












Diagnosis of Scoliosis Using Chest Radiographs with a Semi-Supervised Generative Adversarial Network

준지도학습 방법을 이용한 흉부 X선 사진에서 척추측만증의 진단

Woojin Lee, MD^{1*} , Keewon Shin, PhD^{2*} , Junsoo Lee, BSc^{3,4} ,
Seung-Jin Yoo, MD¹ , Min A Yoon, MD⁵ , Yo Won Choi, MD¹ ,
Gil-Sun Hong, MD⁵ , Namkug Kim, PhD^{2,4*} , Sanghyun Paik, MD^{1**} 


¹Department of Radiology, Hanyang University Hospital, Seoul, Korea

²Department of Bioengineering, Asan Medical Institute of Convergence Science and Technology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea


³Department of Industrial Engineering, Seoul National University, Seoul, Korea


Departments of ⁴Convergence Medicine and ⁵Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea


ORCID iDs


Woojin Lee  <https://orcid.org/0000-0001-5185-8531>

Keewon Shin  <https://orcid.org/0000-0002-5028-5716>

Junsoo Lee  <https://orcid.org/0000-0003-0886-6119>


Seung-Jin Yoo  <https://orcid.org/0000-0002-0779-3889>

Min A Yoon  <https://orcid.org/0000-0003-4033-9060>

Yo Won Choi  <https://orcid.org/0000-0003-0508-9430>

Gil-Sun Hong  <https://orcid.org/0000-0002-0068-9413>

Namkug Kim  <https://orcid.org/0000-0002-3438-2217>

Sanghyun Paik  <https://orcid.org/0000-0002-8276-333X>

Received August 27, 2021
Revised October 3, 2021
Accepted November 8, 2021

*Corresponding author

Sanghyun Paik, MD
Department of Radiology,
Hanyang University Hospital,
222-1 Wangsimni-ro,
Seongdong-gu, Seoul 04763,
Korea.

Tel 82-2-2290-9164

Fax 82-2-2293-2111

E-mail radpsh@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

* † These authors contributed equally to this work.

Purpose To develop and validate a deep learning-based screening tool for the early diagnosis of scoliosis using chest radiographs with a semi-supervised generative adversarial network (GAN).

Materials and Methods Using a semi-supervised learning framework with a GAN, a screening tool for diagnosing scoliosis was developed and validated through the chest PA radiographs of patients at two different tertiary hospitals. Our proposed method used training GAN with mild to severe scoliosis only in a semi-supervised manner, as an upstream task to learn scoliosis representations and a downstream task to perform simple classification for differentiating between normal and scoliosis states sensitively.

Results The area under the receiver operating characteristic curve, negative predictive value (NPV), positive predictive value, sensitivity, and specificity were 0.856, 0.950, 0.579, 0.985, and 0.285, respectively.

Conclusion Our deep learning-based artificial intelligence software in a semi-supervised manner

achieved excellent performance in diagnosing scoliosis using the chest PA radiographs of young individuals; thus, it could be used as a screening tool with high NPV and sensitivity and reduce the burden on radiologists for diagnosing scoliosis through health screening chest radiographs.

Index terms Scoliosis; Mass Screening; Thoracic Radiography; Deep Learning; Artificial Intelligence

INTRODUCTION

Scoliosis is defined as a lateral spinal curvature with a Cobb angle $\geq 10^\circ$ (1). Although this abnormal curvature may be because of an underlying congenital or developmental osseous or neurologic abnormality, approximately 80% of all scolioses are idiopathic with unknown etiology (1). The prevalence of adolescent idiopathic scoliosis is reportedly 2%–4% (2), but we assume that the prevalence rate is increasing as adolescents sit for longer periods.

In adolescence, the progression of idiopathic scoliosis is associated with the rate of spinal growth and the initial scoliosis curvature (3). The interval between follow-up observations can be shortened by up to 4 months in adolescence, that is the period of the rapid skeletal growth (3, 4). Adults whose skeletal growth has been completed should only be monitored for scoliosis if they have a Cobb angle $\geq 30^\circ$ (3, 4). Scoliosis with Cobb angle $> 50^\circ$ is known to increase the rate of back pain and the mortality associated with cardiopulmonary complications (5). Therefore, the most important factor for the timely treatment of idiopathic scoliosis is diagnosing the condition early and setting the treatment policy before bone growth completion.

The Cobb angle of a scoliotic curve is the angle formed by the intersection of two lines, one parallel to the endplate of the superior end vertebra and the other parallel to the endplate of the inferior end vertebra (1). The total error of the Cobb angle measurement is known to be approximately 2° – 7° (6), and the intra-observer variation is somewhat significant (approximately 5° – 10°) (7, 8). However, the Cobb angle remains the most widely used tool to diagnose scoliosis to date.

Scoliosis can be diagnosed with coronal plain radiographs, including chest posteroanterior (PA) radiographs; therefore, additional imaging is not required. Measuring the Cobb angle to determine whether scoliosis is present is an effortless but time-consuming procedure for radiologists. Moreover, it is crucial to observe the changes of the Cobb angle in the follow-up observations, but the inter-observer and intra-observer variations are relatively substantial (6-8). Additionally, radiologists focus on the lung, not the spine, when reading chest radiographs.

Several studies have assessed the use of deep learning-based artificial intelligence (AI) software (9-11) in scoliosis detection. Tan et al. (11) showed good consistency between the automatic and reference measurements of the Cobb angle on spine radiograph images. No statistically significant difference was found between the Cobb angle measurements obtained by the doctor and the system. The mean deviation between the doctor and the system was $1.7^\circ \pm 1.2^\circ$, which was extremely minor in significance.

Diagnosing scoliosis on chest radiographs taken for health examination purposes using a deep learning-based AI software may help in the early diagnosis of scoliosis. It may increase the diagnostic efficiency and the consistency of examination and automatic reading immediately after the examination.

To our knowledge, there have been no studies of deep learning networks developed for scoliosis screening. This study aimed to develop an artificial intelligence network that can detect scoliosis sensitively on screening chest radiographs, effectively diagnoses it in advance, and reduces the screening overload by adequately filtering true negatives (TNs).

MATERIALS AND METHODS

This retrospective study was approved by the appropriate Institutional Review Board of two hospitals and the requirement for patient consent was waived (Asan Medical Center [AMC], IRB No. 2019-0115; Hanyang University Seoul Hospital [HUSH], IRB No. 2020-04-061).

DATASET

Chest PA radiographs of patients with and without scoliosis were used to train the deep-learning-based AI software. To build the center-agnostic model, we collected chest PA radiographs taken from a separate center. We used Picture Archiving and Communication System to search for the chest radiographs. Chest PA radiographs of patients (356 normal and 1209 scoliosis radiographs; age range, 15–35 years), taken for health examinations between January 2004 and December 2019, were randomly collected from HUSH. Additionally, chest PA radiographs of the patients (356 normal and 1666 scoliosis radiographs; age range, 10–18 years), taken for health examinations between January 1997 and November 2018, were randomly collected from AMC. Radiographs with scoliosis were diagnosed based on medical readouts. Initially, cases with poor image quality, artifacts, or surgery were excluded. Then one radiologist (L.W., experience: 3 years) manually drew the Cobb angle on the chest PA radiographs to determine whether the patient had scoliosis.

Table 1. Description of the Dataset

	Hanyang University Seoul Hospital			Asan Medical Center		
	1) Upstream Task Training Set	2) Downstream Task Training Set	3) Downstream Task Test Set	1) Upstream Task Training Set	2) Downstream Task Training Set	3) Downstream Task Test Set
Number of radiographs	853	512	200	1310	512	200
Positive	853	256	100	1310	256	100
Negative	0	256	100	0	256	100
Age, years						
Positive	27.78 ± 5.97	27.11 ± 4.34	21.92 ± 3.59	14.58 ± 2.08	14.87 ± 2.09	16.42 ± 1.17
Negative	-	32.33 ± 1.73	15.35 ± 0.59	-	14.62 ± 2.76	16.50 ± 1.12
Sex						
Positive	F/330 M/268 O/255	F/114 M/30 O/112	F/82 M/18	F/691 M/198 O/421	F/199 M/57	F/74 M/26
Negative	-	F/180 M/76	F/49 M/51	-	F/113 M/143	F/46 M/54

Age = years old ± standard deviation, F = female, M = male, Negative = normal, O = unknown, positive = scoliosis, Sex = gender/number

DATASET TYPE

After merging all cross-hospital data, the dataset was split into three: 1) upstream task training set: learns the representation of a dataset, 2) downstream task training and validation set: trains a vector classifier using GAN, and 3) downstream task test set: evaluation of the downstream task classifier. These datasets were separated independently without overlapping, as the duplicated data from the upstream task and downstream task operations can interfere with model validation. Table 1 shows the demographic description of the dataset by the center and training type.

As shown in the class ratio statistics of both centers, only radiographs with scoliosis were included in the upstream task training to achieve salient feature extraction. A downstream work designed a model for diagnosing the presence or absence of scoliosis. Therefore, both training and test datasets were required to be binary classified as normal and abnormal based on the Cobb angle.

PREPROCESSING OF IMAGES

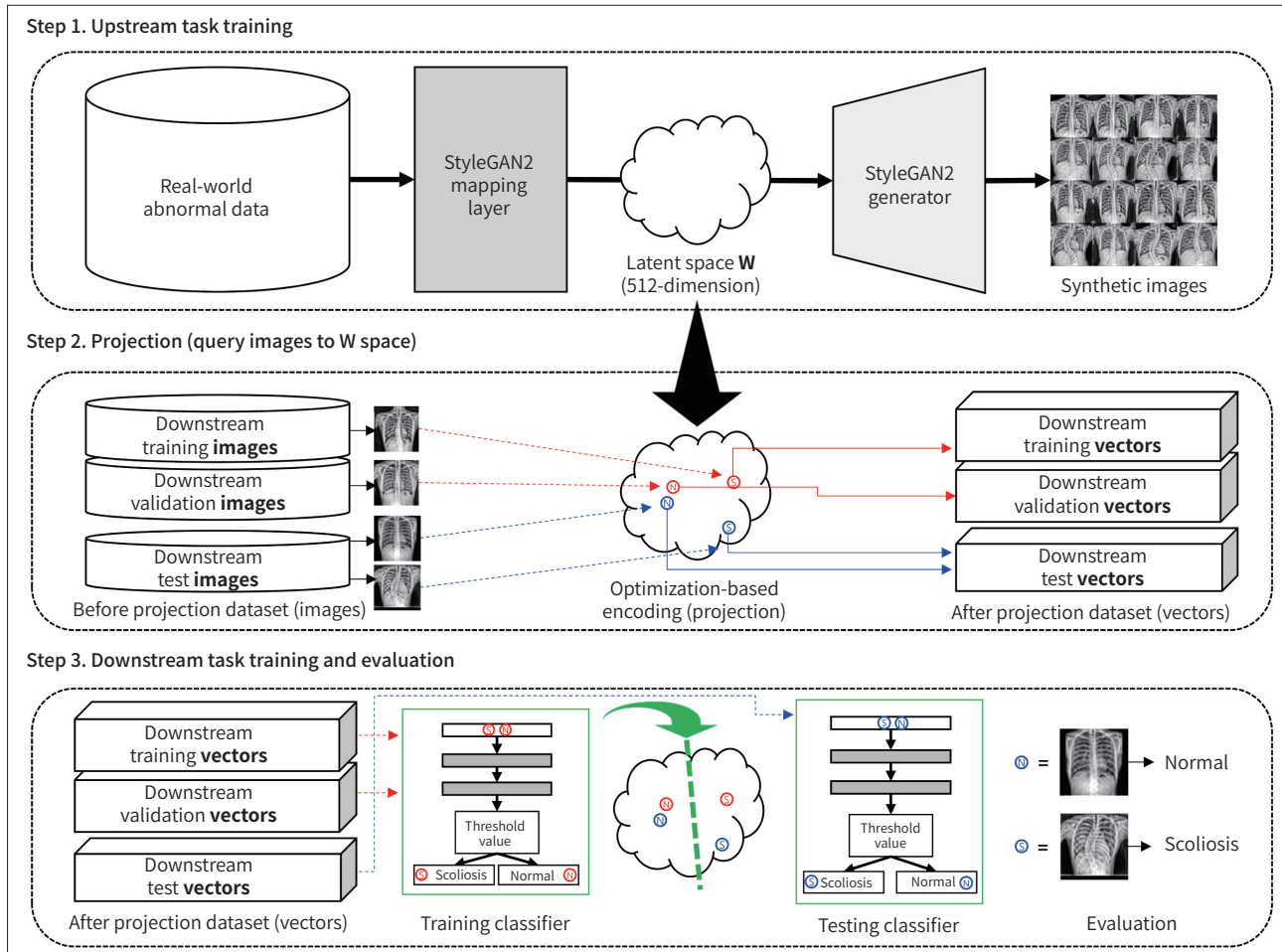
Histogram matching was performed for image standardization, and the top and bottom 2% values were excluded. Contrast limited adaptive histogram equalization (12) was taken to increase the bone contrast compared to the soft tissue. Then, padding was added to the image to preserve the aspect ratio during image resizing. After the images were resized to a 512×512 resolution, the chest PA images were finally converted to 8-bit, 3-channel images. Since the data set was acquired over a long period, it was necessary to verify the chest PA image quality according to the acquisition period. When visual inspection and image histogram analysis were compared by acquisition period, there was no significant difference between our pre-processed chest PA images.

DEVELOPMENT OF A DEEP LEARNING ALGORITHM

Since only 16% of the dataset contained exact Cobb angle (572/3587), it was challenging to derive regression-based prediction using supervised learning models. To overcome this, we proposed a semi-supervised classifier using StyleGAN2-ADA (13). StyleGAN2-ADA is a generative model that can create high-quality images with only about thousands of images by adding augmentation to the discriminator. This generative model forms a multivariate distribution by learning the distribution of images, which is a latent space. Fig. 1 shows a diagram of our proposed method. Our algorithm performs the Scoliosis classification task in the third step through two preceding steps, upstream task and projection: 1) upstream task training (a preceding training step for the downstream task): training StyleGAN2-ADA to extract rich semantics from data distributions, 2) projection: project downstream task training and validation images onto latent space to extract vectors where semantic representations were embedded, and 3) downstream task training and testing (a final step for scoliosis classification of our proposed model): train and evaluate a simple classifier comprising a multi-layer perceptron (MLP) (14) using projected vectors. Pytorch 1.8 was used for the deep learning framework, and single NVIDIA Titan RTX 24G was used for GPU computing.

Fig. 1. Schematics of our proposed method for classifying scoliosis using GAN.

The proposed algorithm consists of three parts, including the 1) upstream task training: representation learning of scoliosis, 2) project query image onto the latent space W , and 3) downstream task training and testing: a final classifier using projected vector.



GAN = generative adversarial network

REPRESENTATION LEARNING OF SCOLIOSIS

To understand the maximum diversity of the distribution of the targeting abnormality, only abnormal samples with various severities of diseases were used for training the StyleGAN2-ADA (13) model as a feature extractor. In the case of scoliosis, 2163 chest PA radiographs with various scoliosis severities from two centers were used to train the generator of the StyleGAN2-ADA. We trained a model that well represented scoliosis using only positive scoliosis samples. The learning rate of GAN was set to $1e^{-4}$ while training. The training was performed without using any pre-trained weights, as popularly used pre-trained weights are trained on human face datasets such as CelebA (15) or FFHQ (16), which has quite distant features from chest PA radiographs. Training quality assessment was conducted both in a qualitative and quantitative manner, with regards to Fréchet inception distance (FID) scores (17) and an expert's visual inspection. The assessment was performed after every 10000 iterations, i.e. when 10000 images were shown to the discriminator of our architecture. As the FID score was a metric devised to calculate the distribution of 3-channel images, we concatenated gray-

scale image channel-wise to calculate relevant metrics. While training GAN, the FID score was converged at an average of 15.0 after 6 million images were shown. The training lasted approximately 4 days with single NVIDIA Titan RTX 24GB GPUs.

PROJECTION (QUERY IMAGE ONTO LATENT SPACE W)

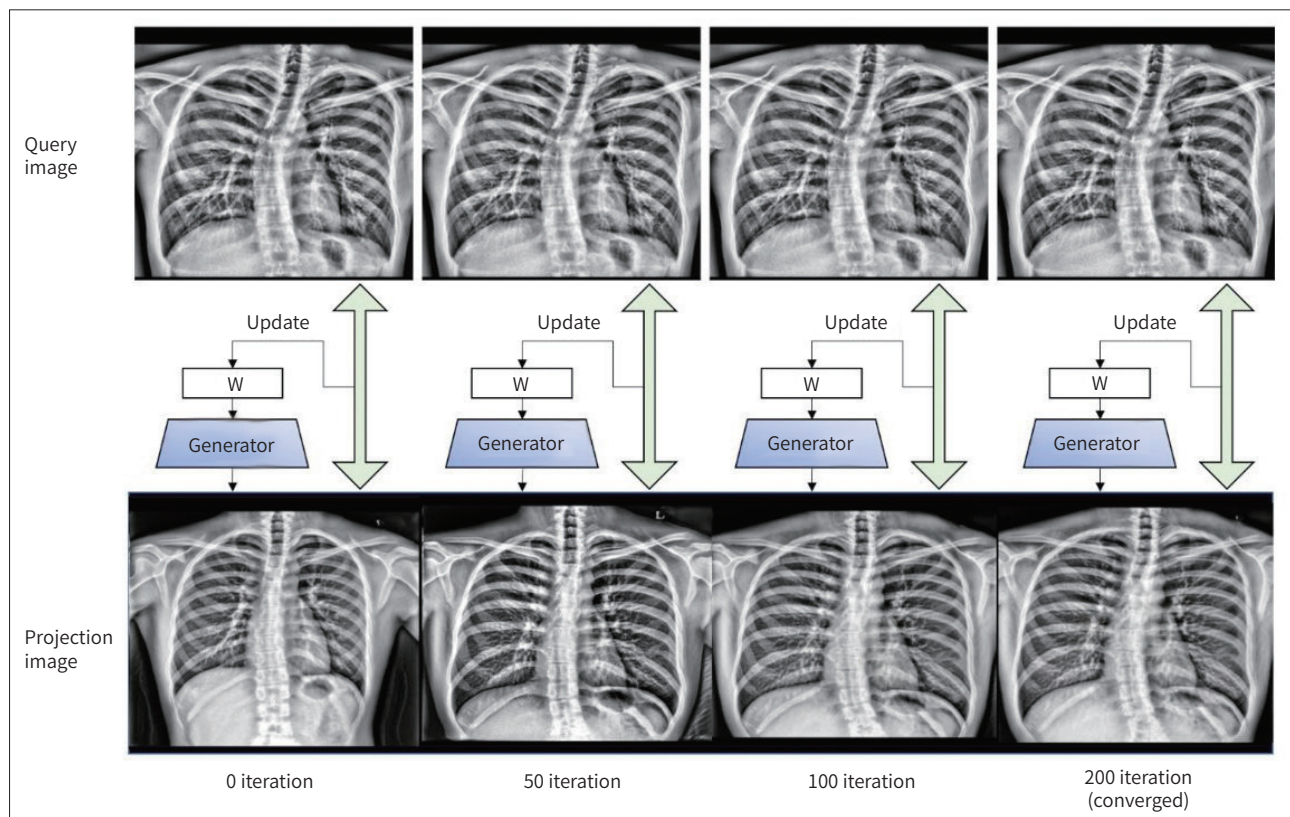
Query images were projected onto intermediate latent space W in an optimization-based method using the trained weights. The mapping layer of StyleGAN2-ADA was used to map gaussian noise Z to relevant latent W . Through the iterative optimization process, input W vector could generate synthetic images that were most similar to the given query image. After 200 iterations of the projection were completed, 512×512 size images were reduced into 512-dimension vectors in which semantics visible in the images were embedded. Fig. 2 shows the optimization-based projection process. This process was a prerequisite for training and evaluating the following downstream tasks. Thus, all data for downstream task training, validation, and test images were projected onto an intermediate latent space W with an optimization-based method using trained weights.

CLASSIFIER USING PROJECTED VECTOR

We constructed another downstream task with MLP trained with latent vectors from weakly

Fig. 2. The optimization-based projection process.

At iteration 0, the projection image is different from the query image because the projection image was initially generated from random W . With increasing iterations, the W values converge, producing a similar projection image as the query image. After 200 iterations, the query image and the projected image presents similar scoliosis characteristics.



labeled chest PA radiographs to validate its distinguishing ability. The MLP model used for the downstream task was a simple two-block architecture consisting of Linear Group Normalization and Gaussian Error Linear Unit, with a sigmoid attached to the end.

Then, we studied how many chest PA radiographs used in training were needed to achieve 95% sensitivity and 95% negative predictive value (NPV) classification capabilities. The ablation studies were compared on a fixed evaluation set comprising latent vectors projected from 100 abnormal and 100 normal chest PA radiographs. Validation data were randomly selected from the training data set at 20%. Binary cross-entropy loss was adapted for training the MLP model. Kingma and Ba (18) was used as an optimization function, and the learning rate was $1e^{-4}$. For an ablation study to evaluate classification performance according to the number of datasets, we conducted the test while increasing the number of downstream training data. We evaluated the model with indicators such as area under the receiver operating characteristic (ROC) curve (AUROC), accuracy (ACC), sensitivity, specificity, NPV, and positive predictive value (PPV).

As a control group, we also trained another convolutional neural network (CNN) using the exact data for training and validation. We did not utilize widely used pre-trained weights, as those were trained on data with notable differences from chest PA radiographs. We used ResNet34 (19) as backbone architecture, binary cross-entropy as the loss function, Adam optimizer, and learning rate of $1e^{-4}$. The training took only a few minutes with a single NVIDIA Titan RTX 24GB GPU.

INTERNAL VALIDATION OF THE SOFTWARE

We constructed a dataset of 100 normal and 100 scoliosis samples to validate the model performance. Since the data was collected from both HUSH and AMC, two test sets were established for each center. For each center, data were first purified by age group of 15–25 years, and 100 normal and 100 scoliosis samples were randomly selected. The selected images were used for the internal validation of the developed diagnosing software. The images used for the internal validation were not included in both the upstream and downstream training datasets.

We analyzed each performance according to the threshold of the likelihood value, distributed between 0 and 1, extracted by the AI model.

The AUROC and NPV of the classifier were mainly evaluated to determine whether the model performances were significantly better. We used the AUROC, ACC, sensitivity, specificity, NPV, and PPV values for quantitative evaluation of the classification into true positive (TP), false positive (FP), TN, and false-negative (FN).

STATISTICAL ANALYSIS

We used ROC comparison to check for statistical significance to compare our method with conventional image classification methods. The significant alpha was considered as 0.05 ($p < 0.05$). Python 3.7, Sklearn 0.23.2, and Numpy 1.16.4. were used for the evaluation of AUROC, ACC, NPV, sensitivity, specificity, and PPV. All statistical evaluations were performed by MED-CALC (MedCalc software, Ostend, Belgium) version 19.1.3.

Table 2. Performance of the Internal Validation Set Based on the Threshold of the Model Output

	AUROC (95% CI)	ACC	SEN	SPE	PPV	NPV
Threshold of model output	0.856 (0.818–0.889)					
0.01		0.527	1.000	0.055	0.514	1.000
0.05		0.635	0.985	0.285	0.579	0.950
0.1		0.663	0.965	0.365	0.601	0.911
0.2		0.705	0.875	0.535	0.653	0.811
0.3		0.748	0.810	0.685	0.720	0.783
0.4		0.768	0.745	0.790	0.780	0.756
0.5		0.783	0.695	0.870	0.842	0.740

ACC = accuracy, AUROC = area under the receiver operating characteristic curve, CI = confidence interval, NPV = negative predictive value, PPV = positive predictive value, SEN = sensitivity, SPE = specificity

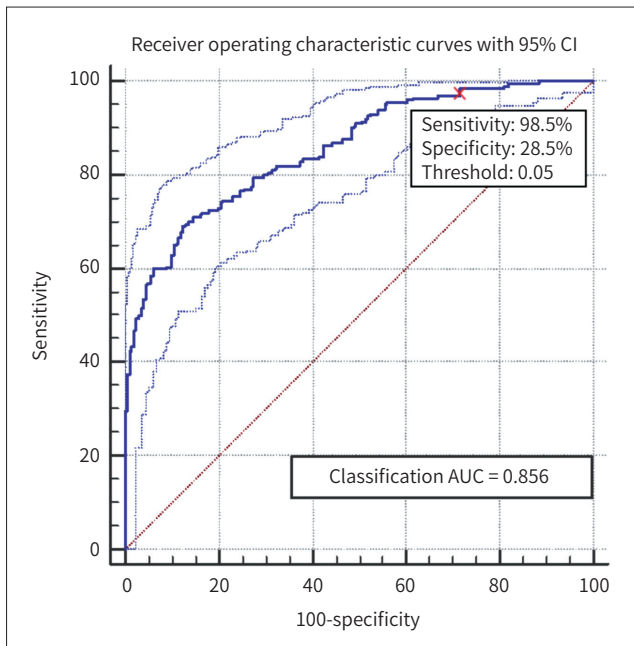


Fig. 3. The receiver operating characteristic curves for downstream tasks classifier based on the number of training samples. The red × sign indicates the selected operating point. AUC = area under the curve, CI = confidence interval

RESULTS

To evaluate the scoliosis classification performance of our proposed model, we calculated with ACC, sensitivity, specificity, PPV, and NPV. We compared the performance by adjusting the threshold of the AI model according to the screening purpose. Threshold was analyzed in the range from 0.01 to 0.5 to set sensitively, and the results are shown in Table 2. Fig. 3 shows the ROC curves with 95% confidence interval (CI), and we selected the threshold as 0.05 with reference to this result. As the threshold increased, ACC, specificity, and PPV tended to increase, while sensitivity and NPV decreased. Our final model screened the suspected scoliosis group with a performance of 0.985 for sensitivity and 0.950 for NPV. We have achieved our target performance through fine-tuning, and this shows that it is possible to classify the use of the GAN generator as a feature encoder. In our test results, 57 TNs total of 200 (28.5% of negative

Fig. 4. Three false-negative cases, including (A) dextroscoliosis with relatively short segmental involvement at the upper thoracic level, (B) dextroscoliosis with upper lumbar level involvement, and (C) thoracic levoscoliosis, the commonly encountered type of idiopathic scoliosis. The measured Cobb angles in the three cases are approximately 35° (somewhat significant), 15°, and 15°, respectively.

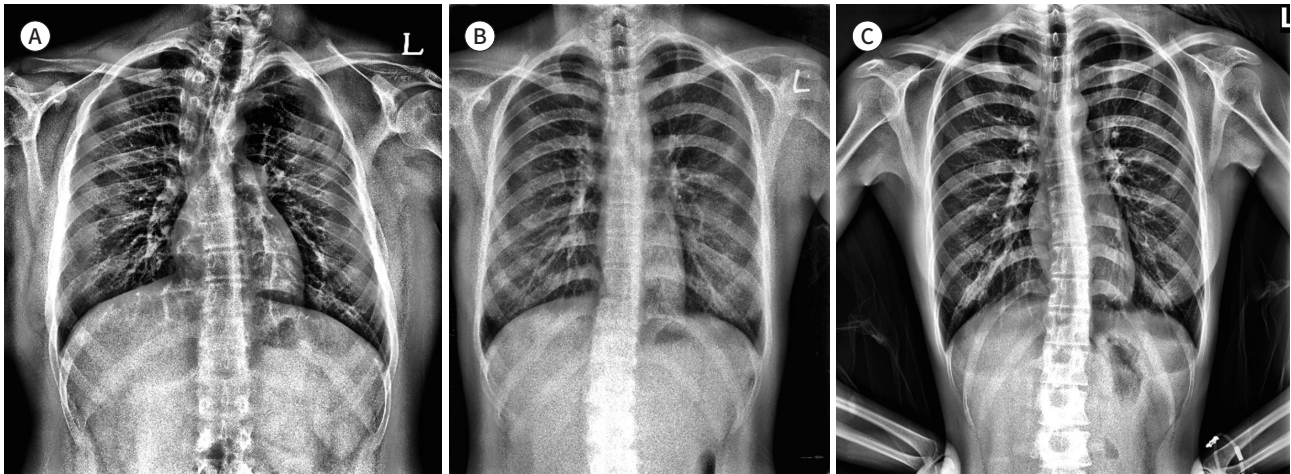
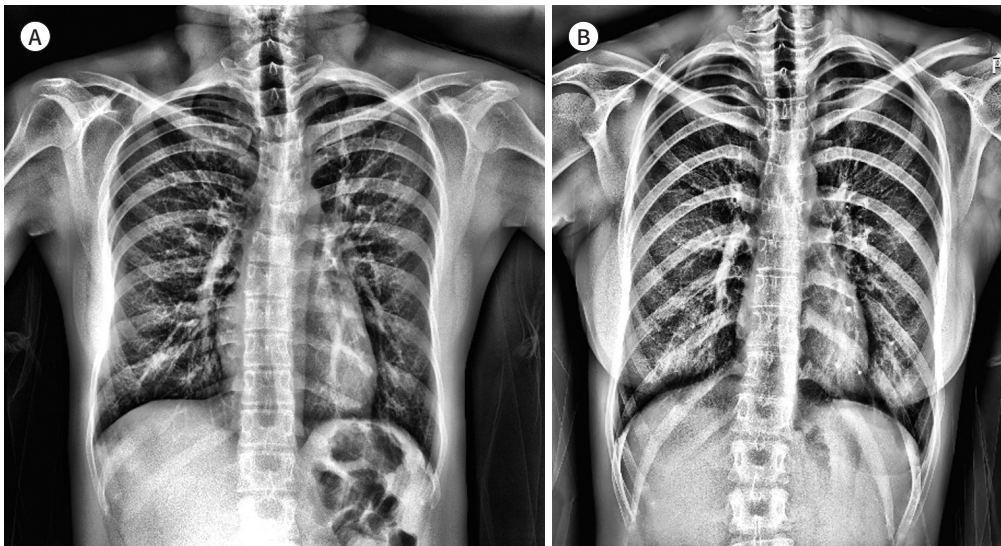


Fig. 5. Two representative true-positive cases of (A) levoscoliosis with relatively short segmental involvement at the upper thoracic level and (B) relatively mild dextroscoliosis with lower thoracic and upper lumbar level involvement. The measured Cobb angles are approximately 20° and 13°, respectively. Both cases present mild but relatively long segmental involvement of scoliosis.



samples) were filtered out as true normal, and 3 FNs (1.5% of positive samples) co-occurred. Although the natural occurrence is not reflected in the test set, we filtered out a relatively large number of true normals. Figs. 4, 5 show these three FN cases and examples of two TP cases.

Lastly, for control group, classification models using traditional CNNs with the same downstream data set failed to train. 512 labeled image samples are considered insufficient for classification training. However, it was possible to learn using pretrained weights from IMAGENET.

DISCUSSION

We developed a scoliosis diagnosing algorithm with high classification performance. At downstream test, the AUROC, NPV, PPV, sensitivity, and specificity were 0.856, 0.950, 0.579, 0.985, and 0.285, respectively, when the downstream task training sample was 512 patients with an even number of normal and scoliosis patients, and 0.05 thresholds of the sigmoid function. And there were three FNs (1.5% of positive samples) at this test. Figs. 4, 5 show these three FN cases and examples of two TP cases. It was challenging to analyze the exact cause of FN sortation with only three cases. Nevertheless, there might be detection errors in short segmental upper thoracic scoliosis cases and mild upper lumbar scoliosis cases. One case of scoliosis with relatively short segmental involvement at the upper thoracic level and one case of scoliosis with upper lumbar level involvement were sorted as FN cases (Fig. 4A, B). However, other similar cases (Fig. 5) were sorted correctly as TP cases. Another FN case was a commonly encountered type of idiopathic scoliosis (Fig. 4C). Further analysis is needed to determine the exact reason for this FN sortation.

Studies have assessed the use of deep learning algorithms (DLAs) in the detection of scoliosis. Tan et al. (11) and Tu et al. (20) applied a deep learning-based system to diagnose scoliosis on spine X-ray images. They found that DLAs can achieve segmentation of lumbar spine radiographs and automatic measurement of the Cobb angle. Liu et al. (21) proposed a keypoint estimation approach to assess the spine curve for idiopathic scoliosis-diagnosis from the small dataset. It can alleviate the overfitting problem of the training network with limited data. Their approach has achieved better performance compared with mainstream pose estimation methods. However, these methods require additional spine X-ray images and cannot be used as a screening tool. Kokabu et al. (22) found that the 3-dimensional depth sensor imaging system with its newly innovated CNNs for regression is objective and has a significant ability to predict the Cobb angle in children and adolescents. This can be used without additional radiation exposure, but requires special equipment, and cannot be used for scoliosis screening. Moreover, Yang et al. (23) showed that DLAs can be trained to detect scoliosis, identify cases with a curve $\geq 20^\circ$, and perform severity grading using unclothed back images with an ACC, sensitivity, specificity, and PPV higher or comparable to those of human experts. However, it requires additional photography and can only detect scoliosis with a curve $\geq 20^\circ$. In contrast, our method used GAN to construct a network for classifying Cobb angles in two steps: the upstream and downstream tasks. Compared to supervised learning, our classifier showed sufficiently sensitive classification performance for the few weak labeled dataset. Of course, there was a limit to training GAN in the upstream task that only scoliosis images were used to increase sensitivity. Unlike the ANOGAN (24, 25) algorithm, which detects abnormalities corresponding to outliers by learning only the normal distribution, our algorithm studied the techniques for better learning disease characteristics by training spectra corresponding to various abnormalities. We planned to conduct research to improve downstream performance based on the training results without selecting upstream tasks in future research.

Our study showed that the diagnostic software could be applied to scoliosis screening and only requires chest PA radiographs that were already taken. Oh et al. (26) reported that chest radiographs for scoliosis screening were observed with 93.94% of sensitivity and 61.67% of

specificity in thoracic curves; however, less than 40% of sensitivity (38.27%, 20.00%, and 25.80%) and more than 95% of specificity (97.34%, 99.69%, and 98.45%) were observed in thoraco-lumbar, lumbar, and double major curves, respectively. The incidence of thoracic curve scoliosis was overestimated and lumbar curve scoliosis was easily missed by chest radiography. Scoliosis screening using chest radiography has limited values, nevertheless, it is useful method for detecting thoracic curve scoliosis. Therefore, it reduces unnecessary radiation exposure and increases the diagnostic efficiency and consistency of examination and automatic reading immediately after the examination. Using our diagnostic software as a scoliosis screening tool for young individuals' chest radiographs, diagnosing scoliosis is possible without placing a burden on radiologists. We expect that the introduction of scoliosis screening on chest radiographs using an AI software will significantly help with the early diagnosis and treatment of scoliosis.

Our study has several limitations. First, the training datasets included only scoliosis cases and normal cases of young individuals without underlying diseases. Therefore, it may be challenging to apply the automatic detection to children or the middle-aged population and in cases with underlying diseases, including lung lesions. However, since this study aimed to screen scoliosis in the young population, we believe that there will be no problem with its practical application. Second, chest radiograph cases with underlying diseases other than scoliosis were excluded from the training and validation datasets. Nevertheless, the cases screened out by the DLAs may have mixed etiologies, such as congenital scoliosis, Marfan syndrome, or neuromuscular scoliosis. Third, there were three FN cases in the downstream task test. We presume that there might be detection errors with relatively less encountered cases. Furthermore, in general, the upper lumbar level is not often fully visualized on chest radiographs. Therefore, further analysis with a larger number of test sets is needed to identify the reason for the errors and improve the performance of the diagnostic software. Fourth, the algorithm has relatively low specificity (0.285). Since this algorithm was developed to perform scoliosis screening, we adjusted the detection threshold as 0.05 rather than 0.5 which has balanced classification performance (Table 2, Fig. 3). Further study is needed to improve the specificity of this algorithm. Fifth, the ACC of the algorithm needs improvement with more training data or external validation, and the application of the DLA platform requires further validation in multicenter and multiethnic trials. Sixth, since our upstream task was trained on anomalous data, we need to validate this method. Schlegl et al. (25) studied anomaly detection using only normal data. On the other hand, our approach was a method to detect normality by learning a normal image distribution. Seventh, the upstream and the downstream data sets were collected randomly in the pool of search result of Picture Archiving and Communication System. But relatively small datasets (1565 training and test sets of HUSH and 2022 of AMC) may reduce the reliability of the algorithm. Further study with a larger dataset is needed to improve the reliability of the algorithm. Lastly, the algorithm diagnosed scoliosis without distinction between right and left thoracic curve pattern. There are reported studies about left thoracic curve patterns and their association with disease. According to Goldberg et al. (27) although an association exists between left thoracic curves and disease, it is not strong enough to determine who should be intensively investigated, to the exclusion of other clinical findings, and it seems inappropriate for a different approach to be adopted on

the basis of scoliosis pattern alone. And according to Wu et al. (28) when a left thoracic curve pattern is present in patients with “idiopathic” scoliosis, especially in male patients or patients with severe curve, strong consideration should be given to the possibility of the presence of neural axis abnormalities. But we assume that the prevalence rate is increasing as adolescents sit for longer periods, and the clinical meanings about gender and curvature side of scoliosis are changing. We planned further study about changes of prevalence and clinical meaning of idiopathic scoliosis.

In conclusion, we proposed GAN as a feature extractor to develop a classifier in chest PA radiographs. This screening tool has high NPV and sensitivity, which will reduce the burden on radiologists diagnosing scoliosis in the health screening chest radiographs.

Author Contributions

Conceptualization, P.S., K.N.; data curation, L.W., S.K., L.J.; formal analysis, L.W., S.K., L.J.; investigation, L.W., S.K., L.J.; methodology, P.S., K.N.; project administration, P.S., K.N.; resources, L.W., S.K., L.J.; software, S.K., L.J.; supervision, Y.S., Y.M.A., C.Y.W., H.G., P.S., K.N.; validation, L.W., S.K.; visualization, L.W., S.K., L.J.; writing—original draft, L.W., S.K.; and writing—review & editing, L.W., S.K., C.Y.W., P.S.

Conflicts of Interest

Yo Won Choi has been a Section Editor of the Journal of the Korean Society of Radiology since 2015; however, he was not involved in the peer reviewer selection, evaluation, or decision process of this article. Otherwise, no other potential conflicts of interest relevant to this article were reported.

Funding

None

Acknowledgments

We would like to thank Editage (www.editage.co.kr) for English language editing.

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. Roach JW. Adolescent idiopathic scoliosis. *Orthop Clin North Am* 1999;30:353-365
3. Baert AL. *Spinal imaging: diagnostic imaging of the spine and spinal cord*. Berlin: Springer Science & Business Media 2007
4. Silva FE, Lenke LG. *Adolescent idiopathic scoliosis*. In Errico TJ, Lonner BS, Moulton AW, eds. *Surgical management of spinal deformities*. Philadelphia: Saunders Elsevier, 2009:97-118
5. Weinstein SL, Zavala DC, Ponseti IV. Idiopathic scoliosis: long-term follow-up and prognosis in untreated patients. *J Bone Joint Surg Am* 1981;63:702-712
6. 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:S8-S22
7. Puijts JE, Hageman MA, Keessen W, van der Meer R, van Wieringen JC. Variation in Cobb angle measurements in scoliosis. *Skeletal Radiol* 1994;23:517-520
8. Morrissy RT, Goldsmith GS, Hall EC, Kehl D, Cowie GH. Measurement of the Cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error. *J Bone Joint Surg Am* 1990;72:320-327
9. Galbusera F, Niemeyer F, Wilke HJ, Bassani T, Casaroli G, Anania C, et al. Fully automated radiological analysis of spinal disorders and deformities: a deep learning approach. *Eur Spine J* 2019;28:951-960
10. Alharbi RH, Alshaye MB, Alkanhal MM, Alharbi NM, Alzahrani MA, Alrehaili OA. Deep learning based algorithm for automatic scoliosis angle measurement. Proceedings of 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS); 2020 Mar 19-21; Riyadh, Saudi Arabia: IEEE; 2020:1-5

11. Tan Z, Yang K, Sun Y, Wu B, Tao H, Hu Y, et al. An automatic scoliosis diagnosis and measurement system based on deep learning. Proceedings of the ROBIO 2018: IEEE International Conference on Robotics and Biomimetics; 2018 Dec 12-15; Kuala Lumpur, Malaysia: IEEE; 2018:439-443
12. Reza AM. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *J VLSI Signal Process Syst Signal Image Video Technol* 2004;38:35-44
13. Karras T, Aittala M, Hellsten J, Laine S, Lehtinen J, Aila T. Training generative adversarial networks with limited data. arXiv 2006.06676 [Preprint]. 2020 [cited 2020 October 7]. Available at. <https://arxiv.org/abs/2006.06676>
14. Gardner MW, Dorling SR. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmos Environ* 1998;32:2627-2636
15. Liu Z, Luo P, Wang X, Tang X. Deep learning face attributes in the wild. In Proceedings of the IEEE International Conference on Computer Vision; 2015 December 13-16; Santiago, Chile: IEEE; 2015:3730-3738
16. Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 16-20; Long Beach, CA, USA: IEEE; 2019:4401-4410
17. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems. Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017); 2017 Dec 4-9; Long Beach, CA, USA: NIPS; 2017:30
18. Kingma DP, Ba J. Adam: a method for stochastic optimization. ArXiv 1412.6980 [Preprint]. 2014 [cited 2019 May 1]. Available at. <https://arxiv.org/abs/1412.6980>
19. He K, Zhang X, Ren S, Sun J. *Identity mappings in deep residual networks*. ECCV 2016. Lecture notes in computer science, vol 9908. Cham: Springer 2016
20. Tu Y, Wang N, Tong F, Chen H. Automatic measurement algorithm of scoliosis Cobb angle based on deep learning. *J Phys Conf Ser* 2019;1187:042100
21. Liu T, Yang Y, Wang Y, Sun M, Fan W, Wu C, et al. Spinal curve assessment of idiopathic scoliosis with a small dataset via a multi-scale keypoint estimation approach. Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers; 2020 Sep 12-17; Virtual Event, Mexico: Association for Computing Machinery; 2020:665-670
22. Kokabu T, Kanai S, Kawakami N, Uno K, Kotani T, Suzuki T, et al. An algorithm for using deep learning convolutional neural networks with three dimensional depth sensor imaging in scoliosis detection. *Spine J* 2021;21:980-987
23. Yang J, Zhang K, Fan H, Huang Z, Xiang Y, Yang J, et al. Development and validation of deep learning algorithms for scoliosis screening using back images. *Commun Biol* 2019;2:390
24. Zenati H, Foo CS, Lecouat B, Manek G, Chandrasekhar VR. Efficient gan-based anomaly detection. ArXiv 1802.06222 [Preprint]. 2018 [cited 2019 May 1]. Available at. <https://arxiv.org/abs/1802.06222>
25. Schlegl T, Seeböck P, Waldstein SM, Langs G, Schmidt-Erfurth U. f-AnoGAN: fast unsupervised anomaly detection with generative adversarial networks. *Med Image Anal* 2019;54:30-44
26. Oh CH, Kim CG, Lee MS, Yoon SH, Park HC, Park CO. Usefulness of chest radiographs for scoliosis screening: a comparison with thoraco-lumbar standing radiographs. *Yonsei Med J* 2012;53:1183-1189
27. Goldberg CJ, Moore DP, Fogarty EE, Dowling FE. Left thoracic curve patterns and their association with disease. *Spine (Phila Pa 1976)* 1999;24:1228-1233
28. Wu L, Qiu Y, Wang B, Zhu ZZ, Ma WW. The left thoracic curve pattern: a strong predictor for neural axis abnormalities in patients with “idiopathic” scoliosis. *Spine (Phila Pa 1976)* 2010;35:182-185

준지도학습 방법을 이용한 흉부 X선 사진에서 척추측만증의 진단

이우진¹ · 신기원² · 이준수^{3,4} · 유승진¹ · 윤민아⁵ · 최요원¹ · 홍길선⁵ · 김남국^{2,4} · 백상현^{1*}

목적 흉부 X선 사진에서 척추측만증을 조기진단 할 수 있는 딥러닝 기반의 스크리닝 소프트웨어를 준지도학습(semi-supervised generative adversarial network; 이하 GAN) 방법을 이용하여 개발하고자 하였다.

대상과 방법 두 곳의 상급종합병원에서 촬영된 흉부 X선 사진에서 척추측만증을 조기진단할 수 있는 스크리닝 소프트웨어를 개발하기 위하여 GAN 방법이 이용되었다. GAN의 훈련과정에서 경증에서 중증의 척추측만증을 보이는 흉부 X선 사진들을 사용하였으며 upstream task에서 척추측만증의 특징을 학습하고, downstream task에서 정상과 척추측만증을 분류하도록 훈련하였다.

결과 수신자 조작 특성 곡선의 곡선하면적(area under the receiver operating characteristic curve), 음성예측도, 양성예측도, 민감도 및 특이도는 각각 0.856, 0.950, 0.579, 0.985, 0.285이었다.

결론 우리가 GAN 방법을 이용하여 개발한 딥러닝 기반의 스크리닝 소프트웨어는 청소년의 흉부 X선에서 척추측만증을 진단하는데 있어서 높은 음성예측도와 민감도를 보였다. 이 소프트웨어가 건강검진을 목적으로 촬영한 청소년의 흉부 X선 사진에 진단 스크리닝 도구로써 이용된다면 영상의학과 의사의 부담을 덜어주며, 척추측만증의 조기진단에 기여할 것으로 생각된다.

¹한양대학교병원 영상의학과,

²울산대학교 의과대학 서울아산병원 아산융합의학원 의공학과,

³서울대학교 산업공학과,

울산대학교 의과대학 서울아산병원 ⁴융합의학과, ⁵영상의학과