An Artificial Intelligence-based Support Tool for Lumbar Spinal Stenosis Diagnosis from Self-Reported History Questionnaire

Frederik Abel[a] MD[*], Eugene Garcia[b*], Vera Andreeva[c] MD, Nikolay Nikolaev[c,d] MD, Serhii Kolisnyk[e] MD PhD, Ruslan Sarbaev[b], Ivan Novikov[b] BS, Evgeniy Kozinchenko[b], Jack Kim[b] PhD, Andrej Rusakov[b] MS, Raphael Mourad PhD[f,b], Darren R. Lebl MD[a]

[a]Hospital for Special Surgery, 535 East 70th Street, New York, NY 10021, USA

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

[c]Federal State Budgetary Institution «Federal Center for Traumatology, Orthopedics and Arthroplasty» of the Ministry of Health of the Russian Federation (Cheboksary), Cheboksary, Russia, 428020

[d]Federal State Budgetary Educational Institution of Higher Education «Chuvash State University named after I.N. Ulyanov», Cheboksary, Russia, 428015

[e]Vinnitsa National Medical University, Pyrohova St, 56, Vinnytsia, Vinnytsia Oblast, Ukraine, 21018

[f]University of Toulouse, CNRS, UPS, 31062 Toulouse, France

*Co-first authors

Mail addresses:

Frederik Abel, MD: abelf@hss.edu

Eugene Garcia: eugene.garcia@remedylogic.com

Vera Andreeva, MD: Vbarieva@mail.ru

Nikolay Nikolaev, MD: nikolaevns@mail.ru

Serhii Kolisnyk, MD PhD: s.p.kolisnyk@vnmu.edu.ua

Ruslan Sarbaev: ruslan.sarbaev@remedylogic.com

Ivan Novikov: ivan.novikov@remedylogic.com

Evgeniy Kozinchenko: evgeniy.kozinchenko@remedylogic.com

Jack Kim, PhD: jack.kim@remedylogic.com

Andrej Rusakov, MS: a.rusakov@remedylogic.com

Raphael Mourad, PhD: raphael.mourad@univ-tlse3.fr

Darren R. Lebl, MD: research@leblspinemd.com

Corresponding author:

Raphael Mourad, PhD, Université Toulouse III Paul Sabatier, 118 Rte de Narbonne, Toulouse 31062, France, Email: raphael.mourad@univ-tlse3.fr

Key words:

lumbar spinal stenosis; diagnosis; artificial intelligence; machine learning; self-reported questionnaire.

Running Title:

AI-Support for Lumbar Stenosis Diagnosis

Declaration of Conflicting Interests:

Eugene Garcia, Ruslan Sarbaev, Ivan Novikov, Evgeniy Kozinchenko, Jack Kim, Andrej Rusakov, and Raphael Mourad are employees of Remedy Logic. The rest of the authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Source of Funding:

R.M. was supported by Université Paul Sabatier and Remedy Logic. S.K. was supported by Vinnitsa National Medical University and Remedy Logic. J.K. and A.R. were supported by Remedy Logic.

Abstract

Objectives

Symptomatic lumbar spinal stenosis (LSS) leads to functional impairment and pain. While radiological characterization of the morphological stenosis grade can aid in the diagnosis, it may not always correlate with patient symptoms. Artificial intelligence (AI) may diagnose symptomatic LSS in patients solely based on self-reported history questionnaires.

Methods

We evaluated multiple machine learning (ML) models to determine the likelihood of LSS using a self-reported questionnaire in patients experiencing low back pain and/or numbness in the legs. The questionnaire was built from peer-reviewed literature and a multidisciplinary panel of experts. Random forest, lasso logistic regression, support vector machine, gradient boosting trees, deep neural networks, and automated machine learning models were trained and performance metrics compared.

Results

Data from 4,827 patients (4,690 patients without LSS: mean age 62.44, range 27 – 84 years, 62.8% females, and 137 patients with LSS: mean age 50.59, range 30 – 71 years, 59.9% females) were retrospectively collected. Among the evaluated models, the random forest model demonstrated the highest predictive accuracy with an area under the receiver operating characteristic curve (AUROC) between model prediction and LSS diagnosis of 0.96, a sensitivity of 0.94, a specificity of 0.88, a balanced accuracy of 0.91 and a Cohen's kappa of 0.85.

Conclusions

Our results indicate that ML can automate the diagnosis of LSS based on self-reported questionnaires with high accuracy. Implementation of standardized and intelligence-automated workflow may serve as a supportive diagnostic tool to streamline patient management and potentially lower healthcare costs.

Introduction

Degenerative lumbar spinal stenosis (LSS) is a significant source of disability in older adults, which affects an estimated 103 million persons annually worldwide[1]. The clinical syndrome of LSS is characterized by chronic lower back and extremity pain, accompanied by loss of mobility and function, which can steadily reduce patients' quality of life[2]. First-line treatments include modification of activity, analgesia, and physical therapy. In cases where conservative treatments fail, decompressive spinal surgery is often considered as an option to relieve symptoms. As such, LSS is associated with a significant socioeconomic burden and high healthcare costs[3,4]. Hence, strategies for simplified, fast, and automated diagnosis based on clinical symptoms may have an important societal impact.

Substantial improvements of standardized diagnostic accuracy are streamlined by artificial intelligence (AI), in particular machine learning (ML) models to develop data-driven algorithms and intelligent automation (IA) to enhance human intelligence and therapeutic-decision making. Most commonly, ML models have successfully been trained to automatically detect and classify LSS based on MRI studies of the lumbar spine, achieving high accuracy levels comparable to those of subspecialist radiologists[5,6]. Other approaches applied ML to determine surgical candidacy for spinal surgery based on lumbar spine MRI's[7] or by using hybrid AI models, that combine features from both imaging and clinical information[8]. More recently, ML methods have been utilized to determine prior authorization approval for LSS surgery based on medical vignettes, which consisted of both clinical data and MRI findings[9]. Although diagnostic imaging is a mainstay for the evaluation of LSS, radiographically affected patients may be clinically asymptomatic. Therefore, clinical symptoms also play a significant role in therapeutic decision-making. Patients with symptomatic LSS can be assessed using self-reported questionnaires[10]. However, ML has not yet been tested to identify such patients based solely on patient questionnaires without any imaging data, which could potentially facilitate intelligence-automated therapeutic decision-making.

Here, we propose a novel AI approach to diagnose LSS from self-reported history questionnaires that assess clinical history, pain character, and mobility of patients. Different ML models were trained on retrospectively collected data from patients with diagnosed LSS (LSS+) and without LSS (LSS-) and their performance compared, including random forest, lasso logistic regression, support vector machine (SVM), gradient boosting trees, deep neural networks (DNN), and automated machine learning (H2O autoML). Finally, key contributor variables for predicting patients with LSS+ were identified.

Materials and Methods

Subjects and Data collection

This multicenter study was performed in two hospitals (BLINDED FOR REVIEW and BLINDED FOR REVIEW), and one outpatient clinic (BLINDED FOR REVIEW). Data of patients were retrospectively collected from self-reported questionnaires from August 2021 to September 2022. Additionally, health records were assessed, including past clinical history, treatments, and results of examinations, for all patients. The selection was limited to patients who presented with primary symptoms of low back pain and/or numbness in the back and legs, and experienced difficulties in performing daily activities. The age criteria for inclusion were set at 20 years or older. The questionnaires were filled out by our clinical administrators based on the answers provided by each interviewed patient. LSS diagnosis (presence/absence) was confirmed by an orthopedic surgeon based on clinical history of each patient and reports from lumbar spine MRI studies, which served as the ground truth.

This retrospective study received institutional review board approval and written informed consent was obtained from all subjects.

Questionnaire

A literature review was performed to identify peer-reviewed medical literature that assess diagnosis and outcome measures of LSS (Supplementary material 1). Items from these articles, as well as relevant items extracted from The Short Form (36) Health Survey (SF-36)[11], EQ-5D[12], and Oswestry Disability Index (ODI)[13], were collected. In total, 205 questions were accumulated, which were then compiled with the input of an expert panel comprising a multidisciplinary team of doctors in the fields of spinal surgery, rehabilitation medicine, interventional and diagnostic radiology. The final self-reported questionnaire included 26 questions (qualitative and continuous outcome variables) including pain, pain severity and type, activities prevented by pain (e.g. pain prevents sleeping rate), but also motor impairment, the use of moving device, history of spinal cord or cauda equina injury, general health, and mental health (e.g, anxiety or depression level) **(Table 1)**.

Data Imputation and Machine Learning

The dataset comprised 4,690 patients without LSS (LSS-; mean age 62.44; range 27 – 84 years; 62.8% females) and 137 patients with LSS (LSS+; mean age 50.59; range 30 – 71 years: 59.9% females), summing up to 4,827 patients in total. The overall percentage of missing data was 29%. To cope with missing data, values were imputed applying the median, where necessary. Subsequently, the data were used to train the machine learning models. Patient data were randomly split into 80% for fine-tuning and training of the machine learning models (3,758 LSS- and 104 LSS+) and 20% for testing predictions (932 LSS- and 33 LSS+). To ensure a balanced training dataset for effective model training, patients without LSS were down-sampled to match the number of patients with LSS, resulting in a total of 208 patients for training (104 LSS- and 104 LSS+). Different machine learning models were trained and compared, including random forest, lasso logistic regression, support vector machine (SVM), gradient boosting trees (XGBOOST), deep neural network (DNN), and automated machine learning (H2O autoML). Variable importance was computed using the mean decrease in accuracy in the out-of-bag sample during training. Hyper-parameters were obtained by fine-tuning with 5-fold cross-validation.

Data and Statistical Analysis

All data analyses, including univariate and bivariate analyses, prediction performance metrics, and plots were done using R (Version 4.2.1, the R Foundation). The following R packages were used for computations and fine-tuning: ranger for random forest and variable importance (https://cran.r-project.org/web/packages/ranger), tuneRanger for hyper-parameter fine-tuning (https://cran.r-project.org/web/packages/tuneRanger/), glmnet for lasso logistic regression (https://cran.r-project.org/web/packages/glmnet), e1071 for SVM (https://cran.r-project.org/web/packages/e1071), XGBOOST for extreme gradient boosting (https://cran.r-project.org/web/packages/xgboost), nnet for deep neural network (https://cran.r-project.org/web/packages/nnet/), and caret for SVM, XGBOOST, and DNN fine-tuning (https://topepo.github.io/caret/). For quantitative variables, differences of means between the LSS+ patients and LSS- patients were tested using the Student's test. For qualitative variables, differences of proportions between the LSS+ patients and LSS- patients were assessed using the Fisher's exact test. To account for multiple tests, the Bonferroni p-value threshold was used and computed as 0.05/34=0.0014. Performance metrics, including area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), sensitivity, specificity, Cohen's kappa, accuracy, and the F1-score, were calculated for all models.

Results

Univariate analysis of LSS predictors

Table 2 displays the differences in assessed outcome variables between patients with LSS (LSS+) and patients without LSS (LSS-). Most predictors exhibited significant p-values, with $p<0.0014$ (Bonferroni threshold). It included predictors describing general health (mean difference of -18 points out of 100, 95% confidence interval (CI) [-20.71; -15.45], $p<0.0001$), mental health (mean difference of -17 points out of 100, 95%CI [-19.98; -14.34], $p<0.0001$), pain severity (mean difference of +25.37 points out of 100, 95%CI [22.64; 28.11], $p<0.0001$), or pain preventing activities, e.g. pain preventing standing (mean difference of +35.7 points out of 100, 95%CI [31.41; 39.99], $p<0.0001$). Additionally, reduced or damaged motor skills, problems with performing daily activities (i.e. problems performing washing or dressing), or use of moving devices were significantly different between LSS+ and LSS- patients. Thus, univariate analysis demonstrated strong associations between LSS and most self-reported history predictors, suggesting their potential use in building a ML model for accurate prediction of LSS.

Prediction of LSS based on Machine Learning

The accuracy of our ML approach in predicting LSS was assessed and compared with the ground truth. For this purpose, the patient history data were randomly split into 80% of patients for training a series of ML models, while the remaining 20% of patients were reserved to estimate the prediction accuracy of the different models. The performance of the random forest, lasso regression, SVM, XGBOOST, DNN, and automated machine learning models is summarized in **Table 3**.

Most ML models showed excellent prediction performances when classifying LSS+ versus LSS- patients. Among the models, the random forest exhibited the highest prediction performance, achieving an AUROC of 0.96 (95% CI [0.949; 0.980]), sensitivity of 0.94, specificity of 0.88, and Cohen's kappa value of 0.85 (**Figure 1A**). The second-best performing model was XGBOOST, which demonstrated an AUROC of 0.96, a sensitivity of 0.97, a specificity of 0.86, and a Cohen's kappa value of 0.88 (**Figure 1D**).

Since the data were highly imbalanced (4,690 LSS+ patients and 137 LSS- patients), we also computed the balanced accuracy, a more suitable metric for imbalanced data. The random forest also showed a high balanced accuracy of 0.91, while the XGBOOST model achieved the highest value of 0.92. Our random forest was chosen as the optimal trade-off between a high AUROC and balanced accuracy for predicting LSS from self-reported history data, demonstrating excellent predictive performance metrics.

Importance of predictors

Next, we conducted an assessment to identify the key predictors of LSS. For this purpose, we computed variable importance using the random forest model to identify the most significant predictors **(Figure 2).** Among the list of predictors, problems with performing daily activities, including washing or dressing, were found to be the most significant predictors for LSS+ patients. Additionally, pain severity, and pain or emotional distress that restricts social interactions and/or activities were significant contributors to patients with LSS.

Discussion

Applications of artificial intelligence (AI), including machine learning (ML) models and intelligent automation (IA), are increasing rapidly in the medical domain and have demonstrated remarkable success in various clinical settings and research areas. AI systems strive to improve diagnostic processes, prognostication, and outcomes in a transparent and observer-independent manner, thereby enhancing therapeutic decision-making. The field of spine surgery, in particular, benefits from AI applications and IA tools, as diagnostics and therapeutic decision-making often require clinical expertise and rely on interpretation of ramified factors such as medical history, imaging, or perioperative data[14].

In this article, we tested a series of ML models to predict symptomatic LSS based on a simple self-reported history questionnaire of patients with low back pain and/or numbness in the back and legs, and having problems performing daily activities. Data from 4,827 patients (4,690 LSS- and 137 LSS+) were collected and key factors identified that revealed strong association with LSS+ patients. Subsequently, different ML models were trained and evaluated on a balanced subset of the self-reported predictors, including random forest, lasso logistic regression, support vector machine (SVM), gradient boosting trees, deep neural networks (DNN), and H2O automated machine learning. Of these, the random forest model demonstrated the highest diagnostic accuracy with a prediction error as measured by the AUROC of 0.96, a sensitivity of 0.94, a specificity of 0.88, a balanced accuracy of 0.91, and a Cohen's kappa of 0.85. The computation of variable importance revealed that problems with performing daily activities, pain severity and emotional distress that restrict social interaction or activity rate as were the most significant contributors to patients with LSS+. As such, self-reported questionnaires may be feasible to predict symptomatic LSS patients in an IA-based manner, ultimately enhancing human therapeutic decision-making.

Previous studies have utilized AI applications in spine surgery, with a focus on radiological features extracted from MRI data and employing various ML models to detect and diagnose LSS[5,6,15]. While radiological characterization of morphological stenosis grade contributes to the diagnosis of LSS, it may not always correlate with pain intensity and functional disability experienced by affected patients[16,17]. Consequently, incorporating AI-guided diagnosis of LSS based on other aspects or in combination with radiological features may be a more comprehensive approach to reflect the clinical syndrome of LSS. However, such approaches have been investigated in only a limited number of studies. Ren *et al.* investigated natural language processing-based ML models based on positive symptoms extracted from electronic health records[18]. Contrasting our study, different models were tested to discriminate patients with LSS from patients with lumbar disc herniation, and found that a Long Short-Term Memory DNN achieved the highest capacity with an AUROC of 0.85 for this task. Another approach used ML algorithms on data of patients that performed five-repetition sit-to-stand tests[19]. The algorithm, a fuzzy rule-based system, achieved a classification accuracy of 96.2% for patients with disc herniation, LSS, and chronic lower back pain. Five-repetition sit-to-stand tests are designed to assess functional impairment, which has been identified as the most significant contributor variable to LSS+ patients in our dataset, followed by pain and emotional distress that restrict social interaction or activity rate.

Accordingly, these predictors have been determined as critical factors for treatment and decision-making in LSS+ patients by previous studies. The presence of disability, along with pain and radiological stenosis grade, has been associated with the likelihood of requiring surgical therapy[20]. Similarly, functional disability and pain severity have been correlated with impairment of health-related quality of life in patients with LSS or lumbar disc herniation[21]. Pain severity has been as identified as the fourth most important contributor to LSS patients in our dataset, suggesting that actual restriction of daily activities may be more significant for the diagnosis of LSS patients. However, it should be considered that pain and functional impairment are closely related, and the inability to perform daily activities likely is a secondary effect of pain. Another significant predictive variable for the diagnosis of LSS was the patient’s mental status, evaluated through our questionnaire based on the patient’s subjective grading of their mental health, encompassing a broader field of psychological factors including depression or anxiety. Prior studies have indicated that these factors, particularly preoperative depression, are associated with increased severity of postoperative LSS-related symptoms and poorer long-term outcomes following decompression[22,23]. Therefore, although it can be challenging to assess in clinical practice, considering the patient's overall mental health could significantly contribute to a more comprehensive management approach and improve the patient's prognosis.

Our study contains several limitations. First, the dataset was comprised 4690 LSS- patients and only 137 LSS+ patients, which required down-sampling to obtain a more balanced sample size. Second, the ground truth of LSS+ diagnosis in this cohort was based on clinical history of each patient and reports from MRI examinations and did not include independent classification by other experts. Third, our ML models solely focused on diagnosis of LSS on basis of self-reported symptoms in questionnaires. Symptoms of LSS may partially overlap those of other concomitant degenerative spinal disorders (e.g. degenerative disc disease, spondylolisthesis) in these patients and differentiation of these entities was not addressed by our models. Additionally, other factors relevant for management of LSS were not investigated, including surgical decision-making or impairment of patients' quality of life. Fourth, our algorithms were not validated in an external patient cohort within this study, potentially limiting their generalizability. Finally, our proposed approach is simplified and did not integrate other features, such as radiological stenosis grade, that may have further increased diagnostic accuracy in detecting LSS patients.

Conclusions

In summary, our results demonstrate that AI can be applied to diagnose symptomatic LSS in patients based on simplified, self-reported history questionnaires with high accuracy even in the absence of any imaging input into the model. Functional impairment and pain/emotional distress that restrict social interaction or activity rate are key contributors to patients with LSS. Implementation of standardized and automated AI-guided workflow may act as an intelligent automation tool to identify patients with LSS using simple self-reported history questionnaires and may more efficiently and cost-effectively help determine which patients required advanced imaging studies such as MRI and consideration for surgery.

Acknowledgements

N/A

References

1. Ravindra VM, Senglaub SS, Rattani A, et al. Degenerative Lumbar Spine Disease: Estimating Global Incidence and Worldwide Volume. *Global Spine J*. 2018;8(8):784-794. doi:10.1177/2192568218770769

2. Otani K, Kikuchi S, Yabuki S, et al. Lumbar Spinal Stenosis Has a Negative Impact on Quality of Life Compared with Other Comorbidities: An Epidemiological Cross-Sectional Study of 1862 Community-Dwelling Individuals. *Sci World J*. 2013;2013:590652. doi:10.1155/2013/590652

3. Ong KL, Auerbach JD, Lau E, Schmier J, Ochoa JA. Perioperative outcomes, complications, and costs associated with lumbar spinal fusion in older patients with spinal stenosis and spondylolisthesis. *Neurosurg Focus*. 2014;36(6):E5. doi:10.3171/2014.4.focus1440

4. Machado GC, Maher CG, Ferreira PH, et al. Trends, Complications, and Costs for Hospital Admission and Surgery for Lumbar Spinal Stenosis. *Spine*. 2017;42(22):1737-1743. doi:10.1097/brs.0000000000002207

5. Hallinan JTPD, Zhu L, Yang K, et al. Deep Learning Model for Automated Detection and Classification of Central Canal, Lateral Recess, and Neural Foraminal Stenosis at Lumbar Spine MRI. *Radiology*. 2021;300(1):130-138. doi:10.1148/radiol.2021204289

6. Jamaludin A, Kadir T, Zisserman A. SpineNet: Automated classification and evidence visualization in spinal MRIs. *Med Image Anal*. 2017;41:63-73. doi:10.1016/j.media.2017.07.002

7. Wilson B, Gaonkar B, Yoo B, et al. Predicting Spinal Surgery Candidacy From Imaging Data Using Machine Learning. *Neurosurgery*. 2021;89(1):116-121. doi:10.1093/neuros/nyab085

8. Mourad R, Kolisnyk S, Baiun Y, et al. Performance of hybrid artificial intelligence in determining candidacy for lumbar stenosis surgery. *Eur Spine J*. 2022;31(8):2149-2155. doi:10.1007/s00586-022-07307-7

9. Barros AD, Abel F, Kolisnyk S, et al. Determining Prior Authorization Approval for Lumbar Stenosis Surgery With Machine Learning. *Global Spine J*. Published online 2023:219256822311558. doi:10.1177/21925682231155844

10. Aizawa T, Tanaka Y, Yokoyama T, et al. New diagnostic support tool for patients with leg symptoms caused by lumbar spinal stenosis and lumbar intervertebral disc herniation: A self-administered, self-reported history questionnaire. *J Orthop Sci*. 2016;21(5):579-585. doi:10.1016/j.jos.2016.07.012

11. Ware JE, Sherbourne CD. The MOS 36-ltem Short-Form Health Survey (SF-36). *Med Care*. 1992;30(6):473-483. doi:10.1097/00005650-199206000-00002

12. Rabin R, Charro F de. EQ-SD: a measure of health status from the EuroQol Group. *Ann Med*. 2001;33(5):337-343. doi:10.3109/07853890109002087

13. Fairbank JCT, Pynsent PB. The Oswestry Disability Index. *Spine*. 2000;25(22):2940-2953. doi:10.1097/00007632-200011150-00017

14. Joshi RS, Lau D, Ames CP. Artificial Intelligence and the Future of Spine Surgery. *Neurospine*. 2019;16(4):637-639. doi:10.14245/ns.1938410.205

15. Su ZH, Liu J, Yang MS, et al. Automatic Grading of Disc Herniation, Central Canal Stenosis and Nerve Roots Compression in Lumbar Magnetic Resonance Image Diagnosis. *Front Endocrinol*. 2022;13:890371. doi:10.3389/fendo.2022.890371

16. Andrasinova T, Adamova B, Buskova J, Kerkovsky M, Jarkovsky J, Bednarik J. Is there a Correlation Between Degree of Radiologic Lumbar Spinal Stenosis and its Clinical Manifestation? *Clin Spine Surg*. 2018;31(8):E403-E408. doi:10.1097/bsd.0000000000000681

17. Weber C, Giannadakis C, Rao V, et al. Is There an Association Between Radiological Severity of Lumbar Spinal Stenosis and Disability, Pain, or Surgical Outcome? *Spine*. 2016;41(2):E78-E83. doi:10.1097/brs.0000000000001166

18. Ren G, Yu K, Xie Z, et al. Differentiation of lumbar disc herniation and lumbar spinal stenosis using natural language processing–based machine learning based on positive symptoms. *Neurosurg Focus*. 2022;52(4):E7. doi:10.3171/2022.1.focus21561

19. Staartjes VE, Quddusi A, Klukowska AM, Schröder ML. Initial classification of low back and leg pain based on objective functional testing: a pilot study of machine learning applied to diagnostics. *Eur Spine J*. 2020;29(7):1702-1708. doi:10.1007/s00586-020-06343-5

20. Kurd MF, Lurie JD, Zhao W, et al. Predictors of Treatment Choice in Lumbar Spinal Stenosis. *Spine*. 2012;37(19):1702-1707. doi:10.1097/brs.0b013e3182541955

21. Gautschi OP, Smoll NR, Joswig H, et al. Influence of age on pain intensity, functional impairment and health-related quality of life before and after surgery for lumbar degenerative disc disease. *Clin Neurol Neurosur*. 2016;150:33-39. doi:10.1016/j.clineuro.2016.08.024

22. McKillop AB, Carroll LJ, Battié MC. Depression as a prognostic factor of lumbar spinal stenosis: a systematic review. *Spine J*. 2014;14(5):837-846. doi:10.1016/j.spinee.2013.09.052

23. Tuomainen I, Pakarinen M, Aalto T, et al. Depression is associated with the long-term outcome of lumbar spinal stenosis surgery: a 10-year follow-up study. *Spine J*. 2018;18(3):458-463. doi:10.1016/j.spinee.2017.08.228

Figure Legends

Figure 1: Prediction performance of lumbar spinal stenosis (LSS) based on self-reported history questionnaires for several machine learning models. Receiver operating characteristic curve (ROC) of the random forest (A), lasso logistic regression (B), support vector machine (SVM, C), XGBOOST (D), deep neural network (DNN, E), and automated machine learning (H2O autoML, F).

Figure 2: Top key variables to predict lumbar spinal stenosis probability. Variable importance was calculated using random forests with permutations.


A
Random forest
AUROC:0.964
CI95%:0.949−0.98
B
Lasso regression
AUROC:0.946
CI95%:0.925−0.967
C
SVM
AUROC:0.913
CI95%:0.878−0.948
D
XGBOOST
AUROC:0.958
CI95%:0.942−0.975
E
DNN
AUROC:0.927
CI95%:0.899−0.955
F
H2O autoML
AUROC:0.956
CI95%:0.938−0.974
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0


Variable Importance
Problems performing daily activities
Problems washing and dressing yourself
Pain severity
Pain or emotional distress restricts social interactions and/or activities
Pain prevents lifting weight
Energy level
Pain prevents standing
Mental health
Pain prevents walking
Problems with strenuous physical activities
General health
Pain prevents sitting
Pain prevents sleeping
Walking causes more severity
Motor skills reduced or damaged
Moving devices
Pain when lifting right leg
Pain type
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07

Tables

Table 1: Self-reported questionnaire with 26 questions to assess manifestation of symptomatic LSS.

Question	Value
1. Lumbar spinal stenosis	Yes/No
2. General health (0 – Worst; 100 – Best general health imaginable)	0 – 100
3. Mental health (0 – Worst; 100 – Best mental health imaginable)	0 – 100
4. Congenital diseases or conditions	Yes/No
5. Spinal cord or cauda equina injury	Yes/No
6. Pain severity (0 – Zero pain; 100 – Extreme pain)	0 – 100
7. Pain type	
Extreme cold	Yes/No
Extreme hot or burning sensation	Yes/No
Itching	Yes/No
Mechanical	Yes/No
Sharp stabbing	Yes/No
Throbbing and/or pulsating	Yes/No
8. Bending forward increases pain severity	Yes/No
9. Standing up increases pain severity	Yes/No
10. Walking increases pain severity	Yes/No
11. Pain increases while sleeping	Yes/No
12. Pain prevents lifting weights (0 – Zero pain; 100 – Extreme pain)	0 – 100
13. Pain prevents sitting (0 – Zero pain; 100 – Extreme pain)	0 – 100
14. Pain prevents sleeping (0 – Zero pain; 100 – Extreme pain)	0 – 100
15. Pain prevents standing (0 – Zero pain; 100 – Extreme pain)	0 – 100
16. Pain prevents walking (0 – Zero pain; 100 – Extreme pain)	0 – 100
17. Daily activities limited by physical pain	Yes/No
18. Pain when lifting right leg:	
- No	Yes/No
- Back pain only	Yes/No
- Pain radiating down the leg	Yes/No
19. Pain when lifting left leg:	
- No	Yes/No
- Back pain only	Yes/No
- Pain radiating down the leg	Yes/No
20. Motor skills reduced or damaged	Yes/No
21. Moving devices	Yes/No
22. Problems performing daily activities (0 – Zero problems; 100 – Extreme problems)	0 – 100
23. Problems washing and dressing yourself (0 – Zero problems; 100 – Extreme problems)	0 – 100
24. Problems with strenuous physical activities (0 – Zero problems; 100 – Extreme problems)	0 – 100
25. Energy level (0 – No energy at all; 100 – Highest energy level imaginable)	0 – 100
26. Pain or emotional distress (0 – Zero; 100 – Extreme pain/emotional distress)	0 – 100

Table 2: Univariate analyses of predictors associated with the diagnosis of LSS.

Variable	LSS- (n=4644)	LSS+ (n=137)	OR	Mean difference	95%CI	*p*-value
General health	57.6	39.51		-18	[-20.71; -15.45]	**<0.0001**
Mental health	66.83	49.67		-17	[-19.98; -14.34]	**<0.0001**
Congenital diseases/conditions	11.68%	36.50%	4.34		[2.97; 6.30]	**<0.0001**
Spinal cord or cauda equina injury	14.71%	50.36%	5.88		[4.10; 8.43]	**<0.0001**
Pain severity	46.09	71.46		25.37	[22.64; 28.11]	**<0.0001**
Pain type						
- Extreme cold	0.23%	0.73%	3.08		[0.07; 21.51]	0.2959
- Extreme hot or burning sensation	7.06%	15.33%	2.37		[1.39; 3.87]	**0.0012**
- Itching	0.58%	0.73%	1.25		[0.03; 7.72]	0.5596
- Mechanical	18.42%	14.60%	0.74		[0.43; 1.21]	0.2604
- Sharp stabbing	18.72%	27.74%	1.66		[1.09; 2.47]	0.0141
- Throbbing and/or pulsating	6.20%	5.11%	0.8		[0.31; 1.72]	0.7198
Bending forward increases pain severity	18.61%	32.12%	2.07		[1.40; 3.02]	**<0.0001**
Standing up increases pain severity	10.13%	27.74%	3.4		[2.25; 5.06]	**<0.0001**
Walking increases pain severity	14.05%	48.18%	5.68		[3.96; 8.15]	**<0.0001**
Pain increases while sleeping	22.37%	48.18%	3.34		[2.30; 4.84]	**<0.0001**
Pain prevents lifting weights	42.36	75.41		33.05	[30.08; 36.03]	**<0.0001**
Pain prevents sitting	32.53	57.15		24.62	[20.76; 28.48]	**<0.0001**
Pain prevents sleeping	26.68	51.69		25.01	[20.22; 29.81]	**<0.0001**
Pain prevents standing	33.92	69.62		35.7	[31.41; 39.99]	**<0.0001**
Pain prevents walking	35.00	69.64		34.64	[30.39; 38.89]	**<0.0001**
Daily activities limited by physical pain	82.20%	98.54%	14.62		[3.95; 122.34]	**<0.0001**
Pain when lifting right leg:						
- No	69.46%	45.26%	0.36		[0.25; 0.52]	**<0.0001**
- Back pain only	10.21%	18.25%	1.96		[1.20; 3.08]	0.0044
- Pain radiating down the leg	4.71%	15.33%	3.66		[2.14; 5.99]	**<0.0001**
Pain when lifting left leg:						
- No	68.14%	43.07%	0.35		[0.25; 0.50]	**<0.0001**
- Back pain only	10.77%	19.71%	2.03		[1.27; 3.16]	0.0021
- Pain radiating down the leg	4.78%	15.33%	3.61		[2.11; 5.91]	**<0.0001**
Motor skills reduced or damaged	57.38%	93.43%	10.56		[5.37; 23.68]	**<0.0001**
Moving devices	5.48%	34.31%	9		[6.04; 13.26]	**<0.0001**
Problems performing daily activities	25.94	59.12		33.18	[30.20; 36.17]	**<0.0001**
Problems washing and dressing yourself	20.77	42.54		21.77	[19.23; 24.32]	**<0.0001**
Problems with strenuous physical activities	34.48	61.54		27.06	[23.13; 30.00]	**<0.0001**
Energy level	61.49	35.56		-25.93	[-29.06; -22.81]	**<0.0001**
Pain or emotional distress	24.12	59.80		35.68	[31.67; 39.68]	**<0.0001**

95%CI = 95% confidence interval; LSS = Lumbar spinal stenosis; OR = Odds ratio.

Table 3: Table of predictive performance metrics for the different models.

Model	Sensitivity	Specificity	Kappa	AUROC	Accuracy	F1	Balanced Accuracy	AUPRC
RF	0.9394	0.8777	0.8485	0.9645	0.8798	0.9338	0.9085	0.8943
Lasso regression	0.8788	0.8809	0.8485	0.9461	0.8808	0.9345	0.8798	0.9008
SVM	0.8485	0.8273	0.7576	0.9129	0.828	0.9028	0.8379	0.9062
XGBOOST	0.9697	0.8627	0.8788	0.9582	0.8663	0.9257	0.9162	0.8971
DNN	0.8788	0.838	0.7576	0.9268	0.8394	0.9097	0.8584	0.9035
H2O autoML	0.9394	0.8691	0.8485	0.9556	0.8715	0.9289	0.9042	0.896

AUPR = Area under precision-recall curve; AUROC = Area under the receiver operating characteristic; DNN = Deep neural network; RF = Random forest; SVM = Support vector machine.