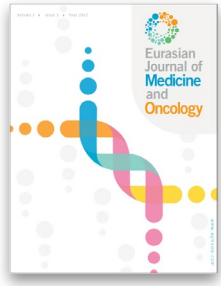
EJMO

DOI: 10.14744/ejmo.2023.59207
EJMO 2023;7(1):42–48



Research Article

Clinical Assessment of CoLumbo Deep Learning System for Central Canal Stenosis Diagnostics

Radoslav Georgiev,[1,2] Marianna Novakova,[2] Kristina Bliznakova[3]

[1]Department of Imaging Diagnostics, Interventional Radiology and Radiotherapy, Medical University, Varna, Bulgaria
[2]Department Medical Imaging, University Hospital St Marina, Varna, Bulgaria
[3]Department of Medical Devices, Electronic and Information Technologies in Healthcare, Medical University, Varna, Bulgaria

Abstract

Objectives: There is a great variability of inter-observer disagreements for central stenosis diagnostics depending on the used classification. This study investigates the level of agreement between lumbar magnetic resonance imaging (MRI) reports created by a deep learning neural network (CoLumbo) and the radiologists' reading.

Methods: A total of 382 (53.4 % females, 46.6 % males and average age 49.52±13.20) prospective consecutive patients in 3 different healthcare centers referred to L-spine MRI for back or leg pain were analyzed by the software CoLumbo for the presence of stenosis on all lumbar levels, by radiologists using it and radiologists not using the dedicated software. In case of disagreement between radiologists, a radiologist-arbiter opinion was used to establish majority opinion. The total number of evaluated levels was 1762.

Results: There were 156 debatable cases of disagreements between radiologists using the software, and radiologists, not using CoLumbo, for the presence of central stenosis. In 18 cases, the arbiter opinion has coincided with that of the radiologist not using the software. In 138 cases, the former has coincided with that of the radiologist using the software CoLumbo. Most of the cases of disagreement are borderline cases. The reported sensitivity and specificity of CoLumbo was 92.70% and 99.04%, respectively.

Conclusion: The study showed that the radiologist using the CoLumbo software achieved best results. The results of the algorithm were inferior but still better than radiologists not using the software in any published study.

Keywords: CoLumbo software, deep learning, Dural Sac, lumbar spinal stenosis, lumbar levels

Cite This Article: *Georgiev R, Novakova M, Bliznakova K. Clinical Assessment of CoLumbo Deep Learning System for Central Canal Stenosis Diagnostics. EJMO 2023;7(1):42–48.*

Lumbar spinal stenosis (LSS) is a type of lumbar degenerative spine disease and is among the most common causes of spine surgery. Its radiographic and anatomical findings is characterized by narrowing of the spinal canal[1] and typically involves L4-L5, L5-S1 levels and less often L3-L4 levels.[2] Narrowing may occur in the central spinal canal, in the area under the facet joints, or more laterally, in the neural foramina. Amongst the features that are specific to the lumbar spinal stenosis are bulging of the intervertebral disk, thickening of the ligamentum flavum, and hypertrophy of the facet joints based on axial view. Features such as loss of disk height, disk protrusion, and facet-joint osteoarthritis, all leading to foraminal stenosis (when stenosis affects the spinal foramen), based on sagittal view. Among the clinical symptoms of lumbar spinal stenosis are lower extremity pain, weakness, and low back pain (LBP) and can lead to a reduction in the quality of life.[3] In case of severe chronic pain, the LSS patients may benefit either of instru-

Address for correspondence: Kristina Bliznakova, MD. Department of Medical Devices, Electronic and Information Technologies in Healthcare, Medical University, Varna, Bulgaria
Phone: +35952677050 **E-mail:** kristina.bliznakova@mu-varna.bg

Submitted Date: December 21, 2022 **Revision Date:** January 31, 2023 **Accepted Date:** February 01, 2023 **Available Online Date:** March 08, 2023

mented spinal fusion surgery[4] or decompression surgery.[5] Complications caused by fusion surgery have been reported including higher morbidity, pseudarthrosis and degeneration of adjacent segments.[6] Therefore, the right and timely diagnosis is extremely important. Based on clinical symptoms, both surgeons and physicians specify the severity of stenosis and make the decision for the type of the lumbar decompression surgery. However, agreement in deciding the severity or level of stenosis and the classification of stenosis among radiologists, neurologists and surgeons may be poor[7-9] as well a poor correlation between clinical symptoms and signs, and radiology findings could be present.[10] These limitations result in a considerable amount of subjective judgment for decision making in lumbar decompression surgery, and level decompressed among surgeons.[11] In addition to this, studies have shown that there is a wide variability in lumbar spinal canal dimensions among patients who do not have clinical spinal stenosis.[2, 12] Karantanas et al.[2] showed that both the somatometric parameters as well as the age have statistically significant correlation with many of the measured indices. Other studies have shown measured L4 canal diameter in the black population for males and females as 15.6 mm and 14.1 mm, respectively.[13] There is a difference in measured LSS parameters between individuals from different sex and age. Twomey et al.[14] compared two adult age groups in both males and females and showed a significant decline in the lumbar spinal canal anteroposterior diameter in both sexes for the older group. Differences in the spinal canal cross-sectional area of the lumbar spinal canal of women and men were reported as well by Griffith et al.[12]

Computer algorithms are increasingly entering medicine, however, they are currently focused mainly on analyzing medical images of the brain, heart, and lungs. Resistance to use computer diagnostic systems is declining,[15] and doctors increasingly appreciate the potential that computer diagnostic products in medicine can provide. In the field of LSS diagnosis, the use of neural network for automated MRI grading turns out to be of invaluable assistance.[16, 17] The use of dedicated software does not change the usual radiologist's workflow. One such application is the CoLumbo software (https://columbo.me/), built to provide confirmation for users to accept or reject an output from optional analysis and is not intended to replace the clinician's diagnosis. The output generated from this software is not intended to be used directly for final diagnosis, which is the sole responsibility of the clinician. It only provides the results with the findings in a text form suitable for further reporting.

The objective of this study is to evaluate and demonstrate the safety of CoLumbo's usage for assessment of findings in the lumbar region and in providing segmentation and measurements. This is achieved by measuring the accuracy of a radiologist using the software versus a radiologist not using the software and the accuracy of the artificial intelligence (AI) algorithm itself. The specific objectives of this clinical testing are to prove: (a) that the accuracy of a radiologist using CoLumbo is not worse than the accuracy of a radiologist not using CoLumbo; (b) the algorithm's accuracy for assessment of lumbar spinal canal stenosis based on MRI images.

Methods

The prospective study "Clinical Trial with Columbo software" was accepted by the Ethical Committee for Clinical Trials of the Ministry of Healthcare of Bulgaria with a protocol ЕКК/СГ-0687 from 06.08.2020.

Patients Data and Radiologists

The target population is patients referred to L-Spine MRI for back and/or leg pain or other spine-related symptoms. To reduce the probability of deviation due to the selection of specific patients, a prospective multicenter study on consecutive patients is conducted to cover various cases and avoid variance by gender, age, or type of disease. The clinical investigation with CoLumbo software was organized in three medical centers from different locations in Bulgaria where the software was installed and only three researchers (each one per medical center) had access to it. The investigation was carried for a period of two months September and October 2020. The number of participants in each center was between 100 and 150, while the total number was 382. All these cases are acquired in the centers involved in the clinical study, where the investigators are working. There was no need for a separate control group since the product under trial does not have a therapeutic effect on patients. Patients below the age of 18 or over 70 years as well as pregnant women and persons with concomitant pathology – scoliosis were not included in the clinical study, since there is a significant difference in the spine's morphology in these persons.

Five different certified radiologists in three different centers participated equally in each of the 3 roles - radiologists using, not using the software, and arbiter. In case of disagreement, a third radiologist arbiter with access to the tool was used to establish majority opinion. All radiologists had more than 20 years experience in the MRI field. Radiologists working with the software and those without the software were independent per specific case for every case, but as a whole, they were rotated between these two roles.

The patient's images used in the trial are from four different

MRI machines, three models (Aera, Signa HDxt, Verio) from two different manufacturers (SIEMENS, GE MEDICAL SYSTEMS). Different machine protocols are used, common to the centers where the study was performed, but still with mandatory axial and sagittal T2 series in 2D and 3D.

The 382 consecutive patients in three different centers referred for L-spine MRI were prospectively analyzed for the presence of central stenosis at all lumbar levels. Among the 382 studies, there were 3 to 5 (4.63 on average) levels with available sagittal and axial images per study; the total number of evaluated levels was 1762.

COLUMBO software

COLUMBO software supports some of the spine's most common pathologies: disc herniation, bulging, stenosis, spondylolisthesis, hypo-, and hyper-lordosis. It is based on AI algorithm originally developed by Georgiev et al.[18]. CoLumbo version 2.0 is a software for visualization and analysis of lumbar spine's medical MRI images. It is an assistant type of software whose main task is to detect a set of common pathologies through the integrated-into-it artificial intelligence. CoLumbo evaluates these pathologies' characteristics and gets the radiologist's attention to them, marking relevant tissues and measurements with different colors in the images and automating part of the report writing. Its current version is intended for use only by radiologists in medical institutions (radiologist and spine surgeons in USA). Convolutional neural networks provide segmentations. Using standard geometric operations like drawing tangents, bisectional lines, and projections, CoLumbo determines standard measurements like distance, area, and angles similar to almost all other AI-based segmentation algorithms. The supported field strength is 1.5T and 3T. All brands and models are supported.

A screen shot from the Reports module of the pre-commercial deep-learning based AI tool - CoLumbo is depicted in Figure 1. This figure shows the segmented Dural Sac (light blue), vertebral body (green), intervertebral discs (dark blue), lamina and spinous process (dark-purple), ligamentum flavum (brown), herniation (red), nerve roots (light red), aorta (purple) and sacrum (light-green). This module is used by the radiologists to evaluate the presence of central stenosis with the assistance of the software.

Segmentation, Measurements and Statistical Analysis

The 382 consecutive patients in three different centers referred for L-spine MRI were prospectively analyzed for the presence of stenosis at all lumbar levels. The grading is based on assessment of both sagittal and axial images. The software segments the tissues in both type of images and radiologists studied a binary 'presence' or 'absence' of stenosis on the images. CoLumbo provides segmentation of the following tissues: (a) vertebra (on axial and sagittal slice

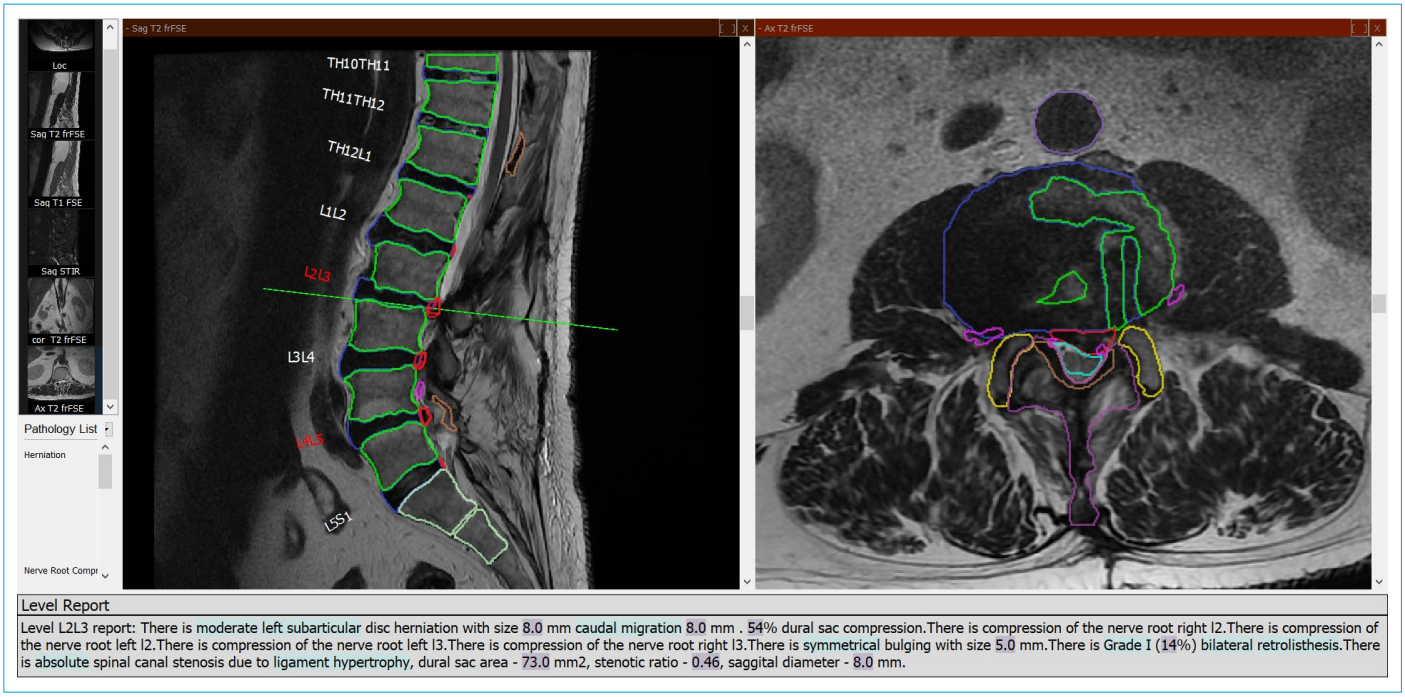


Figure 1. A screenshot from the Report module of the CoLoumbo software: Dural Sac (light blue), vertebrae (green), intervertebral discs (dark blue), lamina and spinous process (dark purple), ligamentum flavum (brown), herniation (crimson), nerve roots (light red), aorta (purple) and sacrum (light green).

around mid-sagittal, 35 mm); (b) part of the disk without the herniation (on axial slice, on sagittal slice, around mid-sagittal, 35 mm); (c) part of the disk with the herniation (on axial slice) without extraforaminal and sequestered part; (d) Dural Sac (on axial slice); (e) ligamentum flavum (on axial slice); (f) nerve roots (on axial slice); (g) aorta and/or iliac artery (on axial slice); (h) sacrum (on sagittal slice). Diagnosis of central stenosis classification is provided based on the Dural Sac cross-sectional area less than 100mm^2.[19]

The performance of radiologists aided and radiologists not assisted by CoLumbo and software performance are evaluated by using accuracy, and level of agreement. Sensitivity, specificity, the positive and the negative predictive values are used to evaluate the software performance. Sensitivity measures the proportion of positives (levels, classified by the software as having central stenosis) that are correctly identified and are given as:

$$Sensitivity = \frac{TP}{TP+FN} \quad (1)$$

Specificity is defined as correctly classified cases that are negative (i.e. the proportion of those levels which do not have central stenosis and are correctly identified as not having central stenosis).

$$Specificity = \frac{TN}{TN+FP} \quad (2)$$

where TP = number of true positives, TN = number of true negatives, FN = number of false negatives, FP = number of false positives. Kappa statistics is used to test the interrater reliability.[20] The Kappa values (0–0.20), (0.21–0.40), (0.41–0.60), (0.61–0.80), (0.81–1.00) corresponded to slight, fair, moderate, substantial, and almost perfect.[20, 21] Further, the positive and negative predictive values (PPV and NPV) were calculated as follows:

$$PPV = \frac{TP}{TP+FP} \quad NPV = \frac{TN}{TN+FN} \quad (3)$$

Results

From the 1762 lumbar levels, there were 156 debatable cases, i.e. disagreements between the radiologists, using the software, and radiologists, not using the software for the presence of central stenosis. In 18 of these lumbar cases, the consensual or predominant opinion has coincided with that of a radiologist not using the software. In 138 lumbar level cases, the former has coincided with that of a radiologist using the software CoLumbo. The average accuracy of radiologists for the presence of central spinal stenosis is shown in Figure 2 for selected age groups, gender and centers.

The measured sensitivity and specificity of the software were

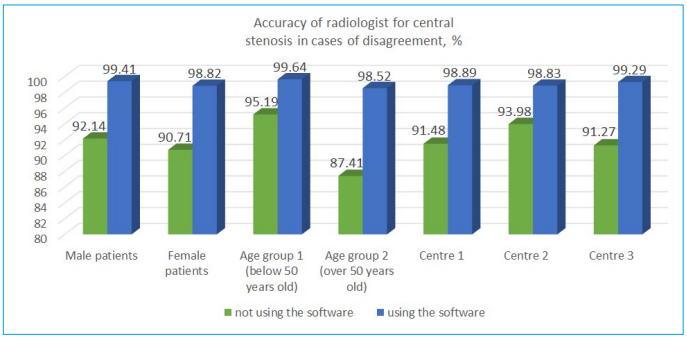


Figure 2. Comparison of the accuracy of radiologist for the presence of spinal stenosis for selected gender and age groups, as well as medical centers.

127/137 (92.70%±4.36%) and 1644/1660 (99.04%±0.47%), respectively. The average sensitivity and specificity of the software for central stenosis derived from the clinical trial are shown in Figure 3. Further, the PPV and NPV were calculated to be 88.81%±5.31% and 99.40%±0.42%, respectively.

From the studied patients, the average patient age was 49.52±13.20, from these female patients were 53.4 %, while male patients were 46.6 %. An example of detected inaccuracy of the algorithm is revealed in Figure 4. The images in the upper two rows of this figure show detected by the algorithm central stenosis with Dural Sac cross-sectional area over 100 mm^2, and measurements of naturally reduced sac at L4/L5 level, respectively. The image in the bottom row of this figure reveals a case of inaccuracy made by a radiologist not using the software, presumably due to the fact that central stenosis is axially visible only at vertebrae level.

The results from the kappa agreement analysis showed an overall interrater reliability of 92.9%, 89.9% and 73% for radiologist using CoLumbo software, CoLumbo software alone and radiologist, without using the software, respectively. The kappa agreement reveals an almost perfect agreement with the majority opinion for CoLumbo and the radiologist and the software itself.[20]

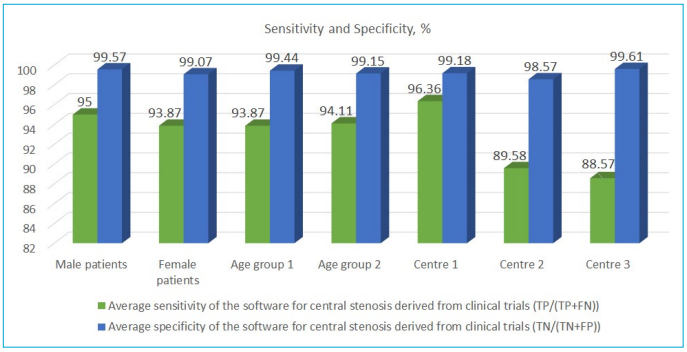


Figure 3. Comparison of the sensitivity and specificity of the software for the presence of spinal stenosis for selected gender and age groups, as well as medical centers.

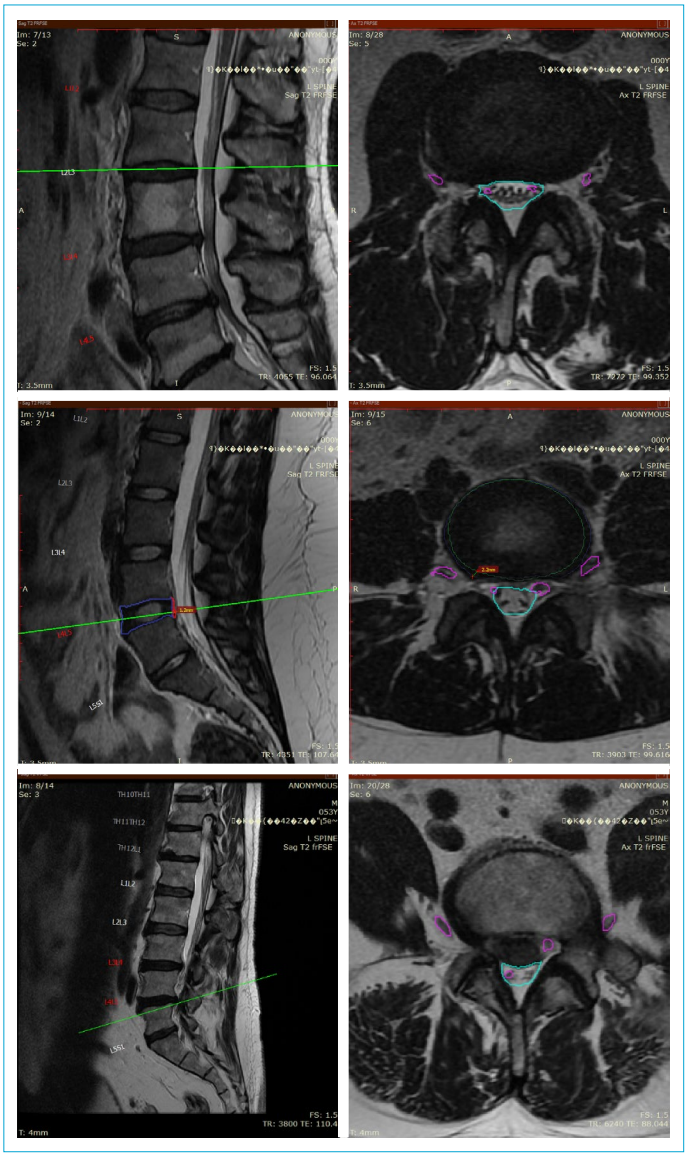


Figure 4. Software performance during measurements and segmentation.

Discussion

Magnetic resonance imaging is the gold diagnostic standard for assessment of the degree of lumbar spinal stenosis and its classification. However, MRI reading is time-consuming,[22] costly, and prone to errors.[23] In this respect, the use of software applications, such as CoLumbo, would reduce the time needed for MRI reading and reporting without decreasing the accuracy of the final report for some pathologies and improving it for others. This prospective study successively demonstrated the evaluation of the software performance, showing very good sensitivity, specificity, the positive and negative predictive values of the software. This inter-reader study also showed an excellent agreement for the radiologists, using CoLumbo versus the majority opinion, which in fact is a promising output in comparison to the lack of agreement between radiologists shown by several inter-observer studies, with kappa varying between 0.26 and 0.65.[24-26] Moreover, a recent review on AI and CAD systems used for diagnosis of low back pain demonstrated similar sensitivity, specificity, as well as accuracy, based on four AI studies related to spinal stenosis.[17] However, all of these are retrospective studies.

Reasons for the 156 debatable cases are summarized as following: (a) disagreements near the classification thresholds/borderline cases, (b) stenosis at vertebral body level (c) reduced sac due to anatomical reasons/variation and (d) stenosis at sacral level. Figure 5 shows two borderline cases. Specifically, the image in the upper row reveals a case with a Dural Sac cross sectional area of 106 mm^2, which can be considered a borderline case. This was a source of disagreement between the radiologist not using the software and the majority opinion. Such cases are a source of both interrater and intrarater disagreements. For the case, shown in the bottom row of Figure 5, the radiologist without the software reported on a lack of stenosis; the radiologist assisted by the software and the software standalone (95 mm^2) reported on a stenosis; the arbiter radiologist reported the case as stenosis.

Another case of disagreement is demonstrated in Figure 6,

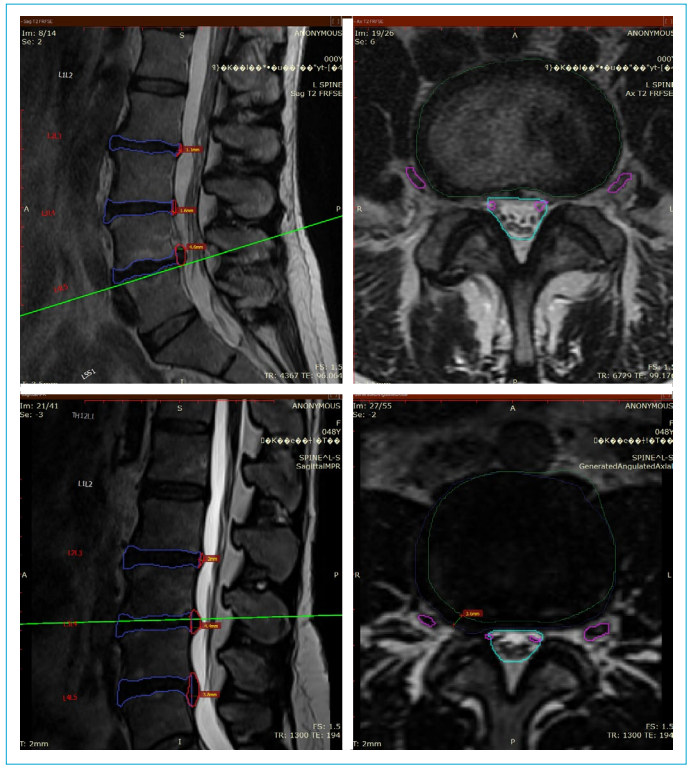


Figure 5. Borderline cases: Dural Sac cross-sectional area upper row 106 mm^2; bottom row 95 mm^2.

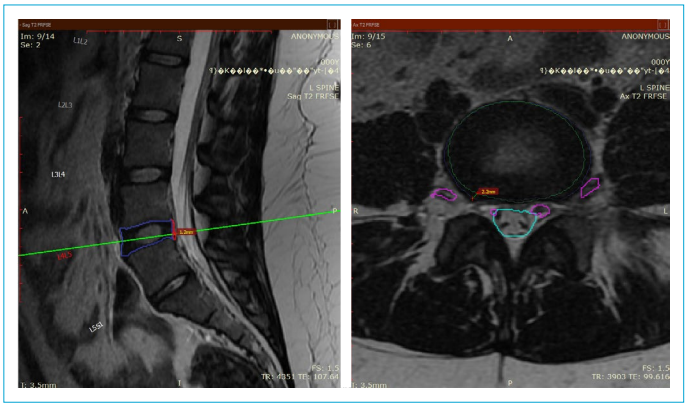


Figure 6. Naturally reduced sac at L4/L5 level.

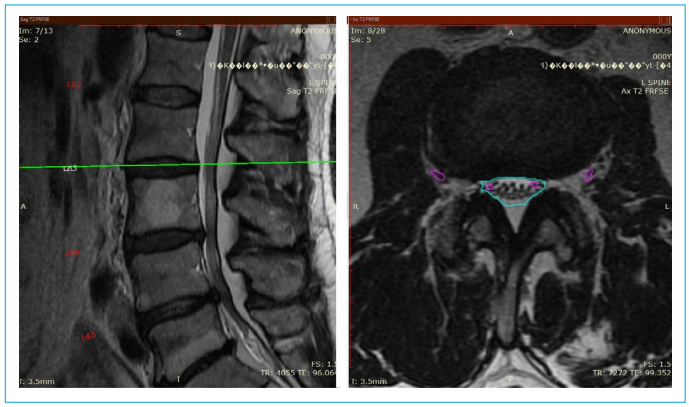


Figure 7. Central Stenosis with Dural Sac cross-sectional area over 100mm^2.

showing naturally reduced sac at L4/L5 level. In this figure, the AI algorithm reported central stenosis, based on the calculated 63 mm^2 cross-sectional area at L4/L5; however majority of radiologists disagreed. The reason could be attributed to the Dural Sac that naturally terminates more cranially. Interestingly, at L5/S1, even though the area is even smaller, the algorithm identifies that the cross-sectional area is logically to be naturally small.

Figure 7 shows central stenosis with a Dural Sac cross-sectional area over 100 mm^2. In this case a different type of disagreement is reported, the cross-sectional area is more than 100 mm^2. The unanimous opinion is that this is a case of central stenosis. The radiologist using the CoLumbo software corrected the algorithm suggestion. This particular case supports the kappa results showing qualitatively why the combination of radiologist assisted by the software gives better results than both the algorithm and the radiologist not using the software.

Finally, Figure 8 reviews central stenosis at sacral level due to epidural lipomatosis. In this case, the software does not report central stenosis at sacral level as the Dural Sac cross sectional area can naturally decrease. However, it is still possible, but the criterion should not be 100 mm^2. A physician may virtually imagine how big the sac should be at the appropriate level.

Limitation of this study. In this study, we used only images of patients, undergoing MRI examination. Other imaging modalities, such as CT scan with contrast dye as well as an electrical test of muscle activity, to validate the presence of stenosis were not used. For the debatable cases, the ground truth was also based on MRI and the arbiter is an MRI radiologist. The fact that data are from three different clinical sites alongside of Bulgaria, received from different MRI systems is also a possible source for discrepancy.

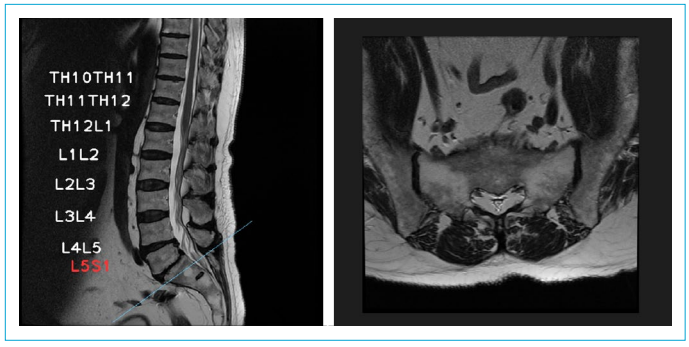


Figure 8. Central stenosis at sacral level due to epidural lipomatosis.

Conclusion

This prospective study showed that the assessment of the radiologists supported by deep learning system for central stenosis classification results in high kappa agreement. The introduction into practice of such AI-based tools would precisely predict the presence of stenosis and thus decrease the observer variability in assessing lumbar spinal stenosis severity based on MRI and its relation to cross-sectional spinal canal area. This would result in timely and effective surgical treatment and improved quality of life for these patients.

Disclosures

Ethics Committee Approval: The study "Clinical Trial with Columbo software" was approved by the Ethics Committee of the Medical University of Varna with a protocol № 108/25.11.2021. The study was also approved by the Ethical Committee for Clinical Trials of the Ministry of Healthcare of Bulgaria with a protocol ЕКК/СГ-0687 from 06.08.2020.

Acknoledgemnets: The authors would like to thank Smart Soft Healthcare (https://columbo.me/)for installing the software, as well in processing and analysis of data.

Peer-review: Externally peer-reviewed.

Conflict of Interest: None declared.

Authorship Contributions: Concept – R.G.; Design – R.G.; Supervision – Smart Soft Healthcare (https://columbo.me/); Materials – R.G., M.N.; Data collection – R.G., M.N.; Data processing – K.B., Smart Soft Healthcare; Analysis and interpretation – R.G., K.B., Smart Soft Healthcare; Literature search – K.B.; Writing – K.B.; Critical review – K.B., R.G.

References

1. Katz JN, Harris MB. Lumbar spinal stenosis. New Engl J Med 2008;358:818–25.
2. Karantanas AH, Zibis AH, Papaliaga M, Georgiou E, Rousogiannis S. Dimensions of the lumbar spinal canal: variations and correlations with somatometric parameters using CT. Eur Radiol 1998;8:1581–5.
3. Ravindra VM, Senglaub SS, Rattani A, Dewan MC, Hartl R, Bisson E, et al. Degenerative Lumbar Spine Disease: Estimating Global Incidence and Worldwide Volume. Global Spine J 2018;8:784–94.
4. Bae HW, Rajaee SS, Kanim LE. Nationwide Trends in the Surgical Management of Lumbar Spinal Stenosis. Spine 2013;38:916–26.
5. Sengupta DK. Dynamic stabilization devices in the treatment of low back pain. Orthop Clin North Am 2004;35:43–56.
6. Li AM, Li X, Yang Z. Decompression and coflex interlaminar stabilisation compared with conventional surgical procedures for lumbar spinal stenosis: A systematic review and meta-analysis. Int J Surg 2017;40:60–7.
7. Drew B, Bhandari M, Kulkarni AV, Louw D, Reddy K, Dunlop B. Reliability in grading the severity of lumbar spinal stenosis. J Spinal Disord 2000;13:253–8.
8. Lewandrowski KU. Retrospective analysis of accuracy and positive predictive value of preoperative lumbar MRI grading after successful outcome following outpatient endoscopic decompression for lumbar foraminal and lateral recess stenosis. Clin Neurol Neurosur 2019;179:74–80.
9. Al-Tameemi HN, Al-Essawi S, Shukri M, Naji FK. Using Magnetic Resonance Myelography to Improve Interobserver Agreement in the Evaluation of Lumbar Spinal Canal Stenosis and Root Compression. Asian Spine J 2017;11:198–203.
10. Amundsen T, Weber H, Lilleas F, Nordal HJ, Abdelnoor M, Magnaes B. lumbar spinal stenosis - clinical and radiologic features. Spine 1995;20:1178–86.
11. Zheng F, Farmer JC, Sandhu HS, O'Leary PF. A novel method for the quantitative evaluation of lumbar spinal stenosis. HSS J 2006;2:136–40.
12. Griffith JF, Huang J, Law S-W, Xiao F, Leung JCS, Wang D. Population reference range for developmental lumbar spinal canal size. Quant Imaging Med Surg 2016;6:671–9.
13. Amonoo-Kuofi HS. The sagittal diameter of the lumbar vertebral canal in normal adult Nigerians. J Anat 1985;140:69–78.
14. Twomey L, Taylor J. Age changes in the lumbar spinal and intervertebral canals. Paraplegia 1988;26:238–49.
15. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp 2018;2:35.
16. Roller BL, Boutin RD, O'Gara TJ, Knio ZO, Jamaludin A, Tan J, et al. Accurate prediction of lumbar microdecompression level with an automated MRI grading system. Skeletal Radiol 2021;50:69–78.
17. D'Antoni F, Russo F, Ambrosio L, Bacco L, Vollero L, Vadalà G, et al. Artificial intelligence and computer aided diagnosis in chronic low back pain: A systematic review. Int J Environ Res Public Health 2022;19:5971.
18. Georgiev N, Asenov A. Automatic segmentation of lumbar spine MRI using ensemble of 2D algorithms. In: Zheng G, Belavy D, Cai Y, Li S, editors. Computational Methods and Clinical Applications for Spine Imaging. Cham: Springer International Publishing; 2019. p. 154–62.
19. Schonstrom N, Lindahl S, Willen J, Hansson T. Dynamic changes in the dimensions of the lumbar spinal canal: an experimental study in vitro. J Orthop Res 1989;7:115–21.
20. McHugh ML. Interrater reliability: The kappa statistic. Biochem Med (Zagreb) 2012;22:276–82.
21. Yuan S, Zou Y, Li Y, Chen M, Yue Y. A clinically relevant MRI grading system for lumbar central canal stenosis. Clin Imaging 2016;40:1140–5.
22. Bhargavan M, Kaye AH, Forman HP, Sunshine JH. Workload of Radiologists in United States in 2006–2007 and Trends Since 1991–1992. Radiology 2009;252:458–67.
23. van Rijn JC, Klemetso N, Reitsma JB, Majoie CB, Hulsmans FJ, Peul WC, et al. Observer variation in MRI evaluation of patients suspected of lumbar disk herniation. AJR Am J Roentgenol 2005;184:299–303.
24. Speciale AC, Pietrobon R, Urban CW, Richardson WJ, Helms CA, Major N, et al. Observer variability in assessing lumbar spinal stenosis severity on magnetic resonance imaging and its relation to cross-sectional spinal canal area. Spine (Phila Pa 1976) 2002;27:1082–6.
25. Lurie JD, Tosteson AN, Tosteson TD, Carragee E, Carrino JA, Kaiser J, et al. Reliability of readings of magnetic resonance imaging features of lumbar spinal stenosis. Spine (Phila Pa 1976) 2008;33:1605–10.
26. Schizas C, Theumann N, Burn A, Tansey R, Wardlaw D, Smith FW, et al. Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on magnetic resonance images. Spine 2010;35:1919–24.