eur Spine J* 2022;31:1969-1978
10. Pan Y, Chen Q, Chen T, Wang H, Zhu X, Fang Z, et al. Evaluation of a computer-aided method for measuring the Cobb angle on chest X-rays. *Eur Spine J* 2019;28:3035-3043
11. Liu J, Yuan C, Sun X, Sun L, Dong H, Peng Y. The measurement of Cobb angle based on spine X-ray images using multi-scale convolutional neural network. *Phys Eng Sci Med* 2021;44:809-821
12. Horng MH, Kuok CP, Fu MJ, Lin CJ, Sun YN. Cobb angle measurement of spine from X-ray images using convolutional neural network. *Comput Math Methods Med* 2019;2019:6357171
13. Qiao J, Liu Z, Xu L, Wu T, Zheng X, Zhu Z, et al. Reliability analysis of a smartphone-aided measurement method for the Cobb angle of scoliosis. *J Spinal Disord Tech* 2012;25:E88-E92
14. Wu C, Meng G, Lian J, Xu J, Gao M, Huang C, et al. A multi-stage ensemble network system to diagnose adolescent idiopathic scoliosis. *Eur Radiol* 2022;32:5880-5889
15. Suri A, Tang S, Kargilis D, Taratuta E, Kneeland BJ, Choi G, et al. Conquering the Cobb angle: a deep learning algorithm for automated, hardware-invariant measurement of Cobb angle on radiographs in patients with scoliosis. *Radiol Artif Intell* 2023;5:e220158
16. Wu C, Ou W, Gao M, Li J, Liu Q, Kang Z, et al. Digital measurement and correlation analysis of coronal and sagittal anatomic parameters in the radiographs of adolescent patients with idiopathic scoliosis. *Eur Spine J* 2023;32:1161-1172
17. Van Goethem J, Van Campenhout A, van den Hauwe L, Parizel PM. Scoliosis. *Neuroimaging Clin N Am* 2007;17:105-115

18. Sardjono TA, Wilkinson MH, Veldhuizen AG, van Ooijen PM, Purnama KE, Verkerke GJ. Automatic Cobb angle determination from radiographic images. *Spine (Phila Pa 1976)* 2013;38:E1256-E1262

컨볼루션 신경망을 이용한 척추측만증의 방사선학적 분석의 임상 적용

오하윤[1] · 김태건[1,2] · 최윤선[1*] · 박미라[3] · 윤라경[1] · 안진경[1]

목적 척추측만증 평가 시 컨볼루션 신경망(convolutional neural network; 이하 CNN)을 이용한 자동콥각도 측정(automated Cobb angle measurement; 이하 ACAM)의 신뢰성과 정확성을 평가하고 측정시간을 비교하였다.

대상과 방법 척추측만증이 의심되는 환자 411명을 대상으로 하였으며, 척추 방사선사진에 대해 ACAM을 수행하였다. 관찰자 1 (두 명의 근골격 영상의학과 의사의 합의)과 관찰자 2 (영상의학과 전공의)가 콥각도를 독립적으로 측정하였다. 콥각도 측정치는 관찰자 1의 측정치를 기준치로 사용하여 분류하였다. 관찰자 간 신뢰성과 상관관계를 급내상관계수(intraclass correlation coefficient; 이하 ICC)와 스피어만 순위 상관계수를 이용하여 평가하였고 ACAM과 관찰자의 정확도와 측정시간을 평가하였다.

결과 ACAM은 관찰자 1과 높은 신뢰성을 보이며 매우 높은 상관관계를 나타냈고(ICC = 0.976, 스피어만의 순위 상관계수 = 0.948) 평균 콥각도 차이는 1.1이었다. 전체적인 정확도는 높았으며(88.2%), 특히 경도(92.2%) 및 중등도(96%) 척추측만증 그룹에서 높았다. 반면 척추 비대칭 그룹에서는 정확도가 낮았고(77.1%) 심한 척추측만증 그룹에서는 높았으며(95%) 관찰자보다 콥각도를 작게 측정하는 경향을 보였다. ACAM은 관찰자에 비해 측정시간이 절반 정도 짧았다($p < 0.001$).

결론 CNN을 이용한 ACAM은 경도나 중등도 척추측만증의 콥각도 측정을 향상시키나, 척추 비대칭이나 심한 척추측만증에 사용하는 데는 제한 점이 있었다. 하지만 측정시간을 현저히 감소시켰다.

[1]을지대학교 노원을지대학교병원 영상의학과,
[2]서울지구병원 영상의학과,
[3]을지대학교 의과대학 예방의학교실