

Deep Learning for Adjacent Segment Disease at Preoperative MRI for Cervical Radiculopathy

Background: Patients who undergo surgery for cervical radiculopathy are at risk for developing adjacent segment disease (ASD). Identifying patients who will develop ASD remains challenging for clinicians.

Research question: Can we develop and validate a deep learning algorithm capable of predicting ASD by using only preoperative cervical MRI in patients undergoing single-level anterior cervical discectomy and fusion (ACDF)?

Methods: In this Health Insurance Portability and Accountability Act-compliant study, retrospective chart review was performed for 1244 patients undergoing single-level ACDF in two tertiary care centers. After application of inclusion and exclusion criteria, 344 patients were included, of whom 60% ($n = 208$) were used for training and 40% for validation ($n = 43$) and testing ($n = 93$). A deep learning-based prediction model with 48 convolutional layers was designed and trained by using preoperative T2-sagittal cervical MRI. To validate model performance, a neuroradiologist and neurosurgeon independently provided ASD predictions for the test set. Validation metrics included accuracy, areas under the curve, and F1 scores. The difference in proportion of wrongful predictions between the model and clinician was statistically tested by using the McNemar test.

Results: A total of 344 patients (median age, 48 years; interquartile range, 41-58 years; 182 women) were evaluated. The model predicted ASD on the 93 test images with an accuracy of 88 of 93 (95%; 95% CI: 90, 99), sensitivity of 12 of 15 (80%; 95% CI: 60, 100), and specificity of 76 of 78 (97%; 95% CI: 94, 100). The neuroradiologist and neurosurgeon provided predictions with lower accuracy (54 of 93; 58%; 95% CI: 48, 68), sensitivity (nine of 15; 60%; 95% CI: 35, 85), and specificity (45 of 78; 58%; 95% CI: 56, 77) compared with the algorithm. The McNemar test on the contingency table demonstrated that the proportion of wrongful predictions was significantly lower by the model (test statistic, 2.000; $P < .001$). Saliency maps were generated to illustrate what part of the image was most important in the algorithm's decision making (Figure 1).

Discussion: A deep learning algorithm that used only preoperative cervical T2-weighted MRI outperformed clinical experts at predicting adjacent segment disease in patients undergoing surgery for cervical radiculopathy. Future validation of the model should be performed before implementation in the clinical workflow.

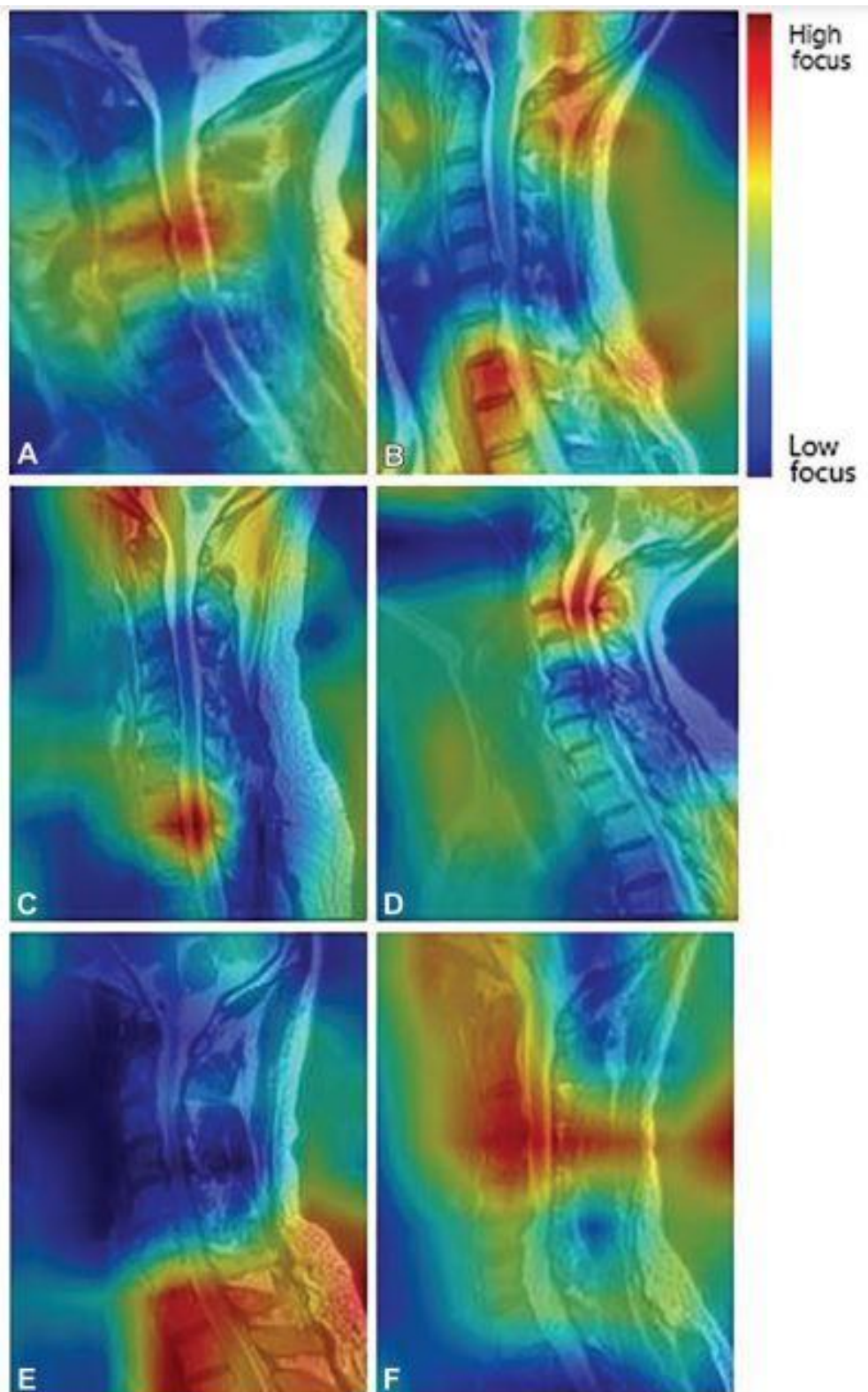


Figure 4: GradCAM saliency maps illustrate which region of the cervical MRI scan positively correlates with the output of the network. GradCAM-based saliency maps are provided for **(A, B)** two cervical MRI scans on the basis of the algorithm accurately predicted that patients would develop adjacent segment disease (ASD) and **(C-F)** four cervical MRI scans on the basis of which the algorithm accurately predicted that patients would not develop ASD. The color bar indicates importance of zones on images for the network in making a prediction. "High focus" zones were considered most important by the network, whereas "low focus" zones are of lesser importance to the network in its decision making. For prediction of patients who would develop symptoms, compression of neural structures appeared to be a main risk factor for the deep learning algorithm.