Neural network model for the detection and classification of central canal stenosis of the lumbar spine in MRI images

Vladislav Tumko
v.tumko@remedylogic.com

Jack Kim
j.kim@remedylogic.com

Burton Drayer
burton.drayer@mountsinai.org

October 2021

Abstract

Background: MRI studies of the lumbosacral spine are conducted to confirm clinical symptom findings indicative of lumbar spinal stenosis (LSS). The efficiency and reliability of radiological identification and grading could be enhanced by automation through the use of neural network models.

Objective: To develop a neural network based imaging model to classify the presence and severity of central canal stenosis and to demonstrate its efficacy as an accurate and consistent diagnostic tool.

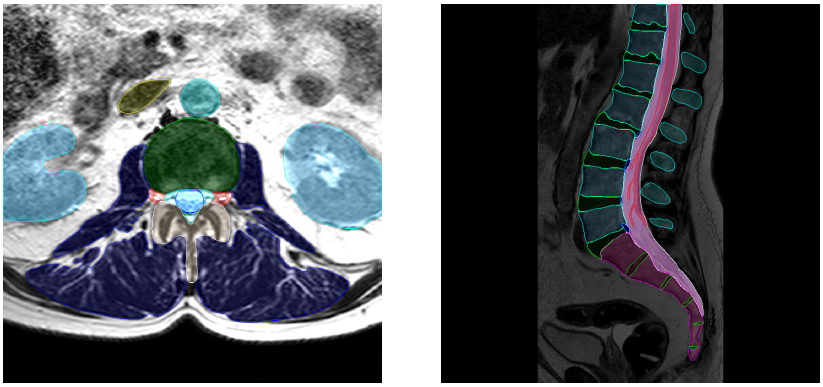
Materials and Methods: A neural network model was developed to automate the classification of the presence and severity of central canal stenosis. The models were trained on 200 annotated MRI studies consisting of T-2 weighted sagittal and axial plane images at each lumbar level. Evaluation of model accuracy was performed on an external validation set of 42 MRI studies graded on a scale of absent, mild, moderate or severe by a panel of 4 radiologists. The reference standard was determined by majority rule and in case of disagreement, adjudicated by a further external radiologist. The radiologists interpretations were then compared to the interpretation of the model.

Results: The model showed comparable performance to the radiologist average both in terms of the determination of presence/absence of stenosis as well as severity classification. For presence/absence, the sensitivity, specificity, AUC-ROC for the neural network model were (0.99, 0.99, 0.97) compared to the radiologist average of (0.93, 0.93, 0.87). For severity grades, the model yielded metrics of (0.94, 0.98, 0.94) compared to the radiologist average of (0.81, 0.94, 0.91).

Conclusion: The neural network model shows comparable performance to radiologist subspecialists for the detection and classification of central canal stenosis. The integration of neural network models in the detection of LSS could bring higher accuracy, efficiency, consistency, and post-hoc interpretability in diagnostic practices.

1 Introduction

Lumbar spinal stenosis (LSS) is one of the most commonly diagnosed spinal pathologies in the United States, with prevalence of degenerative LSS among adults estimated to be in the range of 1.7-13.1% and materially higher among older age groups [3, 8, 13]. The clinical symptoms of LSS may include lower back, lower extremity, and gluteal pain and a substantial loss of mobility [9], which lends it to be the most common reason for spinal surgery in patients above the age of 65 years



(a) axial segmentation (b) sagittal segmentation

Figure 1: Axial and sagittal segmentation of key anatomical objects

[6, 8]. Hence, accurate and reliable diagnosis and subsequent treatment formulation has important economic and social impact.

LSS is characterized by a narrowing of the central spinal canal, lateral recesses, or intervertebral foramen, which in turn causes compression in associated neural and vascular structures [6, 10]. While LSS is defined as a clinical syndrome and diagnosed as such, imaging is commonly used as a confirmatory tool. Magnetic Resonance Imaging (MRI), because of its excellent soft tissue contrast, is frequently used for confirmation of clinical symptom based diagnoses and determining the optimal treatment course. A number of studies [1, 11, 12] have attempted to identify 10-27 core quantitative radiological criteria for the diagnosis of lumbar spinal stenosis. While certain measurements such as the dural sac cross-sectional area are widely recognized in diagnostic import, a definitive relationship between clinical symptom manifestation and specific radiological measurements remain inconclusive, lacking universal consensus [8, 11]. While expert judgment combines qualitative assessment of morphological features as well quantitative measurements, the lack of consensus criteria raises concerns about rating reliability.

In this study we automate this process of assessing central canal stenosis, through a neural network imaging model applied to MRI studies. We demonstrate that for the classification of the presence and severity of central canal stenosis, the performance of the model is comparable to a panel of radiologist subspecialists.

2 Materials and Methods

2.1 Data and Annotation

The training data set consists of 200 MRI studies of T-2 weighted sagittal and axial plane images at each lumbar level. On sagittal plane images, the discs, vertebrae, spinal canal, spinal cord with nerve roots, spinous process, sacrum, and hernia were labeled for segmentation. For axial images, the discs, spinal canal, dural sac, neural foramina, nerve, facet, kidney, aorta, vein, muscles, hernia were labeled. In addition to the key anatomical bodies, the studies were also labeled by radiologist subspecialists on a scale of 0 (absent), 1 (mild), 2 (moderate), 3 (severe) stenosis. An external data set of 42 studies were annotated for final model validation.

2.2 Neural Network Model

We develop a neural network model that identifies key anatomical abnormalities that are strongly indicative of central canal stenosis on both the sagittal and axial planes at each intervertebral levels.

The identified abnormalities are combined to determine the presence and severity of central canal stenosis. See Figure 1.

2.3 Evaluation

An external validation set of 43 MRI studies were graded on a scale of absent, mild, moderate or severe by a panel of 4 radiologist subspecialists. The reference standard was determined by voting rule and in case of disagreement, adjudicated by a further external radiologist. The radiologist interpretations were then compared to the interpretation of the model. Sensitivity, specificity, AUC-ROC, and inter-rater agreement kappa values were calculated.

3 Results

3.1 Binary Classification

Rater	Sensitivity	Specificity	AUC-ROC
Rater_1	0.93	0.93	0.84
Rater_2	0.94	0.94	0.94
Rater_3	0.93	0.93	0.85
Rater_4	0.94	0.94	0.86
Rater_average	0.93	0.93	0.87
Model	0.99	0.99	0.97

Table 1: Binary Classification Performance

In order to compare binary classification results, we combine grades 0 and 1 as absent; 2 and 3 as present. The sensitivity, specificity, and AUC-ROC metrics are shown in Table 1 and Figure 2. For inter-rater agreement, the Fleiss kappa between raters was 0.844 and 0.963 between the model and reference standard. The model shows performance comparable to radiologists and high degree of agreement with the reference standard ratings.

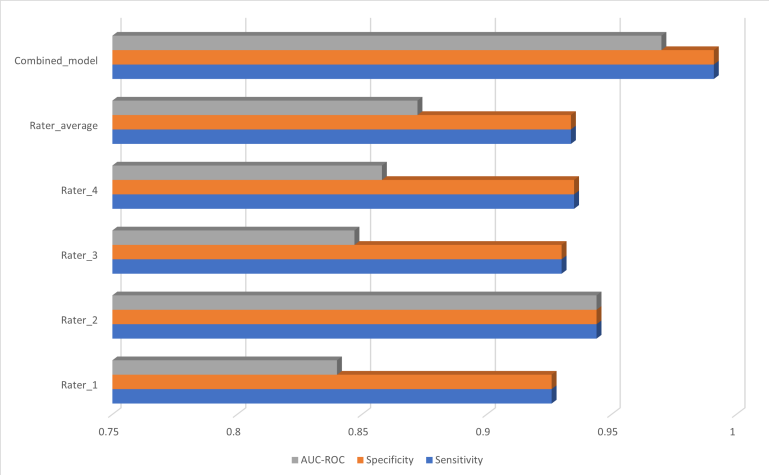


Figure 2: Binary Classification Performance

3.2 Multiclass Classification

The performance metrics for multi-class severity classification are shown in Table 2 and Figure 3. Note that the AUC-ROC is computed as the AUNP weighted average as the classes are imbalanced. For severity ratings, the model performed notably better than the rater average. For the multi-class reliability we compute the ordinal Krippendorff Alpha between raters was 0.897 and 0.963 between the model and reference standard.

Rater	Sensitivity	Specificity	AUC-ROC
Rater_1	0.80	0.93	0.83
Rater_2	0.81	0.94	-
Rater_3	0.81	0.94	0.83
Rater_4	0.83	0.94	0.84
Rater_average	0.81	0.94	0.84
Model	0.94	0.98	0.94

Table 2: Multiclass Performance

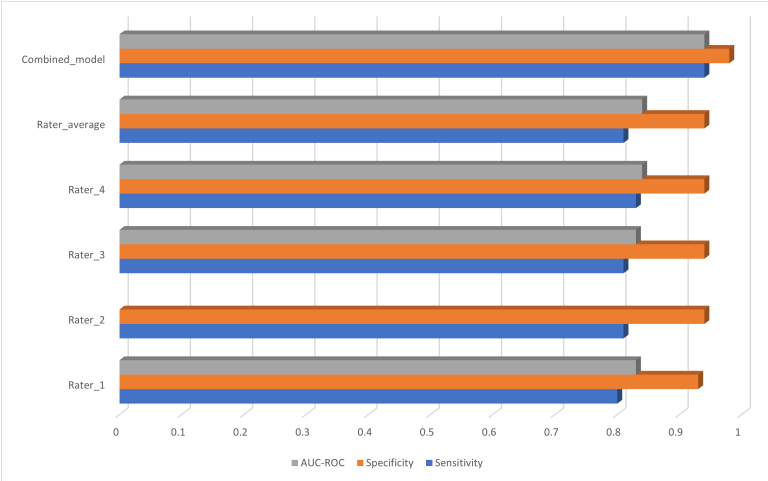


Figure 3: Multiclass Classification Performance

4 Discussion

An neural network imaging model that detects key anatomical abnormalities shows comparable performance to radiologist subspecialists for the detection and severity classification of central canal stenosis. The integration of neural network models in the detection of spinal stenosis could bring accuracy, expediency, consistency, and post-hoc interpretability in diagnostic practices. Research scope should be expanded to lateral recess and foraminal stenosis in future work.

References

[1] Andreisek, G., Deyo, R.A., Jarvik, J.G., Porchet, F., Winklhofer, S.F. and Steurer, J., 2014. Consensus conference on core radiological parameters to describe lumbar stenosis-an initiative for structured reporting. European radiology, 24(12), pp.3224-3232.

[2] De Schepper, E.I., Overdevest, G.M., Suri, P., Peul, W.C., Oei, E.H., Koes, B.W., Bierma-Zeinstra, S.M. and Luijsterburg, P.A., 2013. Diagnosis of lumbar spinal stenosis: an updated systematic review of the accuracy of diagnostic tests. Spine, 38(8), pp.E469-E481.

[3] Deyo, R.A., 2010. Treatment of lumbar spinal stenosis: a balancing act. The spine journal, 10(7), pp.625-627.

[4] Fanuele, J.C., Birkmeyer, N.J., Abdu, W.A., Tosteson, T.D. and Weinstein, J.N., 2000. The impact of spinal problems on the health status of patients: have we underestimated the effect?. Spine, 25(12), pp.1509-1514.

[5] Hallinan, J.T.P.D., Zhu, L., Yang, K., Makmur, A., Algazwi, D.A.R., Thian, Y.L., Lau, S., Choo, Y.S., Eide, S.E., Yap, Q.V. and Chan, Y.H., 2021. Deep learning model for automated detection and classification of central canal, lateral recess, and neural foraminal stenosis at lumbar spine MRI. Radiology, p.204289.

[6] Lurie, J. and Tomkins-Lane, C., 2016. Management of lumbar spinal stenosis. Bmj, 352.

[7] Fairbank, J., Jamaludin, A., Lootus, M., Kadir, T., Zisserman, A., Urban, J., Battié, M., McCall, I. and Genodisc Consortium, 2017, May. AUTOMATION OF READING OF RADIOLOGICAL FEATURES FROM MAGNETIC RESONANCE IMAGES (MRI's) OF THE LUMBAR SPINE WITHOUT HUMAN INTERVENTION IS COMPARABLE WITH AN EXPERT RADIOLOGIST (ISSLS Prize-Bioengineering Science): 56. In Spine Journal Meeting Abstracts (p. 56). LWW.

[8] Kalichman, L., Cole, R., Kim, D.H., Li, L., Suri, P., Guermazi, A. and Hunter, D.J., 2009. Spinal stenosis prevalence and association with symptoms: the Framingham Study. The spine journal, 9(7), pp.545-550.

[9] Katz, J.N. and Harris, M.B., 2008. Lumbar spinal stenosis. New England Journal of Medicine, 358(8), pp.818-825.

[10] Kreiner, D.S., Shaffer, W.O., Baisden, J., Gilbert, T., Summers, J., Toton, J., Hwang, S., Mendel, R. and Reitman, C., 2011. Diagnosis and treatment of degenerative lumbar spinal stenosis. Boulevard, USA: North American Spine Soc.

[11] Majidi, H., Shafizad, M., Niksolat, F., Mahmudi, M., Ehteshami, S., Poorali, M. and 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), p.229.

[12] Steurer, J., Roner, S., Gnannt, R. and Hodler, J., 2011. Quantitative radiologic criteria for the diagnosis of lumbar spinal stenosis: a systematic literature review. BMC musculoskeletal disorders, 12(1), pp.1-9.

[13] Wu, A.M., Zou, F., Cao, Y., Xia, D.D., He, W., Zhu, B., Chen, D., Ni, W.F., Wang, X.Y. and Kwan, K.Y.H., 2017. Lumbar spinal stenosis: an update on the epidemiology, diagnosis and treatment. AME Medical Journal.