Improved Productivity using Deep Learning assisted Reporting for MRI Lumbar Spine

Naresh Kumar; Desmond Shi Wei Lim; Andrew Makmur; Jonathan Jiong Hao Tan; James Thomas Patrick Decourcy Hallinan

Introduction Lumbar spine MRI studies are widely used for back pain assessment. Interpretation involves grading lumbar spinal stenosis, which is repetitive and time consuming. Deep learning (DL) could provide faster and more consistent interpretation. The aim of the study is to assess the speed and interobserver agreement of radiologists for reporting lumbar spinal stenosis with and without DL assistance.

Materials and Methods In this retrospective study, a DL model designed to assist radiologists in the interpretation of spinal canal, lateral recess, and neural foraminal stenoses on lumbar spine MRI scans was utilized. Randomly selected lumbar spine MRI studies obtained in patients with back pain who were 18 years and older over a 3-year period, from September 2015 to September 2018, were included in an internal test data set. Studies with instrumentation and scoliosis were excluded. Eight radiologists, each with 2–13 years of experience in spine MRI interpretation, reviewed studies with and without DL model assistance with a 1-month washout period. Time to diagnosis (seconds) and interobserver agreement (using Gwet k) were assessed for stenosis grading for each radiologist with and without the DL model and compared with an external musculoskeletal radiologist (with 32 years of experience) as the reference standard.

Results Overall, 444 images in 25 patients (mean age, 51 years 6 20 [SD]; 14 women) were evaluated in a test data set. DLassisted radiologists had a reduced interpretation time per spine MRI study from a mean of 124–274 seconds (SD, 25–88 seconds) to 47–71 seconds (SD, 24–29 seconds) (P,.001). DL-assisted radiologists had either superior or equivalent interobserver agreement for all stenosis gradings compared with unassisted radiologists. DL-assisted general and in-training radiologists improved their interobserver agreement for four-class neural foraminal stenosis, with k values of 0.71 and 0.70 (with DL) versus 0.39 and 0.39 (without DL), respectively (both P,.001).

Conclusion Radiologists who were assisted by deep learning for interpretation of lumbar spinal stenosis on MRI scans showed a marked reduction in reporting time and superior or equivalent interobserver agreement for all stenosis gradings compared with radiologists who were unassisted by deep learning.