Clinical validation of an AI system for MRI-based detection of lumbar degenerative pathology

Marie-Pauline Talabard[1], Imad Bouaddi[2], Marion Durteste[3], Rami Naidji[3], David Vittecoq[3], Nor-Eddine Regnard[2,3]

(1) Hôpital Cochin, Paris, France · (2) Réseau d'Imagerie Sud Francilien, Lieusaint, France · (3) Gleamer, Paris, France

Introduction

- Low back pain is the leading cause of years lived with disability worldwide, and MRI is the reference standard for patient assessment [1, 2]

- Lumbar spine MRI interpretation is challenging due to the multiplicity of classification systems and their reliance on subjective visual grading [3]

- Existing AI algorithms are limited in scope or too complex for routine clinical implementation [4, 5]

What is the performance of a novel AI solution for detecting lumbar degenerative pathology on MRI?

Methods

✓ **Data Collection:**

- Retrospective collection of adult lumbar spine MRIs from French private practices
- MRI protocols included sagittal T1, T2 (DIXON/STIR) and axial T2-weighted sequences (2D or 3D)

✓ **Study Design:**

- Standalone AI performance
- Reader study: 1 radiologist read all cases twice with and without AI assistance

✓ **Ground Truth:**

- Two expert musculoskeletal radiologists reviewed all exams (T8-T9 to L5-S1)
- Eight pathologies assessed: disc herniation, disc bulging, foraminal stenosis, facet arthropathy, spondylolisthesis, disc degeneration, endplate degeneration, and c central canal stenosis

✓ **AI Software:**

- LumbarMR (Gleamer, Paris, France) generates a structured report of degenerative pathologies across all intervertebral levels

✓ **Statistical Analysis:**

- Observation-wise sensitivity and specificity (observation = patient × intervertebral level × pathology)
- Per-pathology sensitivity and specificity
- Interpretation time and diagnostic performance with and without AI

Results

Patient characteristic	Value
Patient	
Sample size (n)	199
Age	
Mean ± SD (y)	47.2 ± 14.5
Range (y)	20 – 88
Sex	
Women	116 (58.3%)
Men	83 (41.7%)
Pathology Burden (number of levels with ≥1 pathology)	
None (0)	22 (11.1%)
Low (1-3)	47 (23.6%)
Moderate (4-7)	54 (27.1%)
High (≥8)	76 (38.2%)

Table 1. Characteristics of included patients

Overall AI performance across 10,647 analyzable observations:

Overall sensitivity
78.4%
(95% CI: 76.2–80.6)

Overall specificity
94.4%
(95% CI: 93.9-94.9)

Per-pathology AI performance is shown in Figure 1

Subgroup analyses

- AI sensitivity was significantly higher in older patients (86.0% vs. 73.5%, p < 0.001)

Reader study

- Significant reduction in radiologist interpretation time (116.4 s to 97.2 s, p < 0.001)
- Preserved sensitivity (p = 0.87) and specificity (p = 0.25)

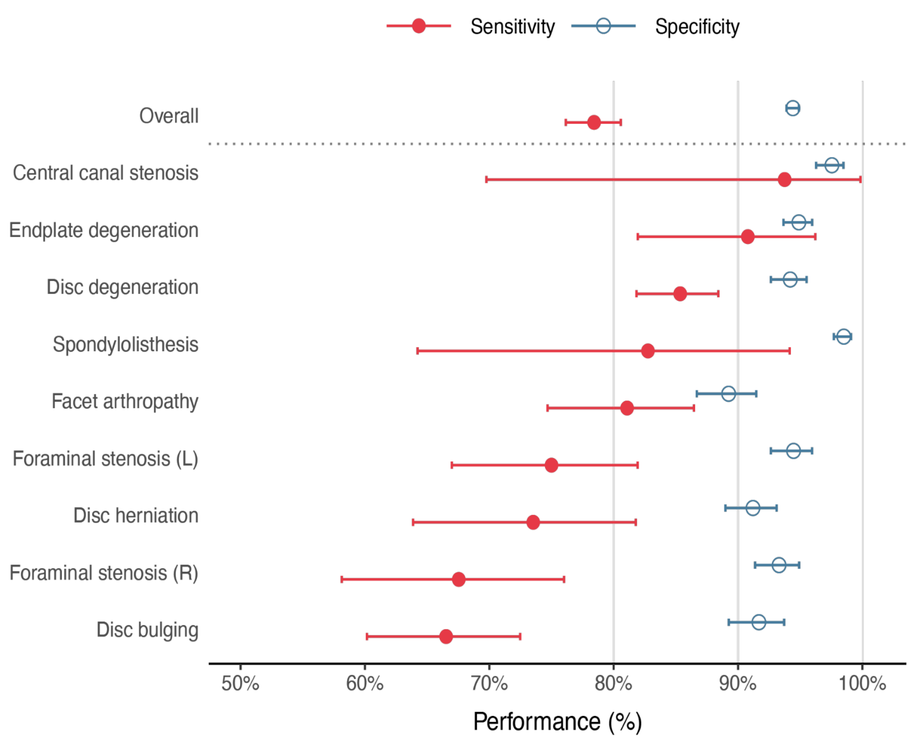

Figure 1. Standalone performance of the AI per pathology

Conclusion

- The AI solution demonstrated robust standalone performance for detecting lumbar degenerative pathology on MRI
- Performance was highest for central canal stenosis and endplate degeneration
- AI assistance reduced radiologist interpretation time by 16.2%

References

[1] Buchbinder et al. Lancet 2018.
[2] de Souza et al. Clinics 2019.
[3] Esposito et al. JOR Spine 2025.
[4] Soydan et al. Glob. Spine J. 2025.
[5] van der Graaf et al. Eur. Radiol. 2025.

deephealth