Deep Learning for Lumbar Spine MRI Reporting: A Welcome Tool for Radiologists

Daichi Hayashi, MD, PhD

Dr Daichi Hayashi is clinical associate professor of radiology at SUNY Stony Brook University and the Musculoskeletal Radiology Fellowship Program director. His research interests include semiquantitative imaging of osteoarthritis, imaging of rheumatologic and orthopedic disorders—including sports injuries and cancers—and AI application for musculoskeletal imaging, including spine, tumors, and fractures. Dr Hayashi has authored or co-authored more than 100 peer-reviewed publications and serves as editorial advisor (previously section editor) for *BMC Musculoskeletal Disorders*.

Lower back pain due to lumbar spinal stenosis is a major reason for seeking medical care in the United States (1). Lumbar spine MRI examinations are a major part of bone radiologists' routine workload in both community and university hospital settings. This is due in part to unnecessary referrals from primary care physicians (2). Interpreting lumbar spine MRI scans one after another can become repetitive and monotonous. Thus, an artificial intelligence (AI) diagnostic tool that reduces radiologists' burden is always welcome.

In their interesting study in this issue of *Radiology*, Hallinan and Zhu et al (3) developed a deep learning (DL) model for automated detection and classification of lumbar central canal, lateral recess, and neural foraminal stenosis. Their article is an excellent example of how AI research studies should be reported. The authors adhered to recommended guidelines for reporting AI studies. This included training, validation, and test sets for their internal data set, followed by evaluation of model performance using an external test set from a different institution outside the authors' own country (4). For the reference standard, an expert musculoskeletal radiologist performed readings for internal data sets, and another experienced subspecialist musculoskeletal radiologist performed readings for the external test set. Four additional readers (two musculoskeletal radiologists and two neuroradiologists) also performed intraobserver and interobserver variability assessments. Very briefly, a two-component DL model was developed. First, a convolutional neural network was trained to detect the region of interest, with a second convolutional neural network for classification. The authors presented clear examples of how visual grading of lumbar spinal stenosis was performed in figure E1 (online) of their article. Finally, their AI algorithm is publicly available. One limitation of the study is that there was no involvement of general radiologists without subspecialization for musculoskeletal radiology or neuroradiology.

Overall, 446 MRI lumbar spine images were included in the study (446 patients; mean age ± standard deviation, 52 years ± 19; 240 women), with 396 patients for training (80%) and validation (9%) and 50 patients (11%) for internal testing. For internal testing, DL model and radiologist central canal recall was greater than 99%, with reduced neural foramina recall for the DL model (85%) and first subspecialist radiologist (84%) compared with the second subspecialist radiologist (97%) ($P < .001$). Also on an internal test set, the DL model showed almost-perfect agreement (κ = 0.92 and 0.96, respectively; $P < .001$) for dichotomous classification (normal or mild vs moderate or severe) of lateral recess and central canal stenosis, similar to that of subspecialist radiologists (κ = 0.92–0.98, $P < .001$). The DL model also showed almost-perfect agreement for dichotomous classification of neural foraminal stenosis (κ = 0.89, $P < .001$), which was slightly reduced compared with that of subspecialist radiologists (κ = 0.94 and 0.95, $P < .001$). Finally, external testing of the DL model based on a data set from a different institution showed substantial levels of agreement between the DL model and a subspecialist radiologist for all regions of interest (κ = 0.95–0.96, $P < .001$).

Although there have already been several published reports on the use of DL models for the evaluation of lumbar spinal stenosis, specifically for grading central canal stenosis and neural foraminal stenosis (5–8), to my knowledge, this study is the first to report automated detection and classification of lateral recess stenosis specifically. This gives it some originality. Describing lumbar lateral recess stenosis is clinically relevant, as it can be specifically targeted for treatment with endoscopic surgical decompression (9). Moreover, compared with SpineNet (2017), a multitask architecture for automated classification of lumbar spinal conditions including grading of central canal stenosis, the DL model presented by Hallinan and Zhu et al (3) showed a seemingly better outcome, with κ values of 0.82 and 0.96 for ordinal and dichotomous classification of central canal stenosis, respectively ($P < .001$). SpineNet had a slightly lower performance, with agreement of 65.7% for ordinal grading of central canal stenosis and 94% (κ = 0.75) for dichotomous grading. Overall, Hallinan and Zhu et al demonstrated the technical feasibility of their DL model for use as a diagnostic tool to evaluate lumbar spinal stenosis in a semiautomated fashion with supervision of radiologists

From the Department of Radiology, Stony Brook University Renaissance School of Medicine, State University of New York, 101 Nicolls Rd, HSc Level 4, Room 120, Stony Brook, NY 11794-8460. Received March 17, 2021; revision requested and received March 29; accepted March 30. **Address correspondence to** the author (e-mail: *Daichi.Hayashi@stonybrookmedicine.edu*).

Conflicts of interest are listed at the end of this article.

See also the article by Hallinan and Zhu et al in this issue.

and with the potential to produce more consistent and less subjective reports.

A key issue with any AI-related productivity-enhancing tool is that as radiologists, we must effectively use them so that nonradiologist colleagues do not think of eliminating the need for our diagnostic expertise. Unfortunately, this is particularly true in the field of orthopedic imaging regarding bones, joints, and the spine. Based on anecdotal evidence in real-life clinical practice, especially at large academic institutions in the United States, some orthopedists can already perform image interpretation and preoperative measurements of joints and the spine at radiography, CT, and MRI themselves using software tools and three-dimensional printers they own. As radiologists, we must provide value-added service to the referrers so that they appreciate our diagnostic lumbar spine MRI report—high-quality, standardized, and appropriately aided by the validated AI-based tool—while, at the same time, optimizing our own efficiency and productivity. Now that technical feasibility has been demonstrated, the next step of the research will be to assess how this AI-aided lumbar spinal stenosis reporting can actually benefit radiologists. Benefits may include a reduction in reading time, more standardized reporting with less intrareader and interreader variabilities (especially for general radiologists who read spine MRI scans), and lower costs—requiring a cost-effective analysis with or without implementation of the AI tool. Finally, we must ask how it may benefit referring clinicians. For example, can it offer improved quality of patient care to treat lower back pain, thus resulting in fewer unnecessary referrals?

For instance, a very recent article published in a journal focusing on spine surgery reported the use of AI for grading lumbar spinal stenosis and compared its performance against the radiologist's stenosis grading (10). The study showed that radiologist and AI grading were equally predictive of a successful outcome of endoscopic decompression surgery. This study was a product of collaborative research between the orthopedic spine surgeon and the radiologist (as well as the vendor of the AI algorithm), providing correlation with surgical and clinically relevant outcomes.

I believe it is this sort of multidisciplinary collaborative AI research that will become increasingly important so that referring clinicians' needs and radiologists' needs align. However, this can be a potentially complicated issue, given the needs for different categories of physicians and allied health care professionals who refer their patients for lumbar MRI (ie, orthopedists, rheumatologists, family physicians, chiropractors, physiatrists). Another way to advance this research would be to add longitudinal analysis with multiple time points and correlate these findings with changes in patients' clinical status to predict outcomes and various therapeutic options.

Disclosures of Conflicts of Interest: Disclosed no relevant relationships.

References

1. Lurie J, Tomkins-Lane C. Management of lumbar spinal stenosis. BMJ 2016;352:h6234.
2. Nevedal AL, Lewis ET, Wu J, et al. Factors influencing primary care providers' unneeded lumbar spine MRI orders for acute, uncomplicated low-back pain: a qualitative study. J Gen Intern Med 2020;35(4):1044–1051.
3. Hallinan JTPD, Zhu L, Yang K, et al. Deep learning model for automated detection and classification of central canal, lateral recess, and neural foraminal stenosis at lumbar spine MRI. Radiology 2021;###(#):###–###.
4. Bluemke DA, Moy L, Bredella MA, et al. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers—from the Radiology Editorial Board. Radiology 2020;294(3):487–489.
5. Jamaludin A, Lootus M, Kadir T, et al. ISSLS Prize in Bioengineering Science 2017: automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. Eur Spine J 2017;26(5):1374–1383.
6. Ishimoto Y, Jamaludin A, Cooper C, et al. Could automated machine-learned MRI grading aid epidemiological studies of lumbar spinal stenosis? Validation within the Wakayama spine study. BMC Musculoskelet Disord 2020;21(1):158.
7. 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.
8. Lu J, Pedemonte S, Bizzo B, et al. Deep Spine: Automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning. Proceedings of the 3rd Machine Learning for Healthcare Conference, PMLR 2018; 85:403–419. http://proceedings.mlr.press/v85/lu18a.html.
9. Lee CH, Choi M, Ryu DS, et al. Efficacy and safety of full-endoscopic decompression via interlaminar approach for central or lateral recess spinal stenosis of the lumbar spine: a meta-analysis. Spine (Phila Pa 1976) 2018;43(24):1756–1764.
10. Lewandrowski KU, Muraleedharan N, Eddy SA, et al. Artificial intelligence comparison of the radiologist report with endoscopic predictors of successful transforaminal decompression for painful conditions of the lumber spine: application of deep learning algorithm interpretation of routine lumbar magnetic resonance imaging scan. Int J Spine Surg 2020;14(s3):S75–S85.