Feasibility of Deep Learning Algorithms for Reporting in Routine Spine Magnetic Resonance Imaging

Kai-Uwe Lewandrowski, Narendran Muraleedharan, Steven Allen Eddy, Vikram Sobti, Brian D. Reece, Jorge Felipe ramírez León and Sandeep Shah

*Int J Spine Surg* 2020, 14 (s3) S86-S97
doi: [https://doi.org/10.14444/7131](https://doi.org/10.14444/7131)
http://ijssurgery.com/content/14/s3/S86

This information is current as of September 14, 2023.

**Email Alerts**  Receive free email-alerts when new articles cite this article. Sign up at: [http://ijssurgery.com/alerts](http://ijssurgery.com/alerts)
Feasibility of Deep Learning Algorithms for Reporting in Routine Spine Magnetic Resonance Imaging

KAI-UWE LEWANDROWSKI, MD,1 NARENDRA MURALEEDHARAN, BASME,2 STEVEN ALLEN EDDY, MD,3 VIKRAM SOBTI, MD, MBA,4 BRIAN D. REECE, MD,5 JORGE FELIPE RAMIREZ LEÓN, MD,6 SANDEEP SHAH, MSEE, MBA7

1Staff Orthopaedic Spine Surgeon Center for Advanced Spine Care of Southern Arizona and Surgical Institute of Tucson, Tucson, Arizona, 2Aptus Engineering, Inc, Scottsdale, Arizona, and Multus Medical, LLC, Phoenix, Arizona, 3Multus Medical, LLC, Phoenix, Arizona, 4Innovative Radiology, PC, River Forest, Illinois, 5The Spine and Orthopedic Academic Research Institute, Lewisville, Texas, 6Fundación Universitaria Sanitas, Bogotá, Colombia, Research Team, Centro de Columna, Bogotá, Colombia, Centro de Cirugía de Mínima Invasión, CECIMIN—Clínica Reina Sofía, Bogotá, Colombia, 7Multus Medical, LLC, Phoenix, Arizona

ABSTRACT

Background: Artificial intelligence is gaining traction in automated medical imaging analysis. Development of more accurate magnetic resonance imaging (MRI) predictors of successful clinical outcomes is necessary to better define indications for surgery, improve clinical outcomes with targeted minimally invasive and endoscopic procedures, and realize cost savings by avoiding more invasive spine care.

Objective: To demonstrate the ability for deep learning neural network models to identify features in MRI DICOM datasets that represent varying intensities or severities of common spinal pathologies and injuries and to demonstrate the feasibility of generating automated verbal MRI reports comparable to those produced by reading radiologists.

Methods: A 3-dimensional (3D) anatomical model of the lumbar spine was fitted to each of the patient’s MRIs by a team of technicians. MRI T1, T2, sagittal, axial, and transverse reconstruction image series were used to train segmentation models by the intersection of the 3D model through these image sequences. Class definitions were extracted from the radiologist report for the central canal: (0) no disc bulge/protrusion/canal stenosis, (1) disc bulge without canal stenosis, (2) disc bulge resulting in canal stenosis, and (3) disc herniation/protrusion/extrusion resulting in canal stenosis. Both the left and right neural foramina were assessed with either (0) neural foraminal stenosis absent, or (1) neural foramina stenosis present. Reporting criteria for the pathologies at each disc level and, when available, the grading of severity were extracted, and a natural language processing model was used to generate a verbal and written report. These data were then used to train a set of very deep convolutional neural network models, optimizing for minimal binary cross-entropy for each classification.

Results: The initial prediction validation of the implemented deep learning algorithm was done on 20% of the dataset, which was not used for artificial intelligence training. Of the 17,800 total disc locations for which MRI images and radiology reports were available, 14,720 were used to train the model, and 3560 were used to validate against. The convergence of validation accuracy achieved with the deep learning algorithm for the foraminal stenosis detector was 81% (sensitivity = 72.4%, specificity = 83.1%) after 25 complete iterations through the entire training dataset (epoch). The accuracy was 86.2% (sensitivity = 91.1%, specificity = 82.5%) for the central stenosis detector and 85.2% (sensitivity = 81.8%, specificity = 87.4%) for the disc herniation detector.

Conclusions: Deep learning algorithms may be used for routine reporting in spine MRI. There was a minimal disparity among accuracy, sensitivity, and specificity, indicating that the data were not overfitted to the training set. We concluded that variability in the training data tends to reduce overfitting and overtraining as the deep neural network models learn to focus on the common pathologies. Future studies should demonstrate the accuracy of deep neural network models and the predictive value of favorable clinical outcomes with intervention and surgery.

Level of Evidence: 3.

Clinical Relevance: Feasibility, clinical teaching, and evaluation study.

Special Issue
INTRODUCTION

In the last 5 years, the development of artificial intelligence (AI) has leapfrogged forward across several industries, ranging from quality control in various manufacturing areas to improved automation of production processes and enhanced diagnostics in medical applications. Examples include voice and facial recognition, landmark localization, autonomic driving, and a wide array of medical imaging modalities. From a clinical standpoint, demonstrating the application of AI and its deep neural network learning algorithms is highly relevant and timely due to the ongoing debate on the necessity of advanced medical and surgical intervention while costs are rising. Establishing new diagnostic imaging criteria with higher sensitivity, specificity, accuracy, and positive predictive value for favorable clinical outcomes with proposed interventions is considered by many to be key to the development of next-generation patient-centered and higher-quality health care service purchasing strategies in general and for the endoscopic spinal procedure in particular. Using AI in closing the well-recognized diagnostic gap in routine lumbar magnetic resonance imaging (MRI) scanning is one such example. Extracting higher-quality diagnostic value from the MRI scan is critical for successful endoscopic spinal surgery because it relies heavily on correctly identifying the pain generator responsible for the patient’s symptoms.

The deployment of severity grading of traditional subjective visual analysis of advanced cross-sectional MRI imaging of the spine by the radiologist not only leaves room for errors but has been shown to lead to the omission of appropriate spine care, which ultimately contributes to overuse in other areas. Repetitive rounds of less-effective physical therapy, interventional, and medical care often provide only short-term pain relief and rarely definitively address the patients’ disability because the pain generator stemming from the underlying anatomical abnormality of the spine has not been fixed. As a result, the patients’ disabilities continue and play out in their private and professional lives with decreased functional capacity due to poorly controlled pain, lack of strength, coordination, or insufficient endurance. The cumulative societal burden due to missed work, ongoing use of medical services, and narcotic dependence is on the radar of every stakeholder in government and the medical insurance and service industry.

Out-of-control runaway costs will likely prompt more rationing of medical services in general and spine care in particular unless clinical evidence is presented on how to realize cost savings with technology advancements that produce more impactful, targeted, and durable solutions for patients that ultimately have the potential to flex down the spending curve during a time in which the demand for such spine care is expected to increase significantly with generations of aging baby boomers coming onto the Medicare rolls. More accurate prognosticators of favorable clinical outcomes with spine intervention and surgery are critical to materializing such cost savings. The need for such cost savings will likely also translate into higher use of more targeted minimally invasive and endoscopic spine decompression and reconstructive surgeries. To provide consistent clinical benefit, identifying the primary pain generator is of utmost importance, and applying routine MRI scanning with higher-level accuracy will probably become more relevant. Therefore, we investigated the feasibility of using deep learning algorithms for routine reporting in spine MRI with the ultimate objective of improving its accuracy and predictive value.

MATERIALS AND METHODS

The premise of this research and development is based on the ability of deep learning neural network models to identify features in MRI data that represent varying intensities or severities of pathologies or injuries in patients.

Patients and Training Data

The training dataset used to develop the neural network models includes lumbar MRI scans from 3560 patients, constituting a total of 17 800 levels. The training data were obtained from 168 different MRI imaging center locations around the United States. The dataset included the disc levels L1-L2, L2-L3, L3-L4, L4-L5, and L5-S1 for each patient. The average age of the 3560 patients was 41.2 years, with a standard deviation of 14.9 years. There were 46% male and 51.9% female patients. The remaining 2.1% chose not to identify their gender. The participating MRI imaging centers provided radiology reports prepared and approved by board-certified radiologists. Each radiologist was required to present a reading for the presence or absence of
annular bulging (circumferential, paracentral, posterior), disc herniation (extrusion, protrusion, sequestration, fragmentation), central canal stenosis (compromise of the thecal sac with presence or absence of ventral epidural fat), and foraminal stenosis (compromise of the left, right, or both neural foramina and nerve roots) for each intervertebral level. For algorithm development and validation, various splits of the dataset were used to ensure that the model was tested against cases that it had not seen during training. Train-test percentages of 70%–30%, 80%–20%, and 90%–10% were used for various models.

Preparation of Training Data

It is essential to extract numerical training data from the MRI imaging data. First, it is crucial to identify the location of each vertebra and disc in the patient’s lumbar spine in order. Lu et al proposed the use of automated segmentation algorithms to automate this process. In their approach, quadrilaterals were drawn to encompass each vertebra visible on the sagittal image. Segmented regions were used to fit a spine curve and localize the centers of each disc, and a series of sagittal and axial slices from the area was used for training and prediction. To extract the disc regions more accurately and to extract the spinal cord profile, a three-dimensional (3D) anatomical model of the lumbar spine was fitted to each of the patient’s MRIs by a team of technicians. The 3D model was fitted such that the boundaries of the vertebrae, discs, and cord line up with the respective boundaries in the MRI images. Sagittal and axial slices were used as reference (Figure 1a and b). The segmentation results in a 3D anatomical model custom to each patient’s lumbar spine (Figure 1c). This allows the use of other MRI image series (for example, T1, T2, sagittal, axial, and transverse reconstructions) to be used to train segmentation models by the intersection of the 3D model through these image sequences. For each disc location, the following classes were extracted from the radiologist report for the central canal: (0) no disc bulge, protrusion, or canal stenosis, (1) disc bulge without canal stenosis, (2) disc bulge resulting in canal stenosis, and (3) disc herniation, protrusion, or extrusion resulting in canal stenosis. One of the following classes was also extracted for each of the left and right neural foramina: (0) neural foraminal stenosis absent, (1) neural foraminal stenosis present. An example is shown in Table 1, in which the L3-L4 location in the side-by-side comparison the radiologist read was converted to class (3) for the central canal, (0) for the left neural foramina, and (1) for the right neural foraminal—which was matched by the algorithm model.

Second, 2 approaches were taken to extract manual radiologist reporting labels for the pathologies at each disc level and, when available, the grading of severity. Similar to Lu’s DeepSPINE model, natural language processing (NLP) was used to extract disc-level locations and pathologies at each location. The NLP model was trained with 5000 manually labeled disc levels. One of the following options was marked for the central canal on the basis of the radiologist’s report: no signs of abnormality, disc bulging without compromise of the thecal sac, disc bulging compressing thecal sac (central canal stenosis), or disc herniation compressing thecal sac (central canal stenosis). One of the following options was labeled for the neural foramina as well: no signs of abnormality, left foraminal stenosis, right foraminal stenosis, or bilateral foraminal stenosis. For example, a report finding that states “L4-L5: Broad-based posterior disc herniation, best seen on sagittal T2 image #8/13 indenting thecal sac and causing mild narrowing of bilateral neural foramina” is labeled as follows: disc herniation compressing thecal sac (central canal stenosis) and bilateral foraminal stenosis.

The NLP algorithm was run on all 17 800 disc levels with radiology reports provided to generate labeled training data for the pathology identification
Deep learning algorithm. Due to known imperfections and accuracy of NLP algorithms, a semi-supervised training process was adopted. Semisupervised training algorithms have been used to improve the accuracy of models when it is unfeasible to prepare supervised training data due to a large sample size or the complexity and labor intensiveness of manually labeling data.68–70 The training process included unsupervised training data generated by the NLP algorithm for the entire dataset along with the 5000 manually labeled and curated labels prepared originally to train the NLP algorithm. Furthermore, due to the tendency for lumbar disc pathologies to be more common in the lower lumbar motion segments, class imbalance was handled by weighting model classes and mixing of disc locations in the training data. Figure 2 depicts the distribution of identified central canal stenosis in the training data at each disc location. Suppressed consistency loss was used as a regularization method to increase robustness towards class imbalance at different levels.71

Models and Architecture

The proposed algorithm operates in three high-level stages. First, each sagittal and axial slice is segmented using a semantic segmentation network using the manually segmented 3D model. The implemented method uses concepts from the one hundred layers tiramisu network proposed by Jégou et al.72 Segmented outputs similar to those in Figure 3a and b are generated for each sagittal and axial slice in the MRI images. The segmented regions are used to extract the disc centers and orientation (using principal component analysis) for each disc location from L5-S1 counting upward until L1-L2. As proposed in the DeepSPINE model,67 stacks of cropped sagittal and axial slices are extracted from MRI images intersecting the disc. The segmented spinal cord is also used to measure the canal midline anterior-posterior (AP) diameter—an objective and measurable metric. The second stage in the pipeline uses 2 separate visual geometry group convolutional networks73 trained with semisupervised methods on cropped sagittal and axial MRI image stacks and radiological findings labeled using NLP and manually. The first network is used to detect and grade

<table>
<thead>
<tr>
<th>Level</th>
<th>Model Output</th>
<th>Radiologist Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1-L2</td>
<td>No disc herniation, neurocompression, or neuroforaminal stenosis is seen at this level.</td>
<td>There is a 1-mm midline disc bulge. There is no central canal narrowing.</td>
</tr>
<tr>
<td>L2-L3</td>
<td>There is moderate posterior disc herniation of the intervertebral disc impinging on the thecal sac. The spinal canal midline AP diameter is 11.8 mm. This is best seen on T2_FSE_TRS image slice No. 4. There is narrowing of the neural foramina bilaterally.</td>
<td>There is a moderate degree of central canal narrowing. There is a posterior annular tear. There is a 3-mm circumferential disc protrusion resulting in abutment of the descending L3 nerve roots bilaterally as well as abutment of the exiting right and left L2 nerve roots.</td>
</tr>
<tr>
<td>L3-L4</td>
<td>There is posterior disc herniation of the intervertebral disc impinging on the thecal sac. The spinal canal midline AP diameter is 12.7 mm. This is best seen on T2_FSE_TRS image slice No. 9. There is narrowing of the right neural foramina with abutment of the exiting right nerve root.</td>
<td>There is a moderate degree of central canal narrowing. There is a 2-mm right foraminally disc protrusion with abutment of the exiting right L3 nerve root and mild narrowing of the right neural foramen.</td>
</tr>
<tr>
<td>L4-L5</td>
<td>There is posterior annular bulging seen compressing the thecal sac, resulting in central canal stenosis. The spinal canal midline AP diameter is 15.9 mm. This is best seen on T2_FSE_TRS image slice No. 16. There is narrowing of the neural foramina bilaterally.</td>
<td>There is mild central canal narrowing. There is a 2-mm right foraminal disc protrusion resulting in mild abutment of the exiting right L4 nerve root and mild narrowing of the right neural foramen.</td>
</tr>
<tr>
<td>L5-S1</td>
<td>There is posterior disc herniation of the intervertebral disc impinging on the thecal sac. The spinal canal midline AP diameter is 19.2 mm. This is best seen on T2_FSE_TRS image slice No. 23. There is narrowing of the neural foramina bilaterally.</td>
<td>There is a 1-mm midline disc bulge with mild effacement of the anterior thecal sac. There is no central canal narrowing. The height of the right neural foramen is 19 mm and the height of the left neural foramen is 17 mm.</td>
</tr>
</tbody>
</table>

Abbreviations: AP, anterior-posterior.
We compared the reports generated by these AI segmentation algorithms to select representative image sections and had the radiologist on our team review the images to confirm the AI read and to compare it with the radiologist report provided by the participating MRI centers. Figure 5a and b is an example diagnostic assessment using the algorithm that is known to have no disc bulging, no central canal stenosis, and no foraminal narrowing. The algorithm reported no canal stenosis and no neural foraminal stenosis, thus matching the known radiologist’s reporting. The algorithm-generated report summary was “L1-L2: No disc herniation, neuro-compression, or neuroforaminal stenosis is seen at this level.” Figure 5c and d is an example diagnostic assessment using the algorithm in which the training radiologist labeled the disc to have a posterior disc protrusion abutting the thecal sac and compromising the neural foramina bilaterally. In comparison, the deep learning algorithm reported “There is posterior herniation of the intervertebral disc impinging on the thecal sac, best seen on T2_FSE_TRS (FSE=fast spin echo) series image #4. The spinal canal midline AP diameter is 10 mm. There is narrowing of the neural foramina bilaterally.” As demonstrated in these 2 examples, the algorithm also indicated the image slice in which the pathology was best demonstrated and reported the measured spinal canal diameter at the affected level.

We used the initial results of the prediction and validation of the implemented deep learning algorithm on the 20% of the dataset that was not used for the AI training. Of the 17,800 total disc locations for which MRI images and radiology reports were available, 14,720 were used to train the model, and

---

**Figure 3.** Representative (a) sagittal and (b) axial segmentation predictions of lumbar spine matching patient’s MRI images in Figure 1. MRI, magnetic resonance imaging.

**Figure 4.** High-level architectural diagram of the implemented deep learning algorithm used to generate automated MRI reports. MRI, magnetic resonance imaging.
3560 were used to validate against. Separate models were developed and trained for the identification of each diagnosis and class: generalized disc bulging, canal stenosis, disc herniation, and foraminal stenosis. The bilateral neuroforamina were assessed independently for stenosis affecting the left and right nerve roots. Therefore, twice as many data points were available to train the foraminal stenosis detector model. The loss functions used to minimize in each model was binary cross-entropy, also known as the log loss function.

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^{N} y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)).$$

Each model was trained for 25 epochs, during which the convergence of binary validation accuracy was observed. Figure 6a is a plot depicting the convergence of validation accuracy achieved with the deep learning algorithm to approximately 81% for the foraminal stenosis detector. The optimization, however, was for the above-mentioned binary cross-entropy loss function. Figure 6b is a plot depicting the convergence of the binary cross-entropy loss across the 25 epochs (an epoch is 1 complete iteration through the entire training dataset) on the foraminal stenosis detector with increasing accuracy. At the end of each epoch, the binary cross-entropy loss was calculated and stochastic gradient descent optimization was used to compute changes to deep neural network model weights to minimize the loss. We observed from the plots that the binary training accuracy continued to increase, whereas the validation accuracy converged to roughly 81% for the foraminal stenosis detector (sensitivity = 72.4.4%, specificity = 83.1%). The convergence of the validation accuracy was apparent after just 5 training epochs. Any gain in training accuracy observed past the validation accuracy convergence was due to overfitting to specific radiology reads and methods; however, these did not affect the overall validation accuracy. Spatial dropouts and other techniques were implemented to minimize overfitting to the specific training dataset. Binary accuracy, test sensitivity, and specificity were recorded for each model on the basis of the results from validating against 20% of the complete dataset; these are summarized in Table 2. The accuracy for the central stenosis detector was 86.2% (sensitivity = 91.1%, specificity = 82.5%) and for the disc herniation detector, 85.2% (sensitivity = 81.8%, specificity = 87.4%).
Table 2. Validation results showing accuracy, sensitivity, and specificity for each segmentation model.

<table>
<thead>
<tr>
<th>Detector Model</th>
<th>Accuracy, %</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulging</td>
<td>84.9</td>
<td>89.4</td>
<td>81.2</td>
</tr>
<tr>
<td>Canal stenosis</td>
<td>86.2</td>
<td>91.1</td>
<td>82.5</td>
</tr>
<tr>
<td>Disc herniation</td>
<td>85.2</td>
<td>81.8</td>
<td>87.4</td>
</tr>
<tr>
<td>Foraminal stenosis</td>
<td>81.0</td>
<td>72.4</td>
<td>83.1</td>
</tr>
</tbody>
</table>

DISCUSSION

The need for more accurate prognosticators of favorable clinical outcomes with lumbar spinal surgery prompted us to investigate the feasibility of using deep learning algorithms for routine reporting in spine MRI. The need for accurate prognosticators of favorable clinical outcomes has been recognized by the North American Spine Society (NASS), which discussed this need in several of its consensus treatment guidelines for common spine problems. The organization provided an in-depth review of the existing literature and graded the clinical evidence by comparing preoperative MRI findings with intraoperative observations directly visualized by the surgeon in open spine surgeries. Thus, vetted and validated sensitivity and specificity numbers ranging between the 60th and 70th percentiles and serving as industry benchmarks for most common spine problems were established. One consequential example of the poor diagnostic value of routine lumbar MRI scans is missed injury to the posterior longitudinal ligament complex in thoracolumbar fractures in patients with acute injury. The integrity of this vital, stabilizing ligamentous complex typically triggers nonoperative care with bracing, whereas injury triggers surgical fixation with spinal fusion, an algorithm that calls for 2 vastly different treatments, when used erroneously, has tremendous unintended downstream consequences that nearly always translate into an ongoing need for care and higher cost. Another such example is herniated disc. A high percentage of asymptomatic, healthy volunteers were found to have disc herniations at multiple lumbar levels, calling into question the positive predictive value of the lumbar MRI scan in patients with painful acute injuries or degenerative abnormalities. In a nutshell, the lumbar MRI scan delivers little information with respect to the leading pain generator. The high false negative rate among patients with sciatica-type back and leg pain is on the order of 30%, and yet the radiologists lacking relevant clinical context of the spine care at the time it is delivered—willingly and knowingly or not—find themselves in the middle of the medical necessity controversy when it comes to determining the need for treatment. What is evident is that there is a tremendous need to improve the accuracy of the interpretation of the MRI scan, particularly when it comes to the application of small, targeted, minimally invasive and endoscopic surgeries that aim to treat only the most relevant pain generator. Higher preoperative diagnostic accuracy is at the center of making these less burdensome and more cost-effective advanced, highly targeted endoscopic outpatient surgical procedures work. Currently, the accuracy of the lumbar MRI scan report in predicting acceptable levels of clinical success with spinal decompression surgery can be raised only with the addition of other ancillary tests, such as a lidocaine-containing transforaminal epidural steroid injection.

In an attempt to improve the diagnostic value of the lumbar MRI scan, a uniform nomenclature of a herniated disc and spinal stenosis was proposed and published. Clear definitions of bulging or herniated disc or disc protrusions were given to avoid the interchangeable and indiscriminate use of these terms without attention to detail or their clinical relevance. Several radiographic classification systems of lumbar spinal stenosis in the central canal, lateral recess, or neuroforamina have been published. This lack of detail in the conventional MRI reporting provided by the radiologist and how it relates to the relevant clinical context motivated us to go beyond traditional subjective visual image interpretation and prompted us to look into the deployment of modern AI to reduce the waste and improve patient outcomes in modern spine care. Therefore, we investigated the feasibility of using deep neural network self-learning...
algorithms to provide written reports that not only could be increasingly more accurate and consistent with traditional verbal reads produced by a radiologist but would improve upon the current industry standards.

The results of this feasibility study of 3560 lumbar MRI scans and 17,800 levels shows that the model report generated by the network models (the RadBot) was capable of producing a verbally spelled-out report by a specific lumbar spinal level comparable in detail as to what is typically seen by the radiologist in terms of detail and scope. From training using 14,240 disc locations across 25 epochs and validating against 3560 disc locations, we observed that the accuracy, sensitivity, and specificity metrics are consistently higher than 85% for the central canal compression detectors. Despite training against double the number of neural foramina (left and right nerve roots), the ability of the model to accurately match foraminal stenosis detection to those from radiologist reads is less than that of central canal compression. This can be due to a few possible reasons—due to the more complex 3D volumetric shape of the neuroforamina and of the nerve roots as compared with the posterior section of the disc and central canal, which are much larger in comparison. Variability in the training reads resulting in differences with the stenosis indicators radiologists use to describe the diagnosis of neural foraminal stenosis may also be another limitation of the RadBot in its current modeling algorithm that we may wish to overcome with additional training of the RadBot if a clinical need arises. The overall accuracy of 81% with the MRI reporting may have several explanations. Still, the most obvious one is that the deep learning network model’s accuracy may not exceed 81% with the current manually segmented 3D models. Redefining this manual segmentation to common clinically relevant painful entities of the lumbar spine may improve the accuracy. Another problem that may reside in the underlying DICOM datasets, often obtained on 1.5-T scanners, which are too noisy. Moreover, an 81% accuracy for across-the-board RadBot reading of lumbar spine MRI scans obtained in our feasibility study is approximately 15% higher than the published interobserver and intraobserver reliability rates obtained on routine reports provided by radiologists on the same scan.91–99 Despite these limitations, the RadBot was able to give the printed MRI report in approximately 8 to 10 minutes, which in today’s health care cost-savings context may save time and prevent overuse owing to improved reporting standards. Future studies will have to demonstrate the reliability of the RadBot readings with k analysis of agreement between the RadBot and the MRI reports provided by a radiologist.

**CONCLUSIONS**

We demonstrated the feasibility of using deep learning algorithms for routine reporting in spine MRI. We found the minimal disparity among accuracy, sensitivity, and specificity, which indicated first that the data were not being overfitted to the training set, and second that the frequency of false negatives and false positives were both consistent and low compared with the true positives and true negatives. In addition, variability in the training data tended to reduce overfitting and overtraining as the deep neural network models learned to focus on the common indicators and ignore differences. In future studies, we will focus on providing RadBot reliability data in correlation with painful entities in patients with spinal injuries and degenerative conditions of the lumbar spine, with the ultimate objective of improving its accuracy and predictive value of favorable clinical outcomes with intervention.

**REFERENCES**

8. Borenstein J, Herkert JR, Miller KW. Self-driving cars
40. Lewandrowski KU, de Carvalho PST, Calderaro AL, et al. Outcomes with transfosseal endoscopic versus percutaneous laser decompression for contained lumbar herniated


Disclosure and COI: The first author has no direct (employment, stock ownership, grants, patents) or indirect conflicts of interest (honoria, consultancies to sponsoring organizations, mutual fund ownership, paid expert testimony). He is not currently affiliated with or under any consulting agreement with any MRI vendor that the clinical research data conclusion could directly enrich. The remaining four authors have received no funding for this study and report no conflicts of interest. This manuscript is not meant for or intended to
push any agenda other than reporting the research data related on automated recognition of common painful spine pathologies by deep neural network learning. The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

**Corresponding Author:** Kai-Uwe Lewandrowski, MD, Staff Orthopaedic Spine Surgeon Center for Advanced Spine Care of Southern Arizona and Surgical Institute of Tucson, AZ 85712. Phone: (520) 204-1495; Fax: (623) 218-1215; Email: busniess@tucsonspine.com.

Published 9 December 2020
This manuscript is generously published free of charge by ISASS, the International Society for the Advancement of Spine Surgery. Copyright © 2020 ISASS. To see more or order reprints or permissions, see http://ijssurgery.com.