An automatic diagnosis system with deep learning algorithm for lumbar spinal canal stenosis using lumbar spine x-rays

Hisataka Suzuki[1,2], Terufumi Kokabu[1], Katsuhisa Yamada[2], Yoko Ishikawa[1], Akito Yabu[1], Takahiko Hyakumachi[1], Yasushi Yanagibashi[1], Norimasa Iwasaki[2]

[1]Department of Orthopedic Surgery, Eniwa hospital, Japan, [2]Department of Orthopedic Surgery, Hokkaido University, Japan

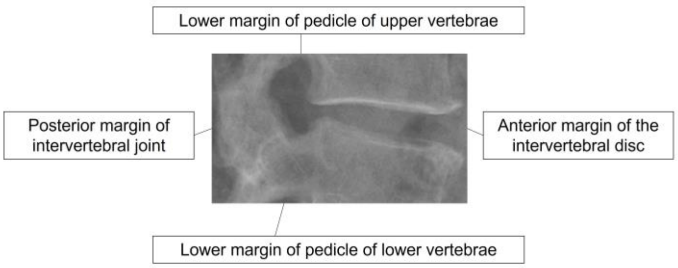
Email of Presenting Author: hisataka_q@yahoo.co.jp

Disclosures: Hisataka Suzuki (N), Katsuhisa Yamada (N), Terufumi Kokabu (N), Yoko Ishikawa (N), Akito Yabu (N), Takahiko Hyakumachi (N), Yasushi Yanagibashi (N), Norimasa Iwasaki (N)

INTRODUCTION: Lumbar spinal stenosis (LSS) is a very common disease in the elderly, and it is reported that one per 1000 persons undergoes the surgery. Although magnetic resonance imaging (MRI) of the lumbar spine is necessary for the diagnosis, it can be difficult for non-specialist to read MRI properly. Furthermore, clinics may not have MRI equipment. In this study, we aimed to develop a system to diagnose the presence or absence of LSS from the lumbar spine radiographs using deep learning.

METHODS: Seventy-five patients who underwent the surgery for LSS other than L1/2 level including degenerative spondylolisthesis at our hospital were enrolled. We randomly divided 75 cases into 50 cases for the internal validation and 25 cases for the external validation. As the region of interest, 4 intervertebral levels from L1/2 to L4/5 were extracted from lateral X-ray images of the lumbar spine according to the margin as Figure 1. Totally, 200 images for the internal validation and 100 images for the external validation were obtained. Each image was annotated the following three types of information; 1, the spinal canal area rate which was calculated by each disc level area divided by L1/2 disc level area measured on axial images of MRI; 2, binary classification for operative level of nonoperative level; 3, binary classification for fusion surgery level of non-fusion surgery level. In the internal validation, five-fold cross validation with 160 data for training and 40 data for validation was performed. Of the five datasets in the internal validation, the learned algorithm with the highest accuracy for annotations 1-3 was used for the external validation.

Figure 1. Representative example of a region of interest extracted from a lateral X-ray image of the lumbar spine.



RESULTS: Five-fold cross-validation of annotation 1 yielded correlation coefficients of 0.47 to 0.67 (All P<0.01). The Area Under the Curve (AUC) calculated from the Receiver Operatorating Characteristic curve (ROC) of Annotation 2 were 0.82 to 0.94. The accuracies were 70% to 90% in Annotation 2. The AUC of Annotation 3 were 0.74 to 0.91. The accuracies were 80% to 93% in Annotation 3. The ROC for Dataset 2 with the highest AUC value in the internal validation are shown in Figure 2. In all of annotation 1, 2 and 3, the heat map showed high feature density in the intervertebral joints and posterior intervertebral discs. In the external validation deriving from learned algorithm with Dataset 2, the correlation coefficient for annotation 1 was 0.58 (P<0.01), and for annotations 2 and 3, the AUC was 0.84 and 0.88 with an accuracy of 72% and 86%, respectively (Table 1).

DISCUSSION: In present study, the correlation coefficient between actual and predicted spinal canal area rate were 0.47 to 0.67 in internal validation, and 0.58 in external validation. In addition, the accuracy for binary classification for operative level of nonoperative level and fusion surgery level of non-fusion surgery level were 70% to 90% and 80% to 93% in internal validation, and 72% and 86% in external validation, respectively. These results indicated that these deep learning algorithms can diagnose LSS based on X-ray images of the lumbar spine. This technology can be useful for clinics without MRI and rural areas where there are no orthopedic surgeons.

SIGNIFICANCE/CLINICAL RELEVANCE: We developed deep learning algorithms that can diagnose LSS based on X-rays of lumbar spine. This technology can be useful for clinics without MRI and rural areas where have no orthopedic surgeons.

Figure 2. Representative images of ROC curves for detecting operative level or nonoperative level and fusion surgery level or non-fusion surgery level in internal validation

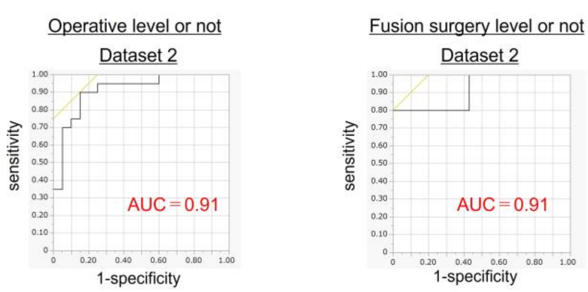


Table 1. Experimental indicators for detecting surgical and fusion surgery indications in external validation.

External validation	Sensitivity	Specificity	PPV	NPV	Accuracy	PLR	NLR
Surgery or not	0.61	0.90	0.90	0.59	0.72	5.91	0.44
Fusion or not	0.75	0.87	0.33	0.98	0.86	5.75	0.29

PPV= positive predictive value, NPV= negative predictive value,
PLR= positive likelihood ratio, NLR= negative likelihood ratio