Retrospective Study

Ligamentum Flavum Average Thickness: A Novel MRI Parameter for Diagnosing Lumbar Central Spinal Stenosis

Yu Wang, MD, Xuanhang Zhang, MD, Jiahao Xie, MD, Kaifeng Liu, MD, Xinyu Liu, MD, Kunqi Li, MD, Jian Xiong, MD, Jianying Zheng, MD, Panjie Xu, MD, Zhongmin Zhang, MD, PhD, and Wei Ji, MD, PhD

From: Division of Spine Surgery, Department of Orthopedics, Nanfang Hospital, Southern Medical University, Guangzhou, China

Address Correspondence: Wei Ji, MD, PhD
Division of Spine Surgery, Department of Orthopedics, Nanfang Hospital, Southern Medical University, Guangzhou, China
E-mail: spineji@126.com

Disclaimer: This study was approved by the Ethic Committee of Nanfang Hospital. This study was sponsored by the National Natural Science Foundation of China (No.82172523) and Natural Science Foundation of Guangdong Province (No. 2022A1515010488). The work was performed in the Division of Spinal Surgery, Department of Orthopaedics, Nanfang Hospital, Southern Medical University, 1838 North Guangzhou Avenue, Guangzhou, China.

Conflict of interest: Each author certifies that he or she, or a member of his or her immediate family, has no commercial association (i.e., consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted article.

Article received: 06-20-2025
Revised article received: 01-30-2026
Accepted for publication: 02-04-2026

Free full article: www.painphysicianjournal.com

Background: Hypertrophy of the ligamentum flavum (LF) is a primary cause of lumbar central spinal stenosis (LCSS). Conventional metrics of LF hypertrophy, however, have limitations. Using ligamentum flavum thickness (LFT), measured at the midpoint of the thicker side of the LF, can miss heterogeneous thickening, while the metric known as the ligamentum flavum area (LFA) is confounded by anatomical variations in LF length. To overcome these limitations, we introduced and validated a novel parameter, ligamentum flavum average thickness (LFAT).

Objectives: To evaluate the diagnostic performance of LFAT for LCSS and compare the metric with conventional MRI parameters.

Study Design: A retrospective observational study.

Setting: A single center in Guangzhou, China.

Methods: This retrospective study analyzed T2-weighted axial MR images from 166 patients with LCSS and 168 asymptomatic controls. Using receiver operating characteristic (ROC) curve analysis, we compared the diagnostic performance of the novel parameter, LFAT—the more hypertrophied side of the ligamentum flavum area (LFA') divided by its curvilinear length—with 3 conventional metrics: LFT—the midpoint thickness of the more hypertrophied LF side, LFA—the total cross-sectional area of the LF, and LFA'.

Results: ROC analysis revealed that of all the parameters, the LFAT demonstrated the highest diagnostic accuracy, with an area under the curve (AUC) of 0.909 (95% CI, 0.879–0.940). This performance was significantly superior to that of LFT (AUC = 0.852), LFA' (AUC = 0.851), and LFA (AUC = 0.782) (all $P < 0.05$). At an optimal cutoff of 3.35 mm, LFAT yielded a sensitivity of 86.7% and a specificity of 81.5%.

Limitations: Major limitations of this study include its retrospective, single-center design, the use of a control group comprising patients with hip pathologies rather than healthy volunteers, and an exclusive focus on hypertrophy of the ligamentum flavum (LFH) as a contributor to LCSS.

Conclusion: With its superior diagnostic accuracy over traditional metrics, LFAT shows considerable promise as a more objective and reliable parameter for the clinical diagnosis of LCSS.

Key words: Ligamentum flavum average thickness, ligamentum flavum thickness, ligamentum flavum area, ligamentum flavum hypertrophy, lumbar central spinal stenosis, receiver operating characteristic analysis

Pain Physician 2026: 29:E247-E255

Lumbar central spinal stenosis (LCSS), a degenerative condition of the spine, manifests primarily in older adults, reaching a prevalence of 23.6% in this demographic (1-4). LCSS is a clinical syndrome caused by the narrowing of the spinal canal, which leads to compression of the contained neural elements, primarily the cauda equina and nerve roots. This compression typically manifests as lower back pain, radiating pain into the lower limbs, and, most characteristically, neurogenic claudication (3,5). The severity of these symptoms can profoundly impact patients' quality of life. Although LCSS arises from a combination of factors, hypertrophy of the ligamentum flavum (LFH) stands out as the most critical element (6). Previous research has established that ligamentum flavum thickness (LFT) is influenced by such factors as age, inflammation, and mechanical stress (4,7). Elucidating the pathophysiological mechanisms of LFH is paramount for developing effective treatments and mitigating the debilitating symptoms of LCSS. A prerequisite for all related research and clinical work, however, is the accurate measurement of ligamentum flavum (LF) thickness. The conventional method, LFT, measures the midpoint thickness on the more hypertrophied side of the LF (4,8). This approach, however, suffers from a significant drawback: by relying on a single-point measurement, it may fail to represent the LF's overall thickness, especially because LFH is typically a heterogeneous and uneven process. Figs. 1A and 1B display 2 patients with nearly identical LFT values (approximately 4.3 mm). However, Patient A exhibits marked heterogeneous thickening, in which the LF near the base of the spinous process is significantly less than the measured LFT. In contrast, Patient B demonstrates diffuse, uniform hypertrophy. Visually, the overall hypertrophic burden in Patient B appears significantly more severe than in Patient A, yet LFT fails to distinguish between these 2 distinct morphological patterns. To address this limitation, Kim et al (9) introduced a novel morphological parameter in 2017: ligamentum flavum area (LFA), which quantifies the entire cross-sectional area of the LF at the L4-L5 level. While LFA provides a more comprehensive assessment than LFT does, LFA is beset by 2 significant limitations. First, due to the frequent discontinuity and asymmetric hypertrophy of the left and right LF, an LFA value cannot precisely represent the severity of hypertrophy on a single side. Fig. 1E illustrates a representative case demonstrating that the LF can manifest significant asymmetry between the left and right sides. Consequently, a measurement method using a total area would artificially dilute the data. Second, and more critically, LFA is confounded by interindividual variability in LF cross-sectional length, a factor likely tied to patient-specific anatomy and body size. For example, the LFT of the patient in Fig. 1C (4.32 mm) is considerably smaller than that in Fig. 1D (5.36 mm); however, the LFA of the former (180.5 mm^2) is significantly larger than that of the latter (151.8 mm^2). The larger LFA in the former is driven primarily by a longer cross-sectional curvilinear length, demonstrating that a larger area does not necessarily imply more severe pathological hypertrophy. Consequently, these variations undermine the reliability of using area alone as a true indicator of LF thickness.

In our own attempt to address the aforementioned limitations, we developed a novel parameter: ligamentum flavum average thickness (LFAT). This metric is calculated by dividing the more hypertrophied side of the ligamentum flavum area (LFA') at the L4-L5 level by its corresponding cross-sectional length. By design, LFAT quantifies the average LF thickness, offering a more comprehensive assessment than the single-point LFT while remaining unaffected by the interindividual variations in LF length that confound LFA. We therefore hypothesized that LFAT would provide a more accurate quantification of LFT than traditional parameters, which in turn would translate to a superior performance in diagnosing LCSS.

Methods

Patients

This retrospective study was approved by the Nanfang Hospital Institutional Review Board. We reviewed patients diagnosed with LCSS between 2020 and 2024. The inclusion criteria for the patient group were as follows: (1) age 60 years or older; (2) presence of clinical symptoms and signs consistent with LCSS, such as lower back pain or neurogenic claudication; (3) availability of MR images of the lumbar spine obtained within 12 months of the LCSS diagnosis; and (4) the most severe stenosis located at the L4-L5 level. Patients with a history of spinal trauma, prior lumbar surgery, or other spinal interventions such as kyphoplasty or vertebroplasty were excluded. For comparison, the control group consisted of age-matched individuals who had undergone pelvic or lumbar MRI for hip-related pathologies. (That protocol typically includes the L4-L5 spinal level.) This cohort was chosen to represent an age-

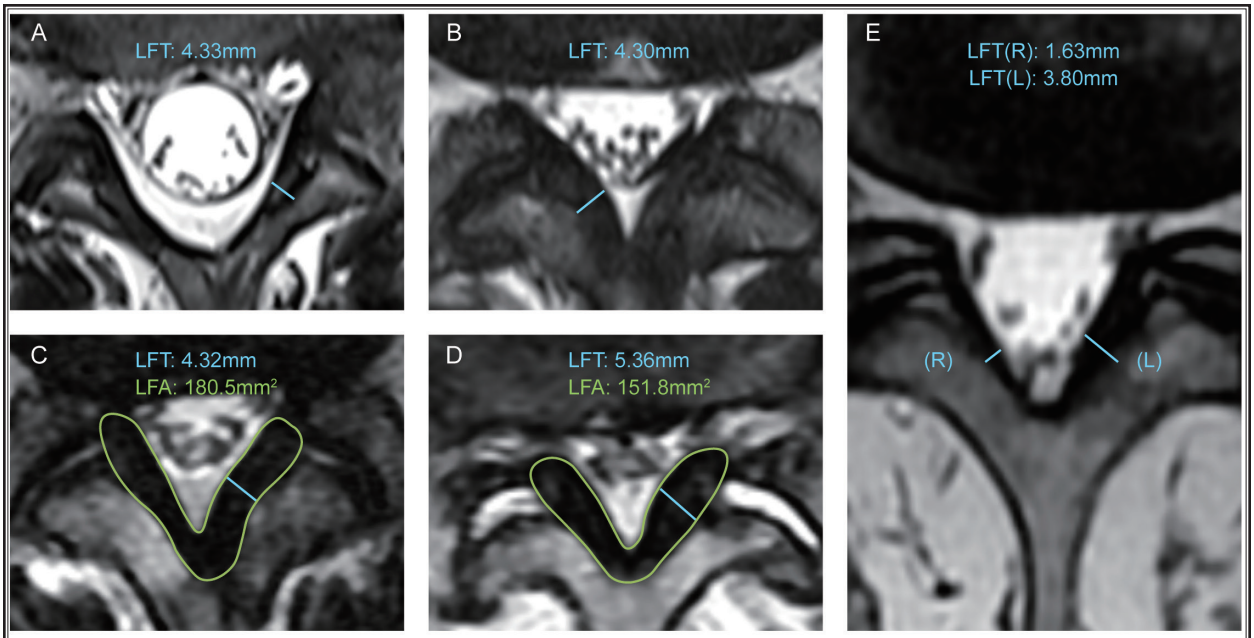

Fig. 1. *Illustration of limitations in conventional MRI parameters.*
(A, B) Limitations of ligamentum flavum thickness (LFT): Patients A and B share similar LFT values (approximately 4.3 mm) but exhibit distinct hypertrophic patterns (heterogeneous vs. uniform), which single-point measurement fails to differentiate.
(C, D) Limitations of ligamentum flavum area (LFA): Patient C has a larger LFA (180.5 mm²) than Patient D (151.8 mm²) solely because of greater ligament length, despite Patient D having a greater LFT (5.36 mm vs 4.32 mm).
(E) Asymmetry: A representative case showing significant side-to-side difference in LF thickness (3.80 mm vs. 1.63 mm).

matched, nonstenotic population, since the process of recruiting healthy volunteers for research MRI scans faced significant practical and ethical constraints. To minimize potential confounding factors, all individuals were strictly vetted and confirmed to be asymptomatic for any lumbar-related conditions.

Assessment of Clinical Severity

Given the retrospective nature of the study, standardized patient-reported outcome measures (PROMs), such as the Oswestry Disability Index (ODI) and the Visual Analog Scale (VAS), were not routinely available. Consequently, the Subjective Symptom Score (SSS) derived from admission records was utilized as a surrogate for assessing clinical severity in the LCSS group. The SSS is based on the symptomatic component of the Japanese Orthopaedic Association (JOA) scoring system, encompassing 3 domains: low back pain (0–3 points), leg pain/numbness (0–3 points), and walking ability (0–3 points). The total score ranges from 0 to 9, with lower scores indicating more severe symptoms. Data extraction was performed independently by 2 spinal surgeons, with any discrepancies resolved by consensus.

Image Processing and Analysis

MRI scans were acquired on a 3.0T system (MAGNETOM Verio, Siemens Healthineers). The acquisition parameters for this sequence were as follows: repetition time (TR)/echo time (TE), 3,000/100 ms; slice thickness, 3 mm; interslice gap, 0.9 mm; field of view (FOV), 200 x 200 mm; echo train length (ETL), 4; and matrix size, 448 × 314. Using RadiAnt DICOM Viewer (version 2024.1), 2 experienced spine surgeons measured all parameters on axial T2-weighted turbo spin-echo (TSE) images at the L4-L5 level. Four key diagnostic parameters were defined and quantified as follows (Fig. 2):

LFT: The thickness measured at the midpoint of the more hypertrophied side of the LF. The midpoint was defined as the geometric center along the curvilinear length of the ligament.

LFA: The total cross-sectional area of the LF.

LFA': The cross-sectional area of only the more hypertrophied side of the LF.

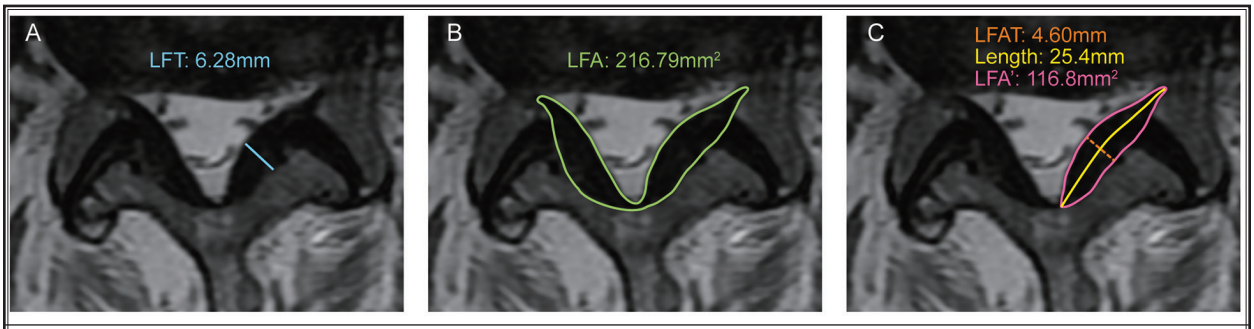

Fig. 2. *Measurement of LF parameters on T2-weighted axial MRI.*
(A) LFT.
(B) LFA.
(C) The more hypertrophied side of ligamentum flavum area (LFA') and its corresponding length. The ligamentum flavum average thickness (LFAT) was calculated as LFA' divided by its length.

LFAT: A novel parameter calculated by dividing LFA' by its corresponding length. For this calculation, LF length was defined as the curvilinear distance along the ligament's midline, measured from its medial insertion on the lamina to its lateral attachment at the facet joint capsule.

The determination of the "more hypertrophied side" was based primarily on visual assessment by the spinal surgeons. In cases wherein visual distinction was ambiguous, measurements were performed on both sides, and the side that yielded the higher value was selected for analysis.

Statistical Analysis

Data were expressed as mean ± standard deviation (SD). Independent Student's t-tests were used to compare all measurement parameters between: (1) the control and the LCSS groups and (2) men and women within each group. Statistical significance was defined as $P < 0.05$. Pearson correlation analysis assessed relationships between parameters within both the control and LCSS groups. Additionally, Spearman's correlation coefficients (ρ) were calculated to assess the relationships between the four diagnostic parameters (LFT, LFA, LFA', and LFAT) and the SSS in the LCSS group, since the SSS represented ordinal data. The diagnostic performance of the parameters was evaluated using receiver operating characteristic (ROC) curve analysis. We reported the area under the curve (AUC) with 95% confidence intervals (CIs), optimal cutoff values, sensitivity, and specificity. Pairwise comparisons of AUCs were conducted using the DeLong test for correlated ROC curves, with the resulting *P*-values adjusted by the Holm-Bonferroni correction; a *P*-value < 0.05 was considered statistically significant.

To assess the reliability of the measurements, a random sample of 40 images (20 from each group) was independently remeasured by the same observer 2 weeks later (for intra-rater reliability) and by a second senior spine surgeon blinded to the initial results and patient diagnosis (also for inter-rater reliability). The intraclass correlation coefficient (ICC) was calculated. The degree of inter- and intra-rater reliability was categorized based on ICC values as follows: excellent (> 0.90), good (0.75 – 0.90), moderate (0.50 – 0.75), and poor (< 0.50) (10).

Results

A total of 334 patients were included, comprising 166 patients in the LCSS group and 168 in the control group. Demographic and parameter comparisons are detailed in Table 1. Compared to the control group, the LCSS group exhibited significantly greater values for LFT, LFA, LFA', and LFAT (all $P < 0.001$). However, no significant difference in length was observed between the 2 groups. Within-group comparisons based on gender (Table 2) revealed that men had significantly larger LFA, LFA', and length than women in both groups (all $P < 0.05$). Conversely, LFT and LFAT showed no significant gender-based differences in either group.

As detailed in Table 3, within the control group, Pearson correlation analysis revealed that LF length was significantly and positively correlated with the LFA (r = 0.414) and LFA' (r = 0.544) while showing no significant correlation with LFT and LFAT. LFAT was positively correlated with LFT (r = 0.636), LFA (r = 0.649), and LFA' (r = 0.743). An analogous pattern of correlations was observed in the LCSS group.

The Spearman correlation analysis between the 4 diagnostic parameters and SSS is detailed in Table 4. LFA showed no statistically significant correlation with SSS. In contrast, while LFT, LFA', and LFAT demonstrated statistically significant correlations with SSS, these associations were relatively weak ($\rho < 0.4$).

ROC curve analysis was employed to assess the diagnostic utility of each parameter in identifying LCSS (Fig. 3). Among the evaluated parameters, LFAT demonstrated the most robust diagnostic performance, yielding an AUC of 0.909 (95% CI, 0.879–0.940). This figure was significantly superior to the AUCs of LFT (0.852), LFA (0.782), and LFA' (0.851). The optimal cutoff value for LFAT was determined to be 3.35 mm, which corresponded to a sensitivity of 86.7% and a specificity of 81.5% (Table 5). The DeLong test revealed no statistically significant difference between the AUCs of LFT and LFA'; however, all other pairwise comparisons showed significant differences (Table 6).

Furthermore, all measurement parameters exhibited excellent intra-rater and inter-rater reliability (ICCs > 0.90) (Table 7).

Discussion

LCSS is a degenerative spinal condition that can impair daily activities severely (11,12). Patients typically present with a constellation of symptoms, including lower back pain, radicular leg pain, and neurogenic claudication, which stem from the compression of the cauda equina and nerve roots (13-15). While the etiology of LCSS is multifactorial—involving disc and facet joint degeneration—LFH is recognized as a major pathological contributor (16,17). Pathologically, LFH involves a histological transformation characterized by the loss of elastic fibers and increased fibrosis, triggered primarily by mechanical stress and inflammation (18-23). However, quantifying this hypertrophic change accurately remains a challenge.

Researchers conventionally assess LF thickening using MRI images, specifically by measuring the thickness at the midpoint of the thicker side at the L4-L5 level (LFT) (4,8). This level is often considered the most indicative because the L4/L5 segment experiences the greatest range of motion and mechanical stress, and its LF is typically the thickest among all lumbar segments (24). However, LF thickening is frequently nonuniform, with its thickness varying across different points of the ligament. Consequently, a single-point measurement like the LFT may not be fully representative of the overall hypertrophic condition and can fail to capture the full extent of the pathology.

In 2017, Kim et al (9) introduced a novel measurement parameter known as the LFA, which involved assessing the entire cross-sectional area of the LF at L4-L5

Table 1. *Demographic and parameter comparisons between the control and LCSS groups.*

Variable	Control Group n = 168	LCSS Group n = 166	*P*-value
Gender (male/female)	82 / 86	78 / 88	0.739
Age (years)	67.39 ± 5.77	67.67 ± 6.17	0.667
LFT (mm)	3.29 ± 0.67	4.41 ± 0.86	< 0.001
LFA (mm^2)	94.69 ± 23.59	130.47 ± 38.06	< 0.001
LFA' (mm^2)	44.96 ± 10.22	65.85 ± 17.92	< 0.001
Length (mm)	15.33 ± 2.26	15.55 ± 3.28	0.487
LFAT (mm)	2.95 ± 0.55	4.25 ± 0.85	< 0.001

Data are presented as mean ± SD.
LCSS, lumbar central spinal stenosis; LFT, ligamentum flavum thickness; LFA, ligamentum flavum area; LFA', the more hypertrophied side of ligamentum flavum area; LFAT, ligamentum flavum average thickness.

Table 2. *Gender-based parameters comparisons within each group.*

	Control Group			LCSS Group		
	Men	Women	*P*-value	Men	Women	*P*-value
Age (years)	68.09 ± 6.10	66.73 ± 5.39	0.129	67.46 ± 5.73	67.86 ± 6.56	0.676
LFT (mm)	3.38 ± 0.65	3.21 ± 0.69	0.091	4.44 ± 0.88	4.39 ± 0.85	0.724
LFA (mm^2)	101.23 ± 22.95	88.46 ± 22.60	< 0.001	137.67 ± 39.79	124.01 ± 35.43	0.021
LFA' (mm^2)	47.96 ± 10.10	42.09 ± 9.53	< 0.001	68.94 ± 18.56	63.11 ± 16.97	0.036
Length (mm)	15.99 ± 2.34	14.70 ± 2.01	< 0.001	16.1 ± 3.49	15.06 ± 3.01	0.041
LFAT (mm)	3.01 ± 0.51	2.88 ± 0.58	0.114	4.31 ± 0.86	4.21 ± 0.83	0.463

Data are presented as mean ± SD.
LCSS, lumbar central spinal stenosis; LFT, ligamentum flavum thickness; LFA, ligamentum flavum area; LFA', the more hypertrophied side of ligamentum flavum area; LFAT, ligamentum flavum average thickness.

level in an axial MRI. This study showed that the LFA had higher sensitivity and specificity than the LFT did, suggesting that LFA might be a more reliable predictor of LCSS than LFT. However, LFA possesses 2 inherent limitations. Firstly, LFT often presents with asymmetry between the left and right sides, a condition likely caused by uneven mechanical stresses (25,26). This asymmetric nature has been characterized in several studies; for instance, Safak et al (25) established a statistically significant difference between the sides, while Abbas et al (4) described a right-sided tendency and Kolte et al (26) noted a left-sided predilection. Consequently, a method using a total area would artificially dilute the data. For example, if one side is severely thickened while the other is normal, the measurement will be "averaged down," masking the true severity and creating a statistically misleading value that correlates poorly with clinical presentation, potentially leading to a false-negative diagnosis. To create a more accurate and clinically robust metric, we focused exclusively on the cross-sectional area of the more hypertrophied side, which we defined as LFA' in our methodology. This approach ensured that our measurement directly captured the clinically relevant "worst-case scenario." In our study, the AUC for LFA' (0.851) was significantly greater than that for LFA (0.782), demonstrating the former's superior diagnostic performance.

However, LFA' also has a limitation: as an area-based parameter, its strong dependence on the cross-sectional curvilinear length of the LF. This inherent correlation means LFA' can be confounded by LF length between individuals. Our findings substantiated this concern. In the control group, we observed a positive

Table 3. *Pearson correlations among measured parameters for the control and LCSS groups.*

Variable	LFT	LFA	LFA'	Length	LFAT
LFT	—	0.392*	0.618*	0.078	0.763*
LFA	0.537*	—	0.740*	0.596*	0.374*
LFA'	0.470*	0.820*	—	0.659*	0.664*
Length	-0.100	0.414*	0.544*	—	-0.102
LFAT	0.636*	0.649*	0.743*	-0.143	—

Values above the diagonal represent Pearson's correlation coefficients (r) for the LCSS group. Values below the diagonal represent coefficients for the control group.
LCSS, lumbar central spinal stenosis; LFT, ligamentum flavum thickness; LFA, ligamentum flavum area; LFA', the more hypertrophied side of ligamentum flavum area; LFAT, ligamentum flavum average thickness.
*Correlation is significant at the 0.01 level (2-tailed).

Table 4. *Spearman correlation analysis between diagnostic parameters and Subjective Symptom Score in the LCSS group.*

Parameter	Spearman's Correlation Coefficient (ρ)	*P*-value
LFT	-0.290	< 0.001
LFA	-0.121	0.120
LFA'	-0.228	0.003
LFAT	-0.318	< 0.001

LCSS, lumbar central spinal stenosis; LFT, ligamentum flavum thickness; LFA, ligamentum flavum area; LFA', the more hypertrophied side of ligamentum flavum area; LFAT, ligamentum flavum average thickness.
The Subjective Symptom Score ranges from 0 to 9, with lower scores indicating more severe symptoms.

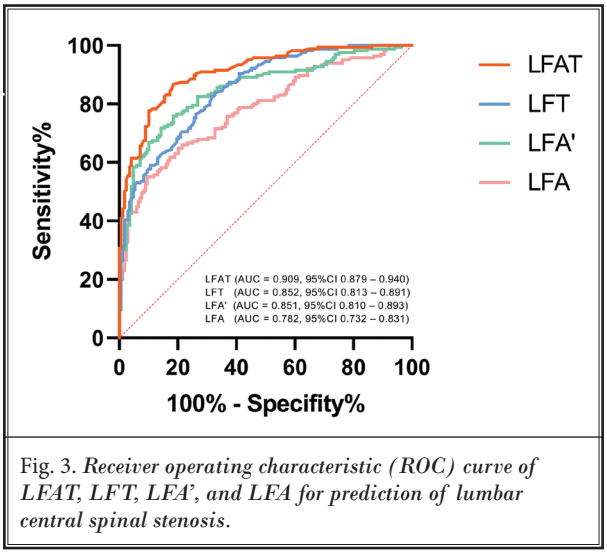

Fig. 3. *Receiver operating characteristic (ROC) curve of LFAT, LFT, LFA', and LFA for prediction of lumbar central spinal stenosis.*

Table 5. *Diagnostic performance and optimal cutoff values for each diagnostic parameter.*

Parameter	Cutoff	AUC (95% CI)	Sensitivity	Specificity
LFT (mm)	3.55	0.852 (0.813 – 0.891)	83.0%	68.5%
LFA (mm^2)	123.28	0.782 (0.732 – 0.831)	55.2%	90.5%
LFA' (mm^2)	55.13	0.851 (0.810 – 0.893)	72.1%	85.7%
LFAT (mm)	3.35	0.909 (0.879 – 0.940)	86.7%	81.5%

AUC, area under the curve; CI, confidence interval; LCSS, lumbar central spinal stenosis; LFT, ligamentum flavum thickness; LFA, ligamentum flavum area; LFA', the more hypertrophied side of ligamentum flavum area; LFAT, ligamentum flavum average thickness.

Table 6. *DeLong's test for pairwise AUC comparisons of 4 diagnostic parameters.*

Comparison	AUC Difference	95% CI	Holm-Bonferroni Corrected *P*-value
LFT - LFA	0.071	0.024 – 0.117	0.006
LFT - LFA'	0.001	-0.041 – 0.044	0.961
LFT - LFAT	-0.057	-0.092 – -0.023	0.004
LFA - LFA'	-0.069	-0.100 – -0.039	< 0.001
LFA - LFAT	-0.128	-0.174 – -0.082	< 0.001
LFA' - LFAT	-0.058	-0.095 – -0.022	0.005

AUC Difference = AUC (Method 1) - AUC (Method 2).
CI, confidence interval; LCSS, lumbar central spinal stenosis; LFT, ligamentum flavum thickness; LFA, ligamentum flavum area; LFA', the more hypertrophied side of ligamentum flavum area; LFAT, ligamentum flavum average thickness.

Table 7. *Intra-rater and Inter-rater reliability of measurement parameters.*

Parameter	Reliability Type	ICC	95% CI
LFT	Intra-rater	0.952	0.909 – 0.975
	Inter-rater	0.907	0.828 – 0.951
LFA	Intra-rater	0.962	0.928 – 0.980
	Inter-rater	0.953	0.911 – 0.975
LFA'	Intra-rater	0.948	0.902 – 0.972
	Inter-rater	0.930	0.870 – 0.963
Length	Intra-rater	0.934	0.877 – 0.965
	Inter-rater	0.917	0.847 – 0.956
LFAT	Intra-rater	0.930	0.870 – 0.963
	Inter-rater	0.945	0.898 – 0.971

ICC, intraclass correlation coefficient; CI, confidence interval; LFT, ligamentum flavum thickness; LFA, ligamentum flavum area; LFA', the more hypertrophied side of ligamentum flavum area; LFAT, ligamentum flavum average thickness.

correlation between LF length and both LFA (r = 0.414) and LFA' (r = 0.544). This confounding effect was further highlighted by the gender-based differences in our data. Specifically, while LFT and LFAT showed no significant differences between men and women, LF length, LFA, and LFA' were all significantly greater in men. Given that men generally possess a larger body habitus, it is highly probable that LF length is influenced by overall body size. Consequently, LF length acts as a significant confounding variable, which limits the diagnostic utility of measurements like LFA and LFA'.

Recognizing that the cross-sectional length of the LF confounded area-based measurements (LFA, LFA'), we sought to develop a more accurate metric. To that end, we introduced a novel parameter, LFAT. LFAT quantifies the average thickness of the LF, thereby remaining unaffected by uneven thickening or variations in cross-sectional length. In our study, LFAT showed no significant correlation with LF length. We hypothesized that LFAT would demonstrate superior precision and diagnostic efficacy. This hypothesis was confirmed by ROC analysis, wherein LFAT achieved the highest AUC of 0.909, significantly outperforming LFT (AUC = 0.852), LFA' (AUC = 0.851), and LFA (AUC = 0.782).

The superior diagnostic performance of LFAT offers substantial benefits for both clinical decision-making and radiological workflow integration. The high accuracy of this parameter (AUC = 0.909) establishes an objective basis for diagnosis, minimizing false-positive and false-negative outcomes, particularly in equivocal cases with ambiguous surgical indications, thereby facilitating more informed clinical decisions. In practice, LFAT can be implemented readily using standard measurement tools within PACS/DICOM viewers. Furthermore, deep learning algorithms can be subsequently developed to enable the automated computation of LFAT, providing immediate and reproducible quantitative data. This automated approach would streamline diagnostics, reduce the workload for radiologists, and enhance consistency by eliminating inter-observer variability.

Notably, the diagnostic performance of LFT in our study (AUC = 0.852) was considerably higher than that reported by Kim et al (AUC = 0.720) (9). We attribute this discrepancy primarily to a key methodological difference: our protocol defined LFT as the thickness at the midpoint of the LF, whereas Kim et al utilized the maximum thickness. Unlike selecting the thicker of the 2 sides—a decision that captures the dominant pathology—selecting a single "thickest point" along a ligament is inherently less representative. Any single point is just one of many, and the thickest point's location can vary unpredictably among individuals. It might occur at a location that does not impinge on the spinal cord or nerve roots, thus weakening its clinical correlation. In contrast, the midpoint provides a more standardized and consistent landmark, a practice supported by other studies (4,8), and arguably offers a more stable representation of the ligament's overall condition. This rationale is strongly supported by our own findings. The AUC for our midpoint-based LFT was not only substantially higher than that for LFA but was also remarkably close to that of LFA', our optimized area-based metric. This finding confirms that LFT, when

measured at the midpoint, serves as a relatively robust and valuable diagnostic parameter.

Another noteworthy discrepancy exists between the modest diagnostic performance of LFA in our study (AUC = 0.782) and the higher efficacy (AUC = 0.830) reported by Kim et al (9). Indeed, in our analysis, LFA demonstrated the lowest diagnostic efficacy of all tested parameters. We attribute this difference not to a simple contradiction but to a critical limitation of LFA that our study successfully uncovered: its susceptibility to confounding by LF length. Our data provided compelling evidence for this flaw. We found a positive correlation between LFA and LF length (r = 0.414)—a likely surrogate for body size—and observed significant gender-based differences in LFA that were absent in our normalized parameter, LFAT. Consequently, we postulated that the superior performance of LFA in the study by Kim et al (9) might have stemmed from a more homogeneous study population, wherein the confounding effect of body size was minimized, thus potentially inflating its perceived diagnostic accuracy. In contrast, the performance of LFA in our more diverse cohort likely represents a more realistic and generalizable assessment of its true utility. This finding critically underscores the value of LFAT, which, by design, normalizes for these anatomical variations. Its robust and superior AUC (0.909) validates it as a more universally applicable and reliable biomarker for diagnosing LFH-induced LCSS.

The Spearman correlation analysis revealed that LFAT demonstrated the strongest correlation with clinical symptoms (ρ = -0.318), in contrast to the non-significant LFA (P = 0.120), suggesting that normalizing cross-sectional area by length effectively mitigated anatomical confounding. However, it is crucial to note that despite being the strongest predictor, the correlation between LFAT and SSS remains relatively weak. This finding aligns with the well-documented radiological-clinical mismatch (27,28). Symptoms of LCSS are multifactorial, driven not only by static compression but also by dynamic instability, inflammation, and venous congestion—factors that static MRI metrics cannot fully capture (29,30).

Furthermore, the discrepancy between the high diagnostic accuracy (AUC = 0.909) and weak clinical correlation of LFAT can be attributed to the nature of the pathology: the high diagnostic efficacy indicates LFAT's ability to effectively distinguish pathological from healthy states based on a clear anatomical threshold; however, once this pathogenic threshold is crossed, symptom severity does not increase in a simple linear proportion to the degree of anatomical narrowing. Therefore, given this weak linear correlation, establishing a severity grading system based solely on LFAT lacks sufficient predictive validity at the individual patient level. We recommend LFAT as a robust binary diagnostic tool to identify significant stenosis rather than as a grading standard for assessing clinical severity linearly.

Limitations

This study has several limitations that warrant acknowledgment. First, the retrospective, single-center design and the specific Chinese cohort may limit the generalizability of our findings to other ethnic groups with different spinal anthropometrics. Consequently, the diagnostic cutoff value of 3.35 mm requires external validation in diverse populations. Second, standardized PROMs such as ODI and VAS were not routinely available; we therefore relied on the retrospective SSS derived from medical records, which, while a valid surrogate, might have lacked the granularity of prospective instruments. Future research should prioritize the inclusion of standardized PROMs to ensure a more robust assessment of clinical severity. Third, our control group comprised patients with hip pathologies rather than healthy volunteers. Chronic hip conditions could conceivably induce compensatory postural changes. Such biomechanical alterations might theoretically cause minor buckling of the LF or affect the optimal positioning and angulation of MRI slices, thereby introducing measurement errors. Although we strictly excluded individuals with lumbar symptoms to minimize confounding, future studies employing a control group of completely asymptomatic volunteers, if ethically feasible, would be valuable to corroborate our findings.

From a methodological perspective, we exclusively evaluated LFH to isolate its specific impact, meaning other major contributors to stenosis, such as disc herniation and facet joint arthropathy, were not quantitatively included. This focused approach may limit the model's applicability to multifactorial stenosis. Finally, there was a potential inter-observer variability in identifying the hypertrophied side. However, our rigorous measurement protocol yielded excellent inter-rater reliability (ICC > 0.90), suggesting those technical variations were effectively minimized.

Conclusion

This study validated LFAT as a novel MRI parameter with superior diagnostic accuracy for LCSS over

conventional metrics. At an optimal cutoff of 3.35 mm, LFAT achieved a sensitivity of 86.7% and a specificity of 81.5%. By offering a more objective and reliable assessment than other metrics used for evaluating the condition, LFAT holds considerable promise for enhancing the clinical diagnosis of LCSS.

Authors' Contributions

Wang Y designed the study, performed the research, and drafted the article. Ji W and Zhang ZM coordinated the resources, supervised the study, and revised the article. Liu KF, Liu XY, Zhang XH, Xie JH, Xu PJ, Xiong J, Zheng JY and Li KQ collected and analyzed the data.

References

1. Kalichman L, Cole R, Kim DH, et al. Spinal stenosis prevalence and association with symptoms: The Framingham Study. *Spine J* 2009; 9:545-550.
2. Haig AJ, Tomkins CC. Diagnosis and management of lumbar spinal stenosis. *JAMA* 2010; 303:71-72.
3. Manchikanti L, Kaye AD, Manchikanti K, Boswell M, Pampati V, Hirsch J. Efficacy of epidural injections in the treatment of lumbar central spinal stenosis: A systematic review. *Anesth Pain Med* 2015; 5:e23139.
4. Abbas J, Hamoud K, Masharawi YM, et al. Ligamentum flavum thickness in normal and stenotic lumbar spines. *Spine (Phila Pa 1976)* 2010; 35:1225-1230.
5. Manchikanti L, Cash KA, McManus CD, Damron KS, Pampati V, Falco FJ. A randomized, double-blind controlled trial of lumbar interlaminar epidural injections in central spinal stenosis: 2-year follow-up. *Pain Physician* 2015; 18:79-92.
6. Yabe Y, Hagiwara Y, Ando A, et al. Chondrogenic and fibrotic process in the ligamentum flavum of patients with lumbar spinal canal stenosis. *Spine (Phila Pa 1976)* 2015; 40:429-435.
7. Park JB, Chang H, Lee JK. Quantitative analysis of transforming growth factor-beta 1 in ligamentum flavum of lumbar spinal stenosis and disc herniation. *Spine (Phila Pa 1976)* 2001; 26:E492-E495.
8. Chokshi FH, Quencer RM, Smoker WRK. The "thickened" ligamentum flavum: Is it buckling or enlargement? *AJNR Am J Neuroradiol* 2010; 31:1813-1816.
9. Kim YU, Park JY, Kim DH, et al. The role of the ligamentum flavum area as a morphological parameter of lumbar central spinal stenosis. *Pain Physician* 2017; 20:E419-E424.
10. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; 15:155-163.
11. Cook CJ, Cook CE, Reiman MP, Joshi AB, Richardson W, Garcia AN. Systematic review of diagnostic accuracy of patient history, clinical findings, and physical tests in the diagnosis of lumbar spinal stenosis. *Eur Spine J* 2020; 29:93-112.
12. Katz JN, Zimmerman ZE, Mass H, Makhni MC. Diagnosis and management of lumbar spinal stenosis: A review. *JAMA* 2022; 327:1688-1699.
13. Beamer YB, Garner JT, Shelden CH. Hypertrophied ligamentum flavum. Clinical and surgical significance. *Arch Surg* 1973; 106:289-292.
14. Rauschning W. Normal and pathologic anatomy of the lumbar root canals. Spine (Phila Pa 1976) 1987; 12:1008-1019.
15. Schonstrom NS, Bolender NF, Spengler DM. The pathomorphology of spinal stenosis as seen on CT scans of the lumbar spine. *Spine (Phila Pa 1976)* 1985; 10:806-811.
16. Lu QL, Wang XZ, Xie W, Chen XW, Zhu YL, Li XG. Macrophage migration inhibitory factor may contribute to hypertrophy of lumbar ligamentum flavum in type 2 diabetes mellitus. *Chin Med J (Engl)* 2020; 133:623-625.
17. Cheung PWH, Tam V, Leung VYL, et al. The paradoxical relationship between ligamentum flavum hypertrophy and developmental lumbar spinal stenosis. *Scoliosis Spinal Disord* 2016; 11:26.
18. Yayama T, Kobayashi S, Sato R, et al. Calcium pyrophosphate crystal deposition in the ligamentum flavum of degenerated lumbar spine: Histopathological and immunohistological findings. *Clin Rheumatol* 2008; 27:597-604.
19. Evans JH, Nachemson AL. Biomechanical study of human lumbar ligamentum flavum. *J Anat* 1969; 105:188-189.
20. Nachemson AL, Evans JH. Some mechanical properties of the third human lumbar interlaminar ligament (ligamentum flavum). *J Biomech* 1968; 1:211-220.
21. Munns JJ, Lee JY, Espinoza Orías AA, et al. Ligamentum flavum hypertrophy in asymptomatic and chronic low back pain subjects. *PLoS One* 2015; 10:e0128321.
22. Sairyo K, Biyani A, Goel VK, et al. Lumbar ligamentum flavum hypertrophy is due to accumulation of inflammation-related scar tissue. *Spine (Phila Pa 1976)* 2007; 32:E340-E347.
23. Sairyo K, Biyani A, Goel V, et al. Pathomechanism of ligamentum flavum hypertrophy: A multidisciplinary investigation based on clinical, biomechanical, histologic, and biologic assessments. *Spine (Phila Pa 1976)* 2005; 30:2649-2656.
24. Altinkaya N, Yildirim T, Demir S, Alkan O, Sarica FB. Factors associated with the thickness of the ligamentum flavum: is ligamentum flavum thickening due to hypertrophy or buckling? *Spine (Phila Pa 1976)* 2011; 36:E1093-E1097.
25. Safak AA, Is M, Sevinc O, et al. The thickness of the ligamentum flavum in relation to age and gender. *Clin Anat* 2010; 23:79-83.
26. Kolte VS, Khambatta S, Ambiye MV. Thickness of the ligamentum flavum: Correlation with age and its asymmetry: An magnetic resonance imaging study. *Asian Spine J* 2015; 9:245-253.
27. Fushimi Y, Otani K, Tominaga R, Nakamura M, Sekiguchi M, Konno SI. The association between clinical symptoms of lumbar spinal stenosis and MRI axial imaging findings. *Fukushima J Med Sci* 2021; 67:150-160.
28. Doualla-Bija M, Takang MA, Mankaa E, Moutchia J, Ongolo-Zogo P, Luma-Namme H. Characteristics and determinants of clinical symptoms in radiographic lumbar spinal stenosis in a tertiary health care centre in sub-Saharan Africa. *BMC Musculoskelet Disord* 2017; 18:494.
29. Shunmugavel A, Martin MM, Khan M, et al. Simvastatin ameliorates cauda equina compression injury in a rat model of lumbar spinal stenosis. *J Neuroimmune Pharmacol* 2013; 8:274-286.
30. Sengupta DK, Herkowitz HN. Lumbar spinal stenosis. Treatment strategies and indications for surgery. *Orthop Clin North Am* 2003; 34:281-295.