



A neural network model for detection and classification of lumbar spinal stenosis on MRI

Vladislav Tumko¹ · Jack Kim¹ · Natalia Uspenskaia¹ · Shaun Honig¹ · Frederik Abel² · Darren R. Lebl² · Irene Hotalen¹ · Serhii Kolisnyk³ · Mikhail Kochnev¹ · Andrej Rusakov¹ · Raphaël Mourad⁴

Received: 31 May 2023 / Revised: 30 October 2023 / Accepted: 4 December 2023 / Published online: 27 December 2023
© The Author(s) 2023

Abstract

Objectives To develop a three-stage convolutional neural network (CNN) approach to segment anatomical structures, classify the presence of lumbar spinal stenosis (LSS) for all 3 stenosis types: central, lateral recess and foraminal and assess its severity on spine MRI and to demonstrate its efficacy as an accurate and consistent diagnostic tool.

Methods The three-stage model was trained on 1635 annotated lumbar spine MRI studies consisting of T2-weighted sagittal and axial planes at each vertebral level. Accuracy of the model was evaluated on an external validation set of 150 MRI studies graded on a scale of absent, mild, moderate or severe by a panel of 7 radiologists. The reference standard for all types was determined by majority voting and in case of disagreement, adjudicated by an external radiologist. The radiologists' diagnoses were then compared to the diagnoses of the model.

Results The model showed comparable performance to the radiologist average both in terms of the determination of presence/absence of LSS as well as severity classification, for all 3 stenosis types. In the case of central canal stenosis, the sensitivity, specificity and AUROC of the CNN were (0.971, 0.864, 0.963) for binary (presence/absence) classification compared to the radiologist average of (0.786, 0.899, 0.842). For lateral recess stenosis, the sensitivity, specificity and AUROC of the CNN were (0.853, 0.787, 0.907) compared to the radiologist average of (0.713, 0.898, 805). For foraminal stenosis, the sensitivity, specificity and AUROC of the CNN were (0.942, 0.844, 0.950) compared to the radiologist average of (0.879, 0.877, 0.878). Multi-class severity classifications showed similarly comparable statistics.

Conclusions The CNN showed comparable performance to radiologist subspecialists for the detection and classification of LSS. The integration of neural network models in the detection of LSS could bring higher accuracy, efficiency, consistency, and post-hoc interpretability in diagnostic practices.

Keywords Lumbar spinal stenosis · MRI · Artificial intelligence · Neural network · Machine learning

Introduction

Lumbar spinal stenosis (LSS) is one of the most commonly diagnosed spinal pathologies in the United States with high prevalence among older age groups [1, 2]. LSS is defined as the narrowing of the spinal canal in the lower part of the back. The clinical symptoms of LSS can include low back, lower extremity and gluteal pain, numbness, a substantial loss of mobility, and severe limitations in one's ability to perform daily activities [3]. Among degenerative spine disease, LSS represents one of most common reason for spinal surgery in patients above the age of 65 years [4], hence, accurate and reliable diagnosis and subsequent treatment has important economic and social consequences [5]. LSS is characterized by a narrowing of the central canal, lateral/

✉ Jack Kim
jack.kim@remedylogic.com

✉ Raphaël Mourad
raphael.mourad@univ-tlse3.fr

¹ Remedy Logic, 1177 Avenue of the Americas, 5th Floor, New York, NY 10036, USA

² Hospital for Special Surgery, 535 East 70th Street, New York, NY 10021, USA

³ Vinnitsa National Medical University, Vinnitsya, Ukraine

⁴ University of Toulouse, 118 Rte de Narbonne, 31062 Toulouse, France

subarticular recesses, or intervertebral foramen, which in turn causes compression of the associated neural structures [6].

While LSS is defined as a clinical syndrome and diagnosed as such, imaging is frequently used as a confirmatory tool prior to determining treatment. Magnetic resonance imaging (MRI) is the mainstay modality for detailed anatomical assessment of the spine with excellent soft tissue contrast and used for confirmation of clinical symptom-based diagnoses and determining the optimal treatment course. MRI is essential for evaluation of LSS and to confirm clinical findings, largely supplanting CT myelogram studies due to its superior soft tissue contrast. A number of studies have attempted to identify core quantitative radiological criteria for the diagnosis of LSS [7–9]. However, lumbar spine MRI interpretation is time-intensive and depends on the individual radiologist or surgeons' expertise and experience, therefore supporting objective and standardized methods of diagnosing and decision-making are desired.

Machine learning (ML) models, including deep convolutional neural networks (CNN), have already been successfully applied for evaluation of LSS and other degenerative changes with high accuracy in various approaches [10, 11]. However, most CNN algorithms rely on one-component models for binary classification (present/absent) of LSS. One recent work applied a two-component CNN to detect stenosis targeting the central canal, lateral recesses, and neural foramina with subsequent grading of the stenosis [12].

Here, the purpose of this study was to develop a three-stage convolutional neural network (CNN) approach to segment anatomical structures, classify the presence of lumbar spinal stenosis (LSS) and assess its severity on spine MRI on axial and sagittal MR images. The classification covers the detection of all three stenosis types—central canal, foraminal, and lateral recess. The performance of the model has been compared to a panel of radiologist subspecialists to test its reliability and accuracy.

Methods

Data set and annotation

External institutional review board approval was obtained to retrospectively review anonymized imaging data. The initial data set consisted of 1635 MRI studies of adult subjects referred for lumbar spine MRI for low back pain. The data set consisted of 45.7% of males (54.3% of females), with age ranging from 18 to 85 years. Each MRI study corresponded to a patient (1635 MRI studies = 1635 patients). Patients with implants or instrumentation, severe scoliosis, and poor image quality were excluded. Each MRI study was acquired using a standard lumbar spine protocol,

including T2-weighted axial and sagittal pulse sequences with balanced labels. First, T2-weighted axial sequence was extracted from each lumbar study. All slices from the lumbar disc levels (L1/L2, L2/L3, L3/L4, L4/L5, L5/S1) were selected from the T2-weighted axial sequence and labeled per slice. On average, around 10 to 15 axial slices were obtained from each study, with a more precise count of approximately 13.3 for this specific dataset (21,702 images in total).

For axial images, muscle tissue, the discs, spinal canal, thecal sac, neural foramina, nerves, nerve roots, lateral recess, facet joints, spinous process, articular process, ligamentum flavum, disc bulging or herniation, arteries, veins, and kidneys were labeled. On sagittal plane images, the discs, vertebral body, spinal canal, spinal cord with nerve roots, and spinous processes were labeled for segmentation. The segmentation labelling was performed by administrators.

In addition to segmentation of the key anatomical structures, the studies were also labeled by musculoskeletal-trained radiologist subspecialists on a scale of 0 (absent), 1 (mild), 2 (moderate), 3 (severe) for LSS to establish the reference standard [13]. The reference standard was determined by majority voting rule and in case of disagreement, adjudicated by a further radiologist. At first, during training, the segmentation model detected the facet joints and spinal canal. Then images were resized and augmented (horizontal flips and rotations). Of the 1635 studies, 1390 were used for CNN weight training, and 245 as a validation set for hyperparameter tuning. The dataset was randomly divided into a training set for training the CNN parameters (1635 studies) and a validation set for hyperparameter tuning (245 studies). This was done while maintaining the same distribution in each subset, split ratio and the rule that images from one study can't be in different sets at the same time. Thus, data splitting was carried out based on studies rather than individual images, such as images from the same MRI study could go to different subsets (to avoid for instance two images from the same MRI study going to both training and testing).

Additionally, an external data set of 150 studies were reserved for final model accuracy assessment. The inclusion/exclusion criteria were identical to the those of the training data set. The external validation set were graded on a scale of absent, mild, moderate or severe by a panel of 7 radiologist subspecialists. The radiologist interpretations were then compared to the interpretation of the model.

The number of images for each stenosis (central, lateral recess, foraminal) for each severity class (absence, mild, moderate, severe) and for each data subset (train, validation, test) is shown in Supplementary Table 1.

Image preprocessing

We employed an in-house tool referred to as the “preprocessor.” This tool accepts an MRI study as input, where it subsequently separates the MRI sequences, identifies the specific body part, determines the projection, and categorizes the imaging type for each sequence. Here, the term “imaging type” refers to the type based on a relaxation time, such as T1-weighted, T2-weighted, and so on.

With the assistance of this tool and the sagittal segmentation model, which outlines the contours of vertebral bodies and intervertebral discs, among other objects (which are the discs, vertebral body, spinal canal, spinal cord with nerve roots, disc material bulge, and sacrum), we can identify the intervertebral disc or vertebral body level that corresponds to a given axial slice.

Neural network model

In this section, we describe the neural network model employed for the evaluation of lumbar stenosis severity. Our approach utilizes a three-stage model, as illustrated in Fig. 1, to assess the severity of three types of lumbar

stenosis: central, lateral recess, and foraminal. This model comprises a total of five sub-models: one segmentation model, one binary multilabel stenosis classification model and three stenosis severity classification models (central, lateral recess, and foraminal). The output of our three-stage model is depicted in Fig. 2.

Segmentation model

The segmentation model is designed with a U-Net architecture and operates on grayscale MRI axial images as its input. Its primary function is to produce a pixel-wise mask delineating 17 anatomical objects within the lumbar region. These objects include the disc, thecal sac, nerve, neural foramina, lateral recess, spinal canal, herniation, nerve roots, vertebral arch, spinous process, facet joint, articular process, ligamentum flavum, muscle, vein, artery, and kidney.

Binary multilabel stenosis classification model

Running in parallel with the segmentation model, the binary multilabel stenosis classification model adopts the RegNet architecture [18], specifically implementing the

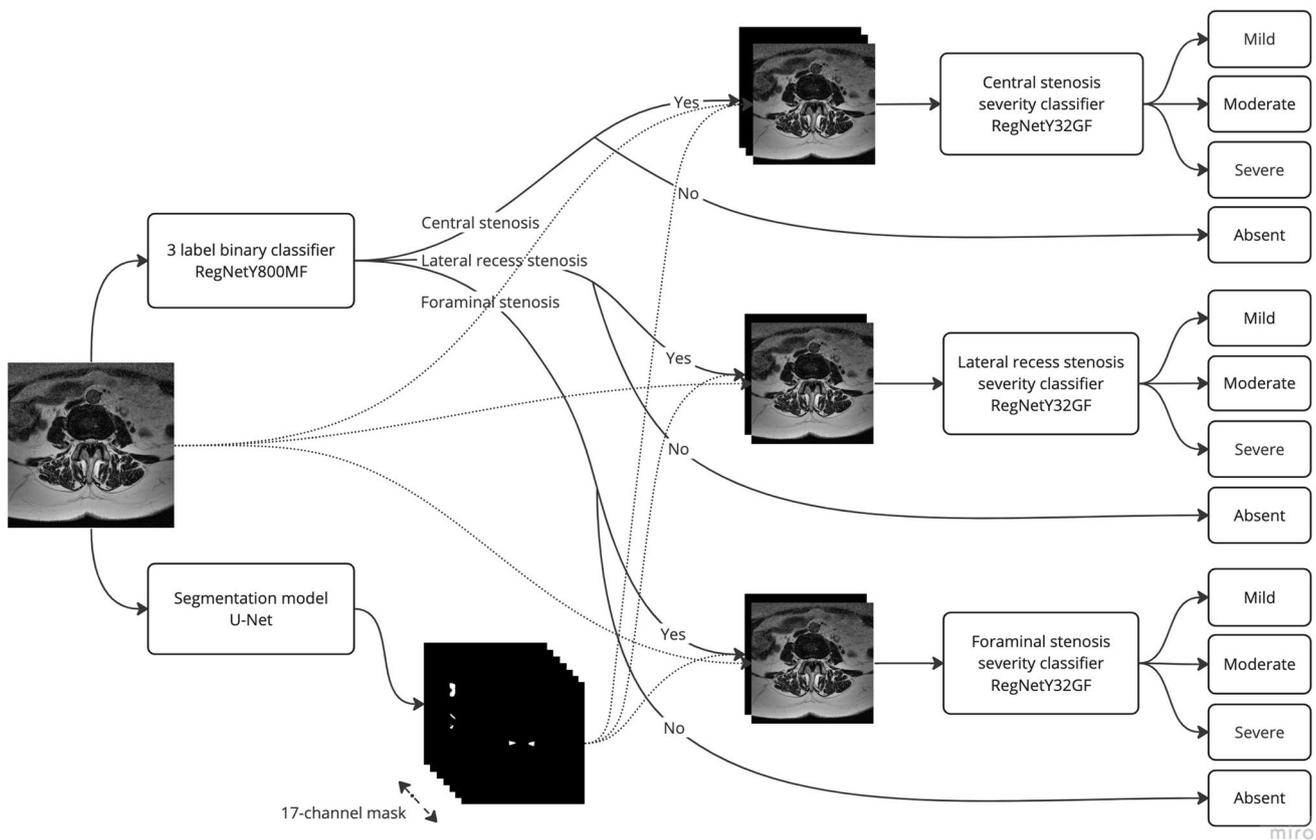


Fig. 1 Schema of the 3-stage neural network approach. It includes the segmentation model (detection of anatomical structures), the binary classification (presence or absence) for central, lateral recess and foraminal stenoses and the severity classification model (mild, moderate, severe)

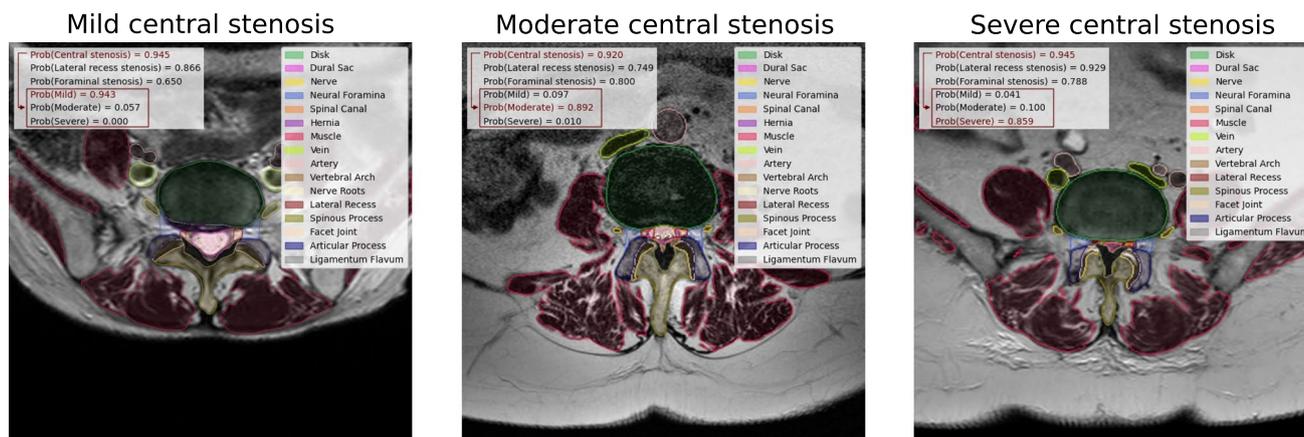


Fig. 2 Illustration of the segmentation model output, the stenosis binary classification output (presence/absence) and the severity classification model output for mild, moderate and severe central stenoses on axial plane from MR images

RegNetY800MF configuration. This model takes the same input axial image and provides an output indicating the presence or absence of each of the three types of lumbar stenosis: central, lateral recess, and foraminal.

Severity classification models

For each type of lumbar stenosis (central, lateral recess, and foraminal), we employ a separate severity classification model based on the RegNetY32GF architecture. Our approach involves customizing the input data by incorporating specific masks that correspond to the type of stenosis under examination. In the instance of central stenosis, we include masks for the thecal sac and spinal canal. In the case of lateral recess stenosis, we add the lateral recess object mask, while for foraminal stenosis, we integrate the neural foramina object mask. These masks are then combined with the original axial image, serving as essential inputs for our severity classification models. The reason behind this tailored approach is that these anatomical objects exhibit narrowing when the corresponding type of stenosis is present.

During the inference process, the segmentation model and the binary multilabel stenosis classification model run independently on the input image. Subsequently, for each type of stenosis, the binary classification model's prediction dictates whether to proceed with the corresponding severity classification model. Specifically, if the binary classification model predicts the absence of stenosis, the corresponding severity classification model is not utilized. However, if stenosis is predicted, both the source image and the mask output from the segmentation model are provided as inputs to the relevant severity classification model. The severity classification model then predicts the severity level, which can be categorized as mild, moderate, or severe for the respective type of lumbar stenosis.

Hyper-parameters of the different models are presented in Supplementary Table 2.

Statistical analysis

The outcomes of the CNN and the radiologists were compared to the reference standard for detection and grading of LSS. For binary classification of LSS, sensitivity, specificity, and area under the receiver operating characteristic curve (AUROC) were calculated. Compared to sensitivity and specificity, AUROC is a measure of prediction performance that does not depend on the discrimination threshold. For the multi-class grading of LSS, these metrics were computed in a one vs. all setting, where averages were weighted by the number of observations in each class (weighted average metrics).

Inter-rater agreement between the reference standard and the model and radiologists was assessed by Cohen's kappa. Level of agreement was defined as follows: 0.0–0.2 (slight), 0.21–0.4 (fair), 0.41–0.6 (moderate), 0.61–0.8 (substantial), 0.81–1.0 (almost perfect).

All confidence intervals for metrics in the *Results* section were calculated for the corresponding model. All calculations were implemented in Python (Version 3.7.4) using scikit-learn (Version 1.0.2).

Results

Binary classification

We first evaluated the presence of stenosis as a classification problem (presence vs absence) and looked at each stenosis type (central, lateral recess and foraminal)

separately. Sensitivity, specificity and AUROC are presented in Fig. 3. ROC curves are shown in Supplementary Fig. 1.

In the case of central canal stenosis, the sensitivity, specificity, and AUROC for the radiologists were 0.786 (95% confidence interval (CI): [0.619, 0.890]), 0.899 (95% CI: [0.878, 0.918]), 0.842 (95% CI: [0.779, 0.905]), respectively, while the metrics for the model were 0.971 (95% CI: [0.847, 0.999]), 0.864 (95% CI: [0.840, 0.885]), 0.963 (95% CI: [0.946, 0.980]). For binary inter-rater agreement, the average Cohen's kappa was substantial between radiologists and reference standard with a κ value of 0.372. The κ value of the CNN model was 0.431.

For lateral recess stenosis, the sensitivity, specificity, and AUROC for the radiologists were 0.713 (95% confidence interval (CI): [0.576, 0.804]), 0.898 (95% CI: [0.877, 0.916]), 0.805 (95% CI: [0.756, 0.854]), respectively, while the metrics for the model were 0.853 (95% CI: [0.689, 0.950]), 0.787 (95% CI: [0.759, 0.813]), 0.907 (95% CI: [0.868, 0.945]). For binary inter-rater agreement, the average Cohen's kappa was substantial between radiologists and reference standard with a κ value of 0.323. The κ value of the CNN model was 0.315.

For foraminal stenosis, the sensitivity, specificity, and AUROC for the radiologists were 0.879 (95% confidence interval (CI): [0.806, 0.932]), 0.877 (95% CI: [0.851, 0.901]), 0.878 (95% CI: [0.846, 0.910]), respectively, while the metrics for the model were 0.942 (95% CI: [0.890, 0.975]), 0.844 (95% CI: [0.817, 0.868]), 0.950 (95% CI: [0.934, 0.965]). For binary inter-rater agreement, the average Cohen's kappa was substantial between radiologists and reference standard with a κ value of 0.596. The κ value of the CNN model was 0.672.

Multiclass classification

Next, we compared the performance over the full LSS grading scale of absence, mild, moderate, and severe. Global metrics are shown in Fig. 4, severity class-specific metrics are shown in Supplementary Table 3, and severity class-specific confusion matrices are shown in Supplementary Table 4. To account for class imbalances, the average metrics were weighted depending on the number of observations per class. In the case of central canal stenosis, the sensitivity, specificity, and AUROC for the radiologists were 0.885 (95% confidence interval (CI): [0.855, 0.912]), 0.792 (95%

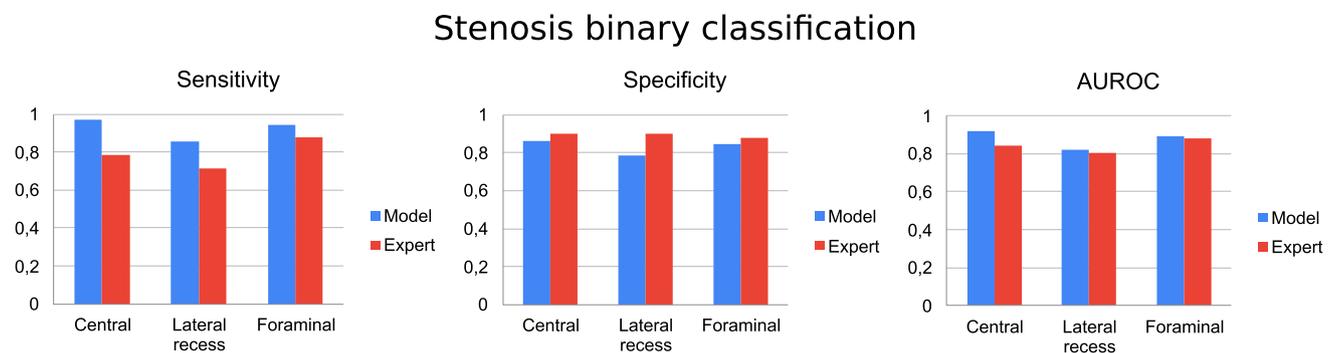


Fig. 3 Sensitivity, specificity, and AUROC comparisons of the radiologist panel and the model for presence/absence classification of lumbar spinal stenosis

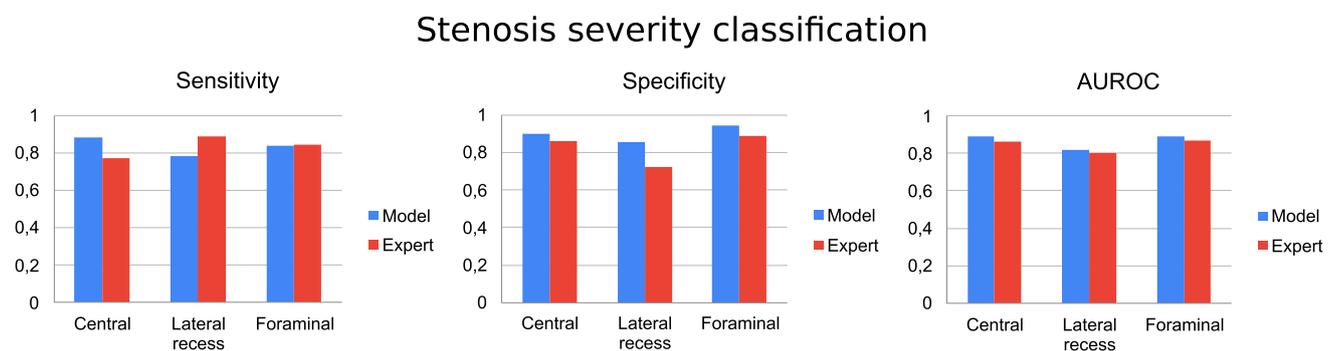


Fig. 4 Sensitivity, specificity, and AUROC comparisons of the radiologist panel and the model for severity classification of lumbar spinal stenosis

CI: [0.631, 0.892]), 0.838 (95% CI: [0.772, 0.902]), respectively, while the metrics for the model were 0.855 (95% CI: [0.821, 0.884]), 0.969 (95% CI: [0.849, 0.998]), 0.912 (95% CI: [0.878, 0.947]). For inter-rater agreement, the average Cohen's kappa between the radiologists and the reference standard was 0.376 and 0.310 between the model and the reference standard.

For lateral recess stenosis, the sensitivity, specificity, and AUROC for the radiologists were 0.887 (95% confidence interval (CI): [0.859, 0.910]), 0.772 (95% CI: [0.589, 0.811]), 0.804 (95% CI: [0.755, 0.884]), respectively, while the metrics for the model were 0.783 (95% CI: [0.745, 0.816]), 0.856 (95% CI: [0.698, 0.950]), 0.819 (95% CI: [0.755, 0.884]). For inter-rater agreement, the average Cohen's kappa between the radiologists and the reference standard was 0.359 and 0.199 between the model and the reference standard.

For foraminal stenosis, the sensitivity, specificity, and AUROC for the radiologists were 0.845 (95% confidence interval (CI): [0.801, 0.884]), 0.890 (95% CI: [0.825, 0.937]), 0.868 (95% CI: [0.831, 0.905]), respectively, while the metrics for the model were 0.836 (95% CI: [0.794, 0.872]), 0.943 (95% CI: [0.895, 0.972]), 0.890 (95% CI: [0.861, 0.918]). For inter-rater agreement, the average Cohen's kappa between the radiologists and the reference standard was 0.620 and 0.637 between the model and the reference standard.

Once again, the model showed comparable performance compared to the average radiologist subspecialist on the panel, across all types of stenosis.

Discussion

The purpose of this study was to develop an artificial intelligence approach to classify the presence and severity of stenosis and to demonstrate its efficacy as an accurate and consistent diagnostic tool. We designed a three-stage convolutional neural network (CNN) approach to segment anatomical structures, classify the presence of lumbar spinal stenoses (central, lateral recess, foraminal) and assess its severity on spine MRI. Evaluation on an external data set demonstrated comparable performance compared to a panel of 7 radiologist subspecialists for the detection and severity classification of LSS. For binary diagnosis of central, lateral recess and foraminal stenosis, the performance of the CNN model was higher as compared to the radiologist average, achieving an AUROC of (0.963, 0.907, 0.950) compared to (0.842, 0.805, 0.878) of the radiologist average, respectively. Accordingly, for multi-class severity grades, the model yielded an AUROC of (0.912, 0.819, 0.890) compared to the radiologist average of (0.838, 0.804, 0.868), for central, lateral recess and foraminal stenosis, respectively.

Our neural network was trained with a large dataset (21,702 images), which is often considered sufficient for training a model with high accuracy, in particular in comparison to previous AI methods that often trained on ~500–10000 images [14, 15]. However, for severe stenoses, the number of images was in the order of several hundred, which led to slightly lower precision.

Previous studies have investigated the performance of various ML models in detecting and grading LSS on lumbar spine MRI. Most of the studies used two experts to assess the models' performance, while our approach consisted of an expert panel of 7–8 radiologists. Hallinan et al. designed a two-stage deep learning model, using the first model to place region of interests at the central canal, lateral recesses, and neuroforamina with subsequent grading by the second model¹¹. Contrasting our results for central canal classification, their model showed higher inter-rater agreements but better performance of the radiologists for both binary (0.98) and multi-class (0.89) classification compared to the deep learning model. On a larger data set of 7108 lumbar spine MRI on axial and sagittal slices, Lu et al. developed a U-Net architecture that was based on weakly supervised natural language processing labels derived from radiology reports. The accuracy of multi-class LSS (normal-mild-moderate-severe) grading reached an average of 70.6% [16]. Another work trained a two-stage CNN to localize the thecal sac on axial T2-weighted MRI studies at a single motion segment (L4-5) with secondary grading of the localized region [17]. The grading consistency was 77.5% between two expert labels, while the model achieved accuracies of 77.9–83.0% compared to the expert labels. Most recently, Su et al. generated a multi-task classification network on a large data set of axial lumbar spine T2-weighted MRI for grading of disc herniation, foraminal stenosis, and LSS, that was graded by two clinicians as reference standard [18]. The models' accuracies ranged from 79.7 to 87.0% for the external and internal data set, respectively.

Compared to previous works, our three-stage model provides the most comprehensive assessment of lumbar stenosis as it combines segmentation, stenosis classification and severity grading. The approach of combining input from a source axial image with the stenosis-related anatomical object mask output from the segmentation model provides the classification model with precise spatial information and a contextual understanding of how the stenosis-related anatomical object interacts with the surrounding structures. This results in more accurate predictions when contrasted with the approach of solely passing the source image to the classification model [19]. In summary, our three-stage model effectively addresses the limitations associated with traditional approaches by seamlessly integrating both segmentation and classification techniques.

The current study has limitations, much of which are theoretical limitations with the nature of the current undertaking—the lack of a “gold standard” or “true north” for comparative analysis of the current model. That is, the lack of a firmly established consensus around the MRI grading scale definitions and the ground truth was established by common radiologic grading. This explains the presence of disagreement between the raters. Second, the labor-intensive manual annotation process limits the number of MRI studies that could be used for training of the neural network model. Furthermore, there was no inter-rater consensus established regarding the labeling process. Third, combining MRI and CT imaging can enhance the accuracy and sensitivity of neural networks in detecting stenosis caused by osteophytes, since CT helps with the visualization of solid structures, making it valuable for detecting osteophytes and bony abnormalities. Fourth, patients with severe stenosis can experience little symptoms [20], thus revealing the need for an AI based not only on imaging but also on symptoms.

Conclusions

In conclusion, we have demonstrated that the integration of CNN models is able to diagnose and grade LSS on lumbar spine MRI with high accuracy, reliability, and post-hoc interpretability comparable or superior to expert radiologists and may function as a supporting diagnostic tool. Combined with the patient symptoms and patient’s conservative therapy failure, this tool could be very valuable for surgery decisions. In future works, the CNN model should be expanded to assess a broader spectrum of degenerative findings that occur within the natural course of LSS, including stenosis of the subarticular/lateral recesses and/or neuroforaminal stenosis.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00586-023-08089-2>.

Declarations

Conflict of interest No authors have any conflicts of interest to disclose. All authors have no financial or non-financial interests directly or indirectly related to this work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Kalichman L, Cole R, Kim DH, Li L, Suri P, Guermazi A, Hunter DJ (2009) Spinal stenosis prevalence and association with symptoms: the Framingham Study. *Spine J* 9(7):545–550
- Wu A-M, Zou F, Cao Y, Xia D-D, He W, Zhu B, Chen D, Ni W-F, Wang X-Y, Kwan K (2017) Lumbar spinal stenosis: an update on the epidemiology, diagnosis and treatment. *AME Med J* 2(5)
- Katz JN, Harris MB (2008) Lumbar spinal stenosis. *N Engl J Med* 358(8):818–825
- Pannell WC, Savin DD, Scott TP, Wang JC, Daubs MD (2015) Trends in the surgical treatment of lumbar spine disease in the United States. *Spine J* 15(8):1719–1727
- Lurie J, Tomkins-Lane C (2016) Management of lumbar spinal stenosis. *BMJ* 352:h6234
- Kreiner DS, Shaffer WO, Baisden JL, Gilbert TJ, Summers JT, Toton JF, Hwang SW, Mendel RC, Reitman CA (2013) An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis (update). *Spine J* 13(7):734–743
- Majidi H, Shafizad M, Niksolat F, Mahmudi M, Ehteshami S, Poorali M, Mardanshahi Z (2019) Relationship between magnetic resonance imaging findings and clinical symptoms in patients with suspected lumbar spinal canal stenosis: a case-control study. *Acta Informatica Medica* 27(4):229
- Steurer J, Roner S, Gnannt R, Hodler J (2011) Quantitative radiologic criteria for the diagnosis of lumbar spinal stenosis: a systematic literature review. *BMC Musculoskelet Disord* 12(1):175
- Andreisek G, Deyo RA, Jarvik JG, Porchet F, Winklhofer SFX, Steurer J, LSOS working group (2014) Consensus conference on core radiological parameters to describe lumbar stenosis - an initiative for structured reporting. *Eur Radiol* 24(12):3224–3232
- Lehnen NC, Haase R, Faber J, Rüber T, Vatter H, Radbruch A, Schmeel FC (2021) Detection of degenerative changes on MR images of the lumbar spine with a convolutional neural network: a feasibility study. *Diagnostics* 11(5):902
- Jamaludin A, Kadir T, Zisserman A (2017) SpineNet: Automated classification and evidence visualization in spinal MRIs. *Med Image Anal* 41:63–73
- Hallinan JTPD, Zhu L, Yang K, Makmur A, Algazwi DAR, Thian YL, Lau S, Choo YS, Eide SE, Yap QV, Chan YH, Tan JH, Kumar N, Ooi BC, Yoshioka H, Quek ST (2021) Deep learning model for automated detection and classification of central canal, lateral recess, and neural foraminal stenosis at lumbar spine MRI. *Radiology* 300(1):130–138
- Schizas C, Theumann N, Burn A, Tansey R, Wardlaw D, Smith FW, Kulik G (2010) Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images. *Spine* 35(21):1919–1924
- D’Antoni F, Russo F, Ambrosio L, Vollero L, Vadala G, Merone M, Papalia R, Denaro V (2021) Artificial intelligence and computer vision in low back pain: A systematic review. *Int J Environ Res Public Health* 18(20):10909
- Park J, Yang J, Park S, Kim J (2023) Deep learning-based approaches for classifying foraminal stenosis using cervical spine radiographs. *Electronics* 12(1):195
- Lu J-T, Pedemonte S, Bizzo B, Doyle S, Andriole KP, Michalski MH, Gonzalez RG, Pomerantz SR (2018) Deep spine: automated lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading using deep learning. In: Doshi-Velez F, Fackler J, Jung K, Kale D, Ranganath R, Wallace B, Wiens J (eds) Proceedings of the 3rd machine learning for healthcare conference, volume 85 of proceedings of machine learning research. PMLR, pp 403–419

17. Won D, Lee H-J, Lee S-J, Park SH (2020) Spinal stenosis grading in magnetic resonance imaging using deep convolutional neural networks. *Spine* 45(12):804–812
18. Su Z-H, Liu J, Yang M-S, Chen Z-Y, You K, Shen J, Huang C-J, Zhao Q-H, Liu E-Q, Zhao L, Feng Q-J, Pang S-M, Li S-L, Lu H (2022) Automatic grading of disc herniation, central canal stenosis and nerve roots compression in lumbar magnetic resonance image diagnosis. *Front Endocrinol* 13:890371
19. Kushchayev SV, Glushko T, Jarraya M, Schuleri KH, Preul MC, Brooks ML, Teytelboym OM (2018) ABCs of the degenerative spine. *Insights Imaging* 9(2):253–274
20. Andrasinova T, Adamova B, Buskova J, Kerkovsky M, Jarkovsky J, Bednarik J (2018) Is there a correlation between degree of radiologic lumbar spinal stenosis and its clinical manifestation? *Clin Spine Surg* 31(8):E403–E408

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.