Original Article

AO SPINE

Global Spine Journal
2025, Vol. 15(4) 2309–2317

DOI: 10.1177/21925682241299332
journals.sagepub.com/home/gsj
Sage

Lumbar Spinal Stenosis Grading in Multiple Level Magnetic Resonance Imaging Using Deep Convolutional Neural Networks

Dongkyu Won, BS[1], Hyun-Joo Lee, MD, PhD[2,3], Suk-Joong Lee, MD, PhD[4]*, and Sang Hyun Park, PhD[1]*

Abstract

Study Design: Retrospective magnetic resonance imaging grading with comparison between experts and deep convolutional neural networks (CNNs).

Objective: The application of deep learning to clinical diagnosis has gained popularity. This approach can accelerate image interpretation and serve as a screening tool to help doctors.

Methods: A comparison was conducted between retrospective magnetic resonance imaging (MRI) grading performed by experts and grading obtained using CNN classifiers. Data were collected from the lumbar axial dataset in the DICOM format. Two experts labeled the sampled images using the same diagnostic tools: localization of patches near the spinal canal, rootlet leveling, and stenosis grading. Comprehensive comparisons were presented for both rootlet cord classification and stenosis grading.

Results: Rootlet-cord classification for the two analyzers was 90.3% and the F1 score was 86.6%. The agreement of Analyzers-Classifiers was 92.7% and 96.8% for data with 90.6% and 95.6% F1 scores, respectively. For stenosis grading, there was an agreement of 89.2% between the two analyzers, resulting in an F1 score of 76.5%. The grades of the Analyzers-Classifiers agreed on 91.5/89.4% of the data, with an F1 score of 78.4/75.7%. Analyzer1 and Analyzer2 classified >74% as grade A (78.8% and 74.4%, respectively), 15.4% and 18.6% as grade B, 4.2% and 6.0% as grade C, and 1.6% and 2.0% as grade D, respectively.

Conclusions: The fully automated deep learning model showed competitive results in stenosis grade diagnosis and rootlet cord classification under similar anatomical conditions. However, abrupt anatomical changes can lead to a puzzle diagnosis based only on images.

[1] Department of Robotics and Mechatronics Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, Korea
[2] Department of Orthopaedic Surgery, School of Medicine, Kyungpook National University, Daegu, Korea
[3] Institute of Medical Device and Robot, Kyungpook National University, Daegu, Korea
[4] Department of Orthopaedic Surgery, Gyeongsang National University College of Medicine and Gyeongsang National University Changwon Hospital, Changwon, Korea

*Suk-Joong Lee and Sang Hyun Park equally contributed to this manuscript as corresponding authors.

Corresponding Authors:
Suk-Joong Lee, M.D., Department of Orthopaedic Surgery, Gyeongsang National University, College of Medicine, Gyeongsang National University Changwon Hospital, 11 Samjeongja-ro, Seongsan-gu, Changwon-si 51472, Korea.
Email: sjleeleesj@gnuh.co.kr

Sang Hyun Park, PhD, Department of Robotics and Mechatronics Engineering, Daegu Gyeongbuk Institute of Science and Technology, 333 Techno Jungang Daero, Hyeonpung-Myeon, Dalseong-Gun, Daegu 42988, Korea.
Email: shpark13135@dgist.ac.kr

Keywords
lumbar, spinal stenosis, multiple level, grading, deep convolutional neural networks

Introduction

Lumbar spinal stenosis is a disease of spinal canal narrowing by hypertrophied ligaments or bone spurs. Even though sagittal image is required to enhance the diagnostic accuracy,[1] dural sac cross-sectional surface area in T2-weighted axial MR images of the lumbar spine is crucial for diagnosis. The grade is frequently determined by the morphological characteristics of the cerebrospinal fluid and rootlet contents in proximity to the dural sac.[2,3]

Artificial intelligence (AI) in medicine has gained popularity owing to its high-precision and time-saving capabilities. AI has the potential to assist doctors in accomplishing repetitive and time-consuming tasks such as reading multiple medical images and automatically identifying medically relevant indicators.

In spinal diseases, approximately 90 studies have proven the usability of deep learning to help clinicians diagnose and predict prognosis and outcome.[4] Lehnen et al. reported the feasibility of using AI to detect degenerative changes by segmenting tissues in the spinal canal and classifying the changes using convolutional neural networks (CNNs).[5] Ouyang et al[6] utilized AI to evaluate a detection model specifically designed for spinal tumors on magnetic resonance imaging (MRI), demonstrating its effectiveness.[6] Wang et al[7] demonstrated the efficacy of a deep learning model for identifying intervertebral foraminal stenosis in the postoperative view.[7] Additionally, Yeh et al. and to analyze the diagnostic performances and errors between human and deep learning models for vertebral fractures on MRI.[8]

Won et al. showed the reliability of AI for the diagnosis of spinal stenosis at the uni-level, exhibiting a strong correlation with clinical experts.[9] These studies collectively highlight the potential of AI in enhancing diagnostic accuracy and clinical decision-making. However, the previous study was limited in its scope, as it focused solely on the L4/5 level; therefore, doctors needed to manually select and input the images of the 4/5 level from multiple images of a single patient into the artificial intelligence system. To broaden the clinical utility of this study, we expanded our training dataset to include all imaging data from the entire lumbar region and its related levels. By incorporating this comprehensive dataset, we developed a multilevel automatic detection system capable of identifying narrow spinal canals across various levels and then classifying the grades of intervertebral stenosis. Thus, the proposed technique can automatically assess the characteristics of all vertebrae and provide grading without the need for manual intervention once MR images are acquired.

The primary objective of this study was to evaluate the agreement between the grades labeled by the two experts and classifiers using a multilevel dataset from each expert. In addition, we investigated the transferability of the deep learning algorithm trained at the lumbar level to the cord level.

Materials and Methods

Dataset

The Institutional Review Board of our hospital approved the use of this dataset. Initially, 13,758 lumbar spine MR images from 542 consecutive patients who performed lumbar spine MRI were used. These images were extracted from at least 15-46 slices, regardless of the rootlet and cord level, between all MR images of each patient. Axial mages were obtained with a slice thickness of 4 mm. The data samples were collected in Digital Imaging and Communications in Medicine (DICOM) format from the Picture Archiving and Communication System, which contained T2-weighted axial images of the lumbar, lower thoracic, and upper sacral spine. All MR images per patient were collected as a set, regardless of level. Furthermore, we collected T2 axial images using Siemens (1.5 T) and used them for the analysis. During the data collection process, there was no communication between the analyzers, and the final collected data were selected as the intersection of the data collected from each analyzer.

Classification by Two Experts

The images were classified and graded by two expert surgeons who assessed the center of the spinal canal, rootlet/cord level, and stenosis grade. Specifically, two analyzers labeled the locations of the spinal canal and determined the level at which the spinal canal was assigned between the rootlet and cord. The images were graded into one of the four levels using an existing grading tool for stenosis grading. Among the collected MR images, images of difficult cases for clinical operations or those lacking quality were discarded in our classifier training. Analyzer1 is a spine surgeon with over 10 years of experience, while Analyzer2 is an orthopedic surgeon with an equivalent level of clinical practice. There was no communication between the experts for any labeling process, such as localization of the spinal canal, rootlet-cord classification, or stenosis grading.

Stenosis Grade Classification Using Deep Convoluted Neural Network (CNN)

Using the labels generated by the two surgeons, we constructed a fully automatic diagnostic framework for stenosis grading. The framework for the stenosis grading is shown in Figure 1. This framework consists of 3 parts: a canal detector for localizing the spinal canal in collected images, which is the

bounding box; a rootlet classifier that classifies the rootlet and cord levels from the patches extracted by the detector; and a stenosis classifier that classifies the stenosis grade of predicted rootlet patches from the rootlet classifier.

To locate the spinal canal, Faster R-CNN,[10] which trains the bounding boxes of the spinal canal labeled by experts, was used. This model consists of one backbone network, region proposal network (RPN), and region of interest (ROI) pooling layer. The backbone network, comprising ImageNet pre-trained ResNet50,[11] extracts the features of the input MR images. Using features from the backbone, the RPN generates bounding-box proposals and probability scores using a regression and classification module. The regression module minimizes the error between the ground truth and the predicted bounding boxes, and the classification module learns whether the class in the predicted bounding box is assigned to the spinal canal. To match the sizes of the feature maps from the RPN, ROI pooling was utilized so that the spinal canal detector could always train feature maps of the same size. The predicted bounding boxes and their scores were predicted by placing the features from the ROI pooling layers into the fully connected layers.

For rootlet-cord classification, we trained a Visual Geometry Group (VGG)[12] network using the extracted bounding box from the detector, where the spinal canal was located. We used a VGG network with 10 convolutional layers, 4 max pooling layers, 1 global average pooling layer, and 1 fully connected layer. As the network learns an image, the convolutional layers extract meaningful features of the rootlet on the input image. To calculate the error between the labels and predictions, we used the binary cross-entropy loss. The model was optimized using an ADAM[13] optimizer for 200 epochs. The learning rate was initially set to 0.0001 and was decreased by 50% when the validation accuracy did not increase from the last best validation accuracy. The best model was selected with high accuracy during validation.

For stenosis grading, we trained a VGG[12] likewise rootlet classifier using the predicted rootlet patches from the rootlet classifier. A fully connected layer with four outputs, that is, the final stenosis grade level, was used in our stenosis classifier. Softmax with cross-entropy loss was calculated to measure the error between the predictions and labels. The model was optimized using the ADAM[13] optimizer and was trained for 100 epochs. The initial learning rate was set to 0.0001 and reduced by 50% when the validation accuracy did not improve during the validation steps. The model with the highest accuracy for the validation dataset was chosen.

In the testing stage, our trained Faster R-CNN detects the bounding box of the spinal canal from a test MR image. The rootlet classifier subsequently classifies whether the patch extracted from the bounding box is at the rootlet or cord. Finally, the stenosis classifier predicted the grade with the

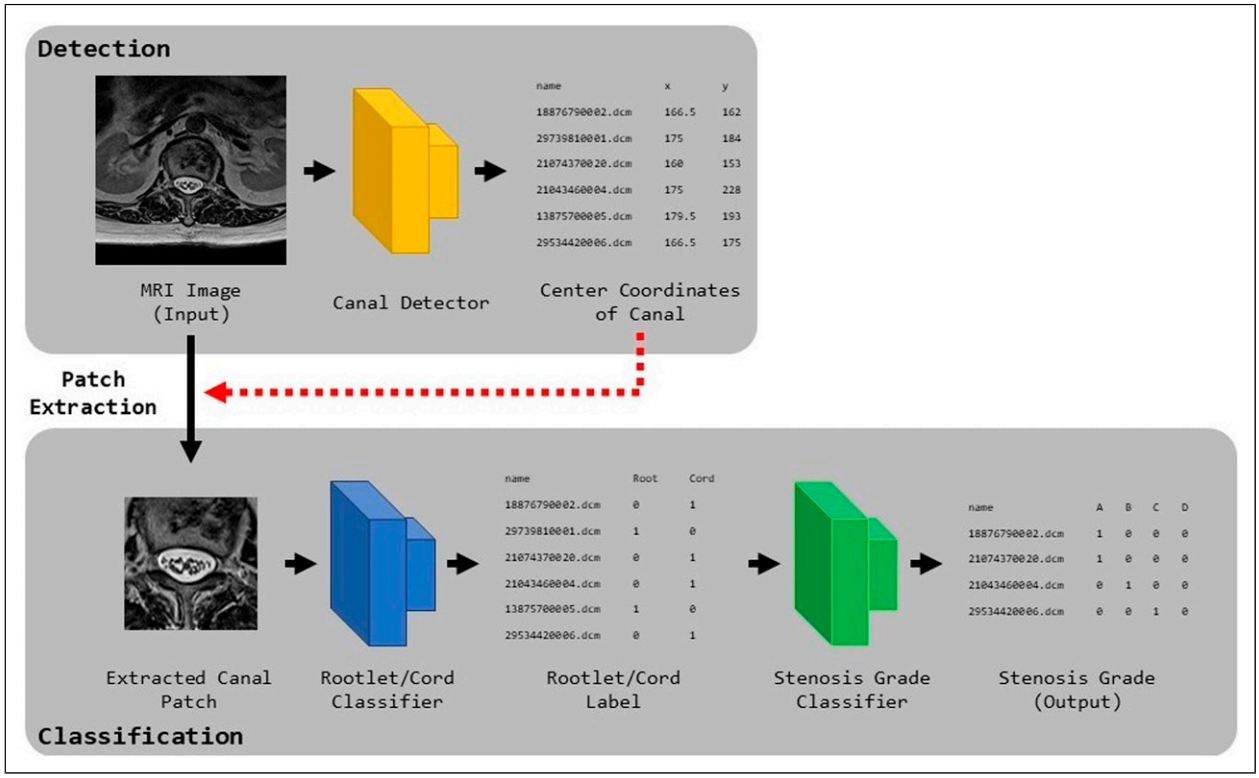

Figure 1. Our deep learning based spinal stenosis grading system. Our grading system consists of 3 stages; Canal Detector, Rootlet/Cord Classifier and Stenosis Classifier. Canal Detector predicts the spinal canal region, and the classifiers predict rootlet/cord level and stenosis grade on localized canal region from the detector, respectively.

highest probability among the four probabilities on the rootlet patches as the final stenosis.

To evaluate our framework, we divided our collected dataset into 10 sets and performed 10-fold cross-validation, where 1-fold training was formed with 7 training, two validation, and one test data. Each set was stratified by constructing an equal distribution of the grade levels.

Since our collected dataset was highly imbalanced (larger number of intervertebral discs and A and B grades on stenosis), we performed weighted sampling and data augmentation. Specifically, weighted sampling samples an even number of levels and grades on the mini-batch to ensure that the model trains equally on each label. For data augmentation, we used random rotation, scaling, translation, and horizontal flipping at the image level, and zero-mean Gaussian noise at the pixel level. Our experiments were performed on a GPU server with an Intel Xeon Silver 4210 2.2 GHz CPU, NVIDIA RTX 2080Ti, and 256 GB RAM, with a PyTorch implementation.

Comparison Between Inter-observer Agreement and Observer-Classifier Agreement

To evaluate our framework, the inter-observer and observer-classifier agreements of the rootlet and stenosis classifiers were measured. The inter-observer agreement reflects the degree of agreement in the diagnostic results between analyzers, whereas the observer-classifier agreement represents the diagnostic agreements between analyzers and classifiers.

For a fair comparison, the inter-observer and observer-classifier agreements were compared at the intersection of the data used in both analyzers' labels. To represent the agreements visually, we generated confusion matrices to illustrate the agreement between the following pairs: Analyzer1 and Analyzer2, Analyzer1 and Classifier1 trained using Analyzer1's labels, and Analyzer2 and Classifier2 trained using Analyzer2's labels. We further provided confusion matrices for each task, rootlet-level classification, and stenosis-grade classification. To validate the efficacy of our framework quantitatively, we utilized accuracy, F1 scores, and a two-sample paired t-test to assess the statistical significance of the variations in diagnostic outcomes.

Table 1. Dataset.

Stage	Analyzer1	Analyzer2	Intersection
Collected	13,758		
Rootlet + cord	11,551	12,442	10,969
Rootlet	8443	9613	7693
Predicted rootlet + cord	11,542	12,436	10,965
Predicted rootlet	8092	8387	7282

Table 2. Confusion Matrix Between Analyzer1 and Analyzer2 in Intervertebral-Cord Classification.

		Analyzer1		
		Rootlet	Cord	
Analyzer2	Rootlet	7693	694	8387 (76.5%)
	Cord	399	2179	2578 (23.5%)
		8092 (73.8%)	2873 (26.2%)	10,965

Results

Number of Labeled Data

The collected dataset size was 13,758, for which the analyzers labeled samples into 3 categories: rootlet, cord, and others (invalid for training owing to interference from screws or deformity). Among these samples, 11,551 and 12,442 slices were labeled as rootlet or cord levels per analyzer, respectively, and the deep learning-based detection models were trained using these data under a cross-validation setting. Consequently, the spinal canal detectors predicted 11,542 and 12,436 rootlet and cord slices, respectively, and the rootlet classifiers predicted 8092 and 8387 rootlet cases, respectively. Finally, 8092 and 8387 samples were used to train the stenosis classifiers (Table 1), and a comparison of the predictions was performed on 7282 intersection cases.

Inter-observer Agreement

The rootlet/cord classifications created by the two analyzers are summarized in Table 2. Analyzer1 and Analyzer2 classified >73% as rootlets and 26.2% and 23.5% as cords, respectively.

The stenosis grading results generated by the two analyzers for the intersection cases are summarized in Table 3. In Table 2, even though the number of read data that Analyzer1

Table 3. Confusion Matrix Between Analyzer1 and Analyzer2 in Stenosis Grading.

		Analyzer1				
		A	B	C	D	
Analyzer2	A	5312	107	3	0	5422 (74.4%)
	B	422	832	18	8	1280 (18.6%)
	C	4	152	263	15	434 (6.0%)
	D	1	31	22	92	146 (2.0%)
		5739 (78.8%)	1122 (15.4%)	306 (4.2%)	115 (1.6%)	7282

Table 4. The F-Score, Accuracy, and *P* Value of Rootlet-Cord and Stenosis Grade Classification Models.

	Rootlet-Cord Classification			Stenosis Grading		
	Analyzer1-Analyzer2	Analyzer1-Classifier1	Analyzer2-Classifier2	Analyzer1-Analyzer2	Analyzer1-Classifier1	Analyzer2-Classifier2
F-score (%)	86.6	90.6	95.6	76.5	78.4	75.7
Accuracy (%)	90.3	92.7	96.8	89.2	91.5	89.4
P	<0.01	0.7128	0.2741	0.01	0.6739	0.2329

and 2 recognized as rootlets was 7693, only 7282 of these were analyzed by the classifier. For further analysis between analyzers and classifiers, only 7282 cases were analyzed (Table 3). Analyzer1 and Analyzer2 classified more than 74% of the samples as Grade A, with Analyzer1 assigning 15.4% to Grade B and Analyzer2 assigning 18.6%. Additionally, they assigned 4.2% and 6.0% to Grade C and 1.6% and 2.0% to Grade D, respectively.

Table 5. Confusion Matrix Between Analyzers and Rootlet-Cord Classifiers.

		Classifier1/Classifier2		
		Rootlet	Cord	
Analyzer1/Analyzer2	Rootlet	7683/8180	409/207	8092/8387
	Cord	385/138	2488/2440	2873/2578
		8068/8318	2897/2647	10965

Observer-Classifier Agreement

The rootlet/cord classification generated by the accuracy and F1-score of the two analyzers is shown on the left side of Table 4. Of these, 90.3% agreed with the two analyzers and the F1 score was 86.6% in the rootlet/cord classification. The accuracy and F-score between Analyzer1-Classifier1 and Analyzer2-Classifier2 are higher than those between Analyzer1-Analyzer2.

The stenosis grading agreement and F1-score are shown on the right side of Table 4. The two analyzers agreed on 89.2% of the data, resulting in an F1 score of 76.5%. The accuracy and F-score between Analyzer1-Classifier1 and Analyzer2-Classifier2 are comparable to or higher than those between Analyzer1-Analyzer2. There was a significant difference between Analyzer1 and 2 in both rootlet/cord classification and stenosis grading. Conversely, the differences between the analyzers and classifiers were not significant.

Tables 5 and 6 show the confusion matrices of the analyzer classifier for rootlet-cord classification and stenosis grading, respectively. Compared to the agreement between Analyzer1 and Analyzer2 in rootlet cord classification, Analyzer1-Classifier1 and Analyzer2-Classifier2 were able to classify most rootlet slices. We have provided the label consistencies of classifiers and analyzers in the supplementary data with agreement percentages (Supplements 1-4 in the Supplementary Information). In stenosis grading, when comparing the label consistency of Classifier1 and Analyzer2 with respect to the labels of Analyzer1, Classifier1 showed higher agreement than Analyzer2 in Grade A (96.0% vs 92.5%, each percentage was calculated based on Analyzer 1 and 2; Supplements 1 and 2 show the percentage numbers) and Grade B (76.5% vs 74.1%), but lower agreements in Grade C (71.2% vs 85.9%) and Grade D (64.3% vs 80%). Conversely, when comparing the label consistency of Classifier2 and Analyzer1 with respect to the labels of Analyzer2, Classifier2 showed higher agreements than Analyzer1 in Grade B (73.2% vs 65%) and Grade C (69.5% vs 60.6%), but lower agreements in Grade A (95.8% vs 97.9%) and Grade D (58.2% vs 63%). For both cases, the classifier agreement was similar to or higher than that of other analyzers in Grades A and B, which included most of the data, but the agreement between analyzers was higher in Grade D, where there was little data available.

Discussion

In summary, this study showed similar diagnostic agreement levels between experts and the agreement between experts and CNN classifiers trained in two specific areas: 1) distinguishing between rootlet and cord levels, and 2) grading stenosis across multiple lumbar levels. In general, the agreement between the analyzers and classifiers ranged from 91.5% to 92.7%, surpassing the level of agreement seen between Analyzer1 and Analyzer2, which was 89.2%. In both the rootlet/cord classification and stenosis grading, analyzers 1 and 2 showed significant differences. However, there were no significant differences between the Analyzers and Classifiers. The agreement between Analyzers and Classifiers for stenosis grading in this study was higher than that in a previous study (77.9%–83%).[6] We confirmed that the decisions generated by deep learning are reasonable for spinal stenosis grading and rootlet level differentiation from the cord level, at least at the entire lumbar level, with one grading system.

The classifier which had been trained at a single level can be applied to other lumbar levels with a higher level of consistency. The classified samples for the lumbar levels are presented in Figures 2 and 3. We speculate that this is possible because of the similar anatomical morphology of other lumbar levels with the single L4/5 level. To develop a more clinical

Table 6. Confusion Matrix Between Analyzers and Stenosis Grade Classifiers.

		Classifier1/Classifier2				
		A	B	C	D	
Analyzer1/Analyzer2	A	5512/5188	226/226	1/8	0/0	5739/5422
	B	205/258	859/937	48/78	10/7	1122/1280
	C	2/6	71/100	218/302	15/26	306/434
	D	0/4	21/18	20/39	74/85	115/146
		5719/5456	1177/1281	287/427	99/118	7282

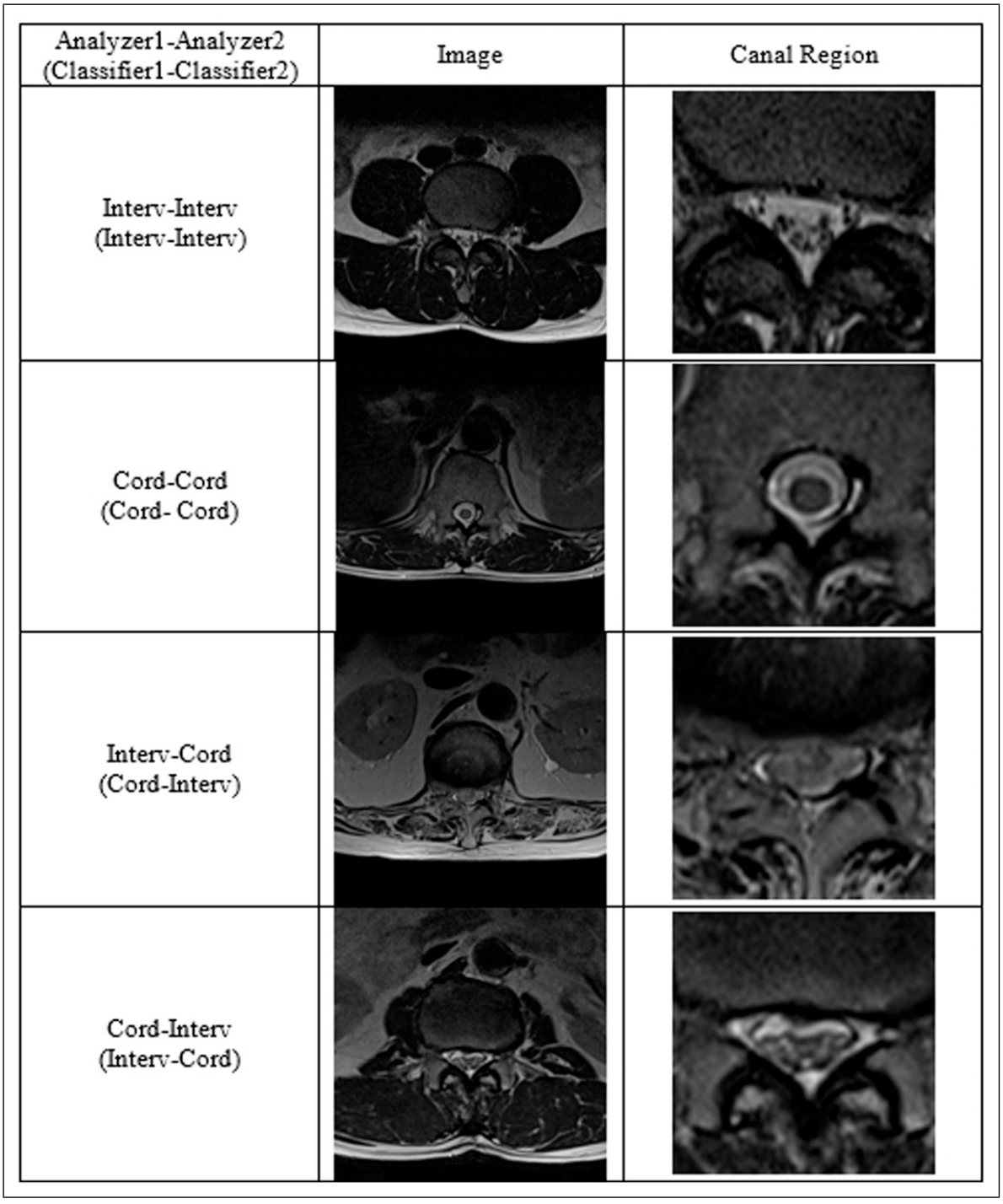

Figure 2. Rootlet/Cord classification cases between analyzers or an analyzer and a classifier.

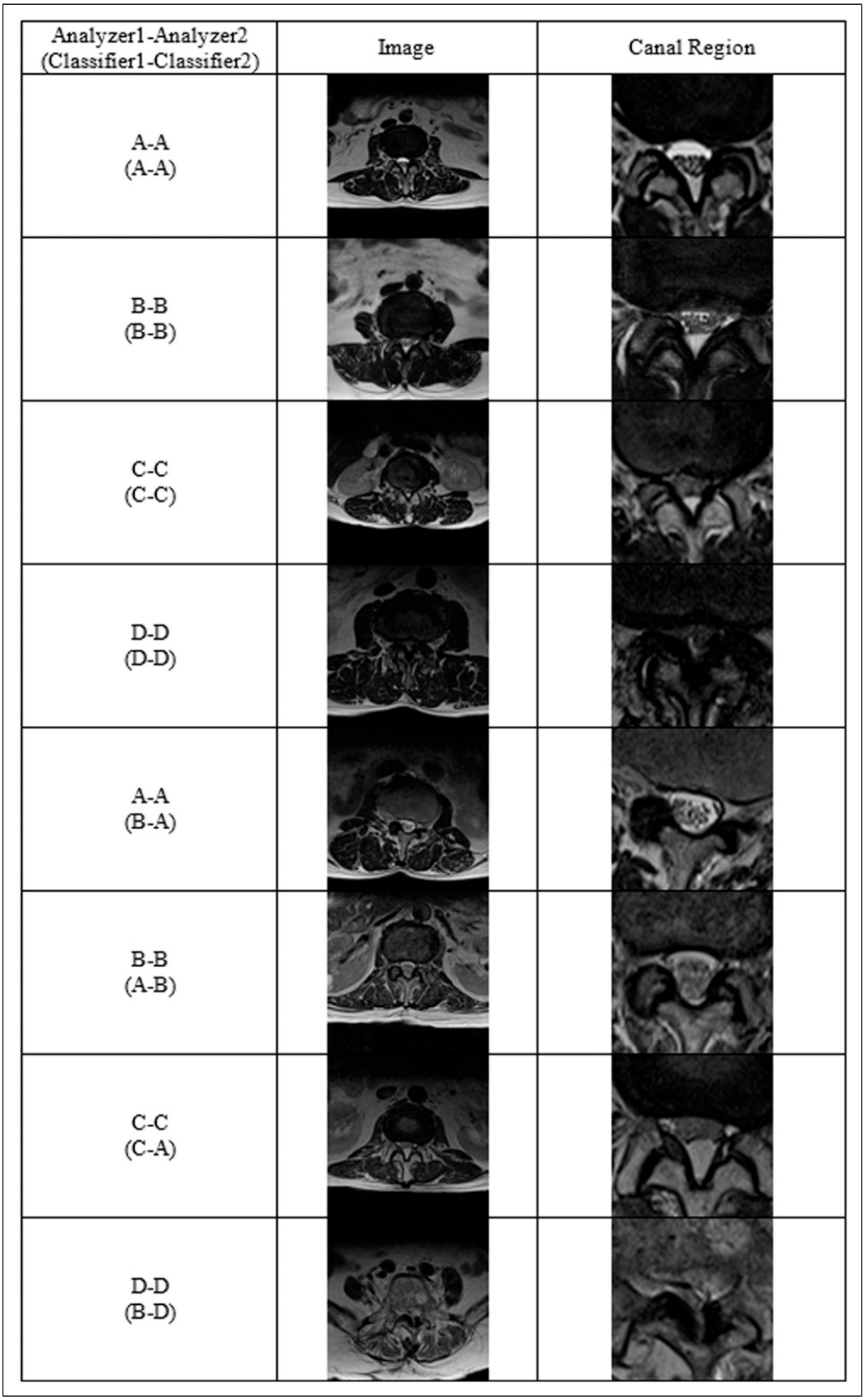

Figure 3. Matched cases between analyzers.

situation-based deep learning model for determining the grade of spinal stenosis, extensibility to multiple levels is necessary. Although single- to multiple-level transfer learning is possible, the specific level that contains the spinal cord results in incorrect categorization by the algorithm. Thus, for a versatile diagnostic tool using deep learning, an additional rootlet classifier with cord compression classification is required at the cord level.

At the spinal cord level, the disagreement between analyzers and classifiers was higher than at other levels where the spinal cord was not present. We assumed that the classifier was confused by spinal cord morphology, which mimicked the aggregated spinal root, as it did not have information on spinal cord morphology. Therefore, we speculated that training data at the cord level is needed to broaden the deep learning model to the thoracic lumbar level.

Image-based diagnosis is a key procedure in the final diagnosis of spinal stenosis. AI-based image interpretation has gained popularity to enhance diagnostic accuracy and for fast detection excluding critical human error.[8,14,15] The spine has multiple levels with a similar shape, and requires considerable time for level checking and diagnostic decisions. As Herzog described, more clinically practical tools are needed to perform relevant analyses.[16]

We found that the highly imbalanced dataset across the grades caused current results, and the data augmentation and class-wise sampling techniques used in this experiment were insufficient to cover the sample diversity of severe cases such as grades C and D. Nevertheless, we found that trained classifiers still perform competitive diagnosis in grades A and B, similar to human analyzers. We expect that more samples or undersampling processes with respect to Grades C and D for sample diversity could resolve the current issues.

This study had several limitations. Despite our efforts to collect a substantial MRI dataset, an imbalance in grades occurred within the training dataset, primarily owing to the relatively smaller number of collected images for Grades C and D compared to Grades A and B. In this study, we did not consider the priority of under-diagnosis or over-diagnosis. However, high sensitivity is often required for screening. To use this model as a screening tool, a loss function using regression or classification is required. Therefore, overdiagnosis should be considered. Furthermore, to improve diagnostic accuracy, the inclusion of sagittal images is essential. We anticipate that the effective utilization of these images will necessitate the development of more advanced segmentation techniques and the application of multiple classifiers. This will be a focus of our future research.

Conclusion

Our trained deep learning model can be used when anatomical conditions are similar. However, anatomical changes can lead to a puzzle diagnosis based only on images. Doctors should be involved in interpreting the model for medical diagnosis.

Appendix

List of Abbreviations

CNN	convolutional neural network
AI	artificial intelligence
MRI	magnetic resonance imaging
DICOM	Digital Imaging and Communications in Medicine
RPN	region proposal network
ROI	region of interest
VGG	Visual Geometry Group

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (**NRF-2022R1F1A1068306).**

Authors' Contributions

HJL and SJL analyzed and interpreted the image data. WD and SJL wrote the manuscript. HJL,SJL and SHP reviewed and edited the manuscript. WD and SHP analyzed the interpreted data using this model. All authors have read and approved the final manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-2022R1F1A1068306).

Ethical Statement

Ethical Approval

The institutional review board of Gyeongsang National University Changwon Hospital approved this study (2021–11–023).

Informed Consent

The requirement for informed consent to participate was waived by the Institutional Review Board of Gyeongsang National University Changwon Hospital. All experiments were performed in accordance with the relevant guidelines and regulations.

ORCID iD

Suk-Joong Lee https://orcid.org/0000-0002-6769-6474

Data Availability Statement

The datasets generated and/or analyzed during the current study are not publicly available because of privacy issues but are available from the corresponding author upon reasonable request.

Supplemental Material

Supplemental material for this article is available online.

References

1. Jaganjac B, Džidić-Krivić A, Bečulić H, Šljivo A, Begagić E, Šišić A. Magnetic resonance morphometry of the lumbar spinal canal in Zenica - doboj Canton in Bosnia and Herzegovina. *Med Glas*. 2023;20(2), 20. doi:10.17392/1575-23
2. Lønne G, Ødegård B, Johnsen LG, et al. MRI evaluation of lumbar spinal stenosis: is a rapid visual assessment as good as area measurement? *Eur Spine J*. 2014;23. doi:10.1007/s00586-014-3248-4
3. Schizas C, Theumann N, Burn A, et al. Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images. *Spine*. 2010;35:1919-1924. doi:10.1097/BRS.0b013e3181d359bd
4. Azimi P, Yazdanian T, Benzel EC, et al. A review on the use of artificial intelligence in spinal diseases. *Asian Spine J*. 2020;14(4):543-571. doi:10.31616/asj.2020.0147
5. Lehnen NC, Haase R, Faber J, et al. Detection of degenerative changes on MR images of the lumbar spine with a convolutional neural network: a feasibility study. *Diagnostics*. 2021;11(5):902. doi:10.3390/diagnostics11050902
6. Ouyang H, Meng F, Liu J, et al. Evaluation of deep learning-based automated detection of primary spine tumors on MRI using the turing test. *Front Oncol*. 2022;12:814667. doi:10.3389/fonc.2022.814667
7. Wang H, Liu Y, Li Y. Study on automatic multi-classification of spine based on deep learning and postoperative infection screening. *J Healthc Eng*. 2022;2022(1):2779686. doi:10.1155/2022/2779686
8. Yeh LR, Zhang Y, Chen JH, et al. A deep learning-based method for the diagnosis of vertebral fractures on spine MRI: retrospective training and validation of ResNet. *Eur Spine J*. 2022;31(8):2022-2030. doi:10.1007/s00586-022-07121-1
9. Won D, Lee HJ, Lee SJ, Park SH. Spinal stenosis grading in magnetic resonance imaging using deep convolutional neural networks. *Spine*. 2020;45(12):804-812. doi:10.1097/BRS.0000000000003377
10. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. 2016;39(6):1137-1149. doi:10.1109/TPAMI.2016.2577031
11. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on computer Vision and pattern recognition (CVPR)*. 2016;2015:770–778. doi:10.1109/CVPR.2016.90
12. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *Paper Presented at International Conference on Learning Representations*; 2014. [location]. Accessed [Month dd,yyyy].https://arxiv.org/pdf/1409.1556.pdf
13. Kingma DP, Ba J. Adam: a method for stochastic optimization. In: *Paper Presented at International Conference on Learning Representations*; 2014. [location]. Accessed [Month dd,yyyy].https://arxiv.org/pdf/1412.6980.pdf
14. Suri A, Jones BC, Ng G, et al. Vertebral deformity measurements at MRI, CT, and radiography using deep learning. *Radiol Artif Intell*. 2021;4(1):e210015. doi:10.1148/ryai.2021210015
15. Tamai K, Terai H, Hoshino M, et al. A deep learning algorithm to identify cervical ossification of posterior longitudinal ligaments on radiography. *Sci Rep*. 2022;12(1):2113. doi:10.1038/s41598-022-06140-8
16. Herzog RJ. Point of view: spinal stenosis grading in magnetic resonance imaging using deep convolutional neural networks. *Spine*. 2020;45(12):813. doi:10.1097/BRS.0000000000003384