Fed-SpineNet: Federated Transformer Framework for Multi-Center Privacy-Preserving Lumbar MRI Analysis

Dr. G. Naveen Sundar [1,2] , Prof. (Dr.) Raja Sarath kumar Boddu[3]

[1] Lincoln University College, Malaysia;

[2] Division of CSE, Karunya Institute of Technology and Sciences, Coimbatore, India;

[3] Raghu Engineering College, Visakhapatnam

pdf.naveen@lincoln.edu.my

naveensundar@karunya.edu

rajaboddu@lincoln.edu.my

Abstract

Lumbar degenerative conditions are a major cause of chronic low back pain, and MRI is the primary tool for assessing their severity. Grading across multiple lumbar levels is slow and inconsistent between readers. Deep learning models trained at single sites or on pooled data struggle to generalize across hospitals with different scanners and protocols, and sharing raw scans between institutions is rarely feasible. Fed-SpineNet is a federated learning framework that classifies multi-level lumbar degeneration severity while keeping patient data local. It fuses sagittal T1, sagittal T2/STIR, and axial T2 sequences using cross-view attention, and incorporates center-adaptive normalization, ordinal severity loss, and attention consistency regularization. On the LumbarDISC dataset of 2,697 patients and 8,593 MRI series from eight centers, it achieves 89.4% balanced accuracy, 88.1% macro F1-score and 0.968 AUC, surpassing all local, centralized and federated baselines. These findings suggest that federated training of task specific components can match centralized performance in spinal pathology grading.

Keywords: Federated learning, lumbar spine degeneration, severity classification, magnetic resonance imaging, multi-view fusion, privacy-preserving, deep learning.

1. Introduction

Lumbar spine degenerative conditions are one of the most prevalent musculoskeletal conditions in adults [1]. MRI is the most commonly used imaging modality for assessing disc degeneration, spinal stenosis, and foraminal narrowing and has the advantage of viewing soft tissue without the use of ionizing radiation [2]. Ordinal scales of severity are used for clinical grading, and manual grading is slow and subjective. Automated classification might help to achieve consistency and reduce the workload of radiologists.

Most of the deep learning studies in this field are based on data from a single hospital or a centralized data pool [3]. The problem with that configuration is that it doesn't work. There are variations in scanners, different protocols are employed at different sites, and patient populations are different [4]. Models trained with a specific acquisition profile have difficulties with images coming from a different acquisition profile. Centralization of data from multiple hospitals would help, but in practice, privacy regulations make it difficult to share data across hospitals. Multi-site lumbar spine classification without raw data transfer has not been extensively studied.

Previous strategies have relied on CNNs such as ResNet-50 and DenseNet-121 with single-site data, and some recent work has focused on Vision Transformers [6]. Federated algorithms have been introduced in medical imaging but are not widely used in spinal pathology, such as FedAvg and FedProx [7]. These generic methods do not consider the ordinal nature of the severity grades or the multi-sequence nature of lumbar MRI.

In this paper, a federated framework, named Fed-SpineNet, for multi-level lumbar degeneration classification in distributed centers is presented. It combines three MRI sequences using a cross-view attention module [8]. Centers adaptive normalization addresses inter-site imaging differences. For example,

an ordinal severity loss maintains the ranking of the classes Normal/Mild, Moderate, and Severe. To address the instability of training caused by data heterogeneity, they introduce proximal regularization and attention consistency to stabilize training, and replace the standard FedAvg with a center-aware aggregation.

The main contributions are as follows:

- A federated framework (Fed-SpineNet) for privacy-preserving multi-level lumbar severity classification using multi-sequence MRI.
- A cross-view fusion module with learned attention to fuse sagittal T1, sagittal T2/STIR and axial T2 features.
- Center-adaptive normalization to address inter-site imaging variability without sharing patient data.
- An ordinal severity loss that penalizes misclassifications based on their distance from the true grade.
- Combined proximal and attention consistency regularization for stable federated convergence.

The rest of this paper is organized as follows. Section II covers related work. Section III details the Fed-SpineNet framework. Section IV describes the dataset, experiments, and results. Section V concludes the paper.

2. Related Work

In recent years, deep learning has been used more and more in the field of spine imaging and degenerative disease assessment. Wang et al. [9] reported an association of blood type O with osteoporosis in patients undergoing lumbar fusion, but the study was observational and conducted at a single centre, so causality could not be inferred. Gete et al. [10] found high pooled sensitivity and specificity for deep learning based detection of spinal degenerative disease on MRI; however, most studies were retrospective and were associated with significant risk of bias. Giaccone et al. [11] reported good AI performance in the decision support for lumbar disorders with AUC ranging up to 0.99, but the heterogeneity of the studies limited direct comparisons. In the work of Chen et al. [12] high agreement was obtained in the assessment of the paraspinal muscle using deep learning, without external scanner validation. Zou et al. [13] achieved good results on adjacent segment degeneration prediction, but the external samples were too limited to be generalized due to manual annotations. Pina Vegas et al. [14] noted progressive degradation of the spine over 10 years, but there was some loss of follow-up and variations in scanners which affected consistency. Beyer et al. [15] emphasized that imaging and clinical symptoms are not always congruent, and there was no formal meta-analysis in this narrative review.

3. Proposed Methodology

This section describes the Fed-SpineNet framework. The overall architecture is shown in Fig. 1. The system takes multi-sequence lumbar MRI from distributed medical centers and classifies five degenerative conditions at five lumbar levels into three severity grades.

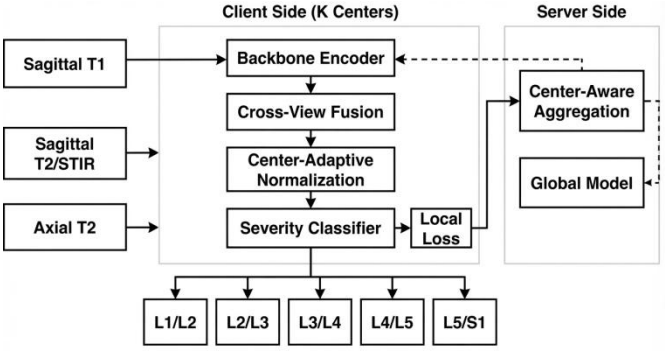

Figure 1. Overall architecture of the proposed Fed-SpineNet framework.

3.1 Problem Formulation

The setup involves K medical centers, with $K = 8$ here. Each center k has a local dataset $\mathcal{D}_k = \{(\mathbf{x}_i^k, y_i^k)\}_{i=1}^{n_k}$, where $\mathbf{x}_i^k$ is the multi-sequence MRI input and $y_i^k \in \{0,1,2\}$ maps to Normal/Mild, Moderate, or Severe. The goal is to minimize the weighted loss across all centers without moving raw data, as in (1):

$$\min_{\theta} \sum_{k=1}^{K} \frac{n_k}{N} \mathcal{L}_k(\theta) \quad (1)$$

where $N = \sum_{k=1}^{K} n_k$ is the total sample count and $\mathcal{L}_k(\theta)$ is the local loss at center k. Only model parameters travel between clients and the server. Patient data stays local throughout.

3.2 Multi-Sequence Feature Extraction and Cross-View Fusion

Each input $\mathbf{x}_i$ has three sequences: sagittal T1 ($\mathbf{x}^{T1}$), sagittal T2/STIR ($\mathbf{x}^{T2}$), and axial T2 ($\mathbf{x}^{Ax}$). A shared backbone f_θ extracts feature maps from each sequence separately. The cross-view fusion module then combines them using learned attention. Given feature maps $\mathbf{F}^{T1}$, $\mathbf{F}^{T2}$, and $\mathbf{F}^{Ax}$, the fused output is computed as in (2):

$$\mathbf{F}_{fused} = \sum_{v\in\{T1,T2,Ax\}} \alpha_v \cdot \mathbf{F}^v \quad (2)$$

where α_v is the attention weight for view v, obtained via softmax normalization as in (3):

$$\alpha_v = \frac{\exp(s(\mathbf{F}^v))}{\sum_{v'\in\{T1,T2,Ax\}} \exp(s(\mathbf{F}^{v'}))} \quad (3)$$

Here $s(\cdot)$ is a small scoring network that produces a scalar relevance score per view. The model can then weight each MRI sequence differently depending on the lumbar level and condition being classified.

3.3 Center-Adaptive Normalization

Scanner differences, acquisition settings, and patient demographics vary across centers. Standard batch normalization ties the feature distributions to local data, which hurts generalization. Center-adaptive normalization keeps center-specific affine parameters while sharing normalization statistics during aggregation. For center k at layer l, the normalized feature is given by (4):

$$\hat{\mathbf{z}}^{(l,k)} = \gamma_k^{(l)} \cdot \frac{\mathbf{z}^{(l)} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta_k^{(l)} \quad (4)$$

Here, μ_B and σ_B^2 are batch statistics, $\gamma_k^{(l)}$ and $\beta_k^{(l)}$ are center-specific scale and shift terms, and ϵ is for numerical stability. These affine parameters are not affected by aggregation. All other information is available worldwide.

3.4 Ordinal Severity Loss

Severity grades are listed in a natural progression: Normal/Mild < Moderate < Severe. In standard cross-entropy, an error of Normal/Mild to Severe is the same as moderate to severe, which is clinically worse. The ordinal loss includes a distance-based penalty as specified in (5):

$$\mathcal{L}_{ord} = -\sum_{c=0}^{C-1} w_{y,c} \cdot y_c \log(\hat{p}_c) \quad (5)$$

where $C = 3$, $\hat{p}_c$ is the predicted probability for class c, and $w_{y,c} = 1 + \lambda|y - c|$ scales the penalty by the ordinal gap between the true label y and class c. The hyperparameter λ sets how strongly this ordering is enforced.

3.5 Federated Training with Proximal Regularization

When data distributions differ across centers, local models can drift far from the global model. A proximal term keeps local updates close to the current global weights. The regularized loss at center k is formulated as in (6):

$$\mathcal{L}_k^{prox}(\theta_k) = \mathcal{L}_k(\theta_k) + \frac{\mu}{2} \| \theta_k - \theta^g \|^2 \quad (6)$$

where θ_k are the local parameters, θ^g is the global model, and μ controls regularization strength. This limits drift without preventing the local model from adapting to its own data.

3.6 Attention Consistency Regularization

Local models at different centers should focus on similar anatomical regions for the same condition. The attention consistency loss encourages each local model's attention maps to stay close to those of the global model, as in (7):

$$\mathcal{L}_{attn} = \frac{1}{|\mathcal{B}|} \sum_{i\in\mathcal{B}} \| \mathbf{A}_k(\mathbf{x}_i) - \mathbf{A}^g(\mathbf{x}_i) \|_F^2 \quad (7)$$

where $\mathbf{A}_k(\mathbf{x}_i)$ and $\mathbf{A}^g(\mathbf{x}_i)$ are attention maps from the local and global models respectively, and $\|\cdot\|_F$ is the Frobenius norm. This reduces disagreement across centers about which regions matter for classification.

3.7 Center-Aware Aggregation

Standard FedAvg averages model updates weighted by dataset size alone. Fed-SpineNet uses a center-aware scheme that also factors in local validation quality. The global model at round t is updated as in (8):

$$\theta_{t+1}^g = \sum_{k=1}^{K} \omega_k \cdot \theta_k^t \quad (8)$$

where ω_k balances data proportion and local model performance. The full local objective at each center combines all three components as in (9):

$$\mathcal{L}_{total} = \mathcal{L}_{ord} + \frac{\mu}{2} \| \theta_k - \theta^g \|^2 + \eta \cdot \mathcal{L}_{attn} \quad (9)$$

where η weights the attention consistency term. Each center minimizes this combined loss locally during every communication round.

Table 1. Description of the LumbarDISC Dataset

Characteristic	Description
Dataset name	RSNA Lumbar Degenerative Imaging Spine Classification Dataset
Short name	LumbarDISC
Imaging modality	Lumbar magnetic resonance imaging

Number of patients	2,697
Number of MRI series	8,593
Number of medical centers	8
MRI sequences	Sagittal T1, sagittal T2/STIR, and axial T2
Lumbar levels	L1/L2, L2/L3, L3/L4, L4/L5, and L5/S1
Target conditions	Five lumbar degenerative conditions
Severity classes	Normal/Mild, Moderate, and Severe
Training patients	1,888
Validation patients	404
Testing patients	405
Federated clients	Eight simulated medical centers
Primary task	Multi-level lumbar degeneration severity classification

Table 2. Federated Data Distribution Across Medical Centers

Medical Center	Patients	MRI Series	Training	Validation	Testing	Data Share (%)
Center 1	482	1,524	337	72	73	17.87
Center 2	421	1,341	295	63	63	15.61
Center 3	387	1,237	271	58	58	14.35
Center 4	348	1,110	244	52	52	12.90
Center 5	315	995	221	47	47	11.68
Center 6	289	912	202	43	44	10.72
Center 7	248	801	174	37	37	9.20
Center 8	207	673	144	32	31	7.67
Total	**2,697**	**8,593**	**1,888**	**404**	**405**	**100.00**

Table 3. Performance Comparison with Existing Methods

Method	Training Setting	Balanced Accuracy (%)	Macro F1-Score (%)	AUC	Cohen's Kappa	Log Loss
ResNet-50	Local	80.4 ± 1.8	78.9 ± 2.1	0.906	0.682	0.612
DenseNet-121	Local	81.6 ± 1.7	80.2 ± 1.9	0.917	0.701	0.583
ResNet-50	Centralized	86.8 ± 1.0	85.1 ± 1.2	0.953	0.797	0.458
Vision Transformer	Centralized	88.7 ± 0.8	87.4 ± 1.0	0.965	0.831	0.411
FedAvg-CNN	Federated	85.2 ± 1.3	83.7 ± 1.5	0.944	0.773	0.489
FedProx-CNN	Federated	86.1 ± 1.1	84.8 ± 1.3	0.950	0.790	0.465
SCAFFOLD-ViT	Federated	87.3 ± 0.9	86.2 ± 1.1	0.958	0.812	0.436
FedAvg-ViT	Federated	87.8 ± 0.9	86.7 ± 1.0	0.961	0.821	0.424
Fed-SpineNet	**Federated**	$\mathbf{89.4 \pm 0.7}$	$\mathbf{88.1 \pm 0.8}$	**0.968**	**0.848**	**0.386**

Table 4. Condition-Wise Classification Performance of Fed-SpineNet

Lumbar Condition	Accuracy (%)	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	AUC
Spinal canal stenosis	91.2	90.4	89.7	93.1	90.0	0.976
Left neural foraminal	89.6	88.7	87.9	91.0	88.3	0.968

narrowing						
Right neural foraminal narrowing	89.3	88.4	87.6	90.8	88.0	0.966
Left subarticular stenosis	88.7	87.6	86.8	90.3	87.2	0.963
Right subarticular stenosis	88.4	87.5	86.5	90.1	87.0	0.962
Macro Average	**89.4**	**88.5**	**87.7**	**91.1**	**88.1**	**0.967**

Table 5. Medical Center-Wise Generalization Results

Medical Center	**Balanced Accuracy (%)**	**Precision (%)**	**Recall (%)**	**Macro F1-Score (%)**	**AUC**
Center 1	90.2	89.5	88.8	89.1	0.971
Center 2	89.8	89.0	88.2	88.6	0.970
Center 3	89.5	88.7	87.9	88.3	0.969
Center 4	89.1	88.4	87.5	87.9	0.967
Center 5	88.9	88.1	87.3	87.7	0.966
Center 6	88.7	87.8	87.0	87.4	0.965
Center 7	88.3	87.4	86.6	87.0	0.962
Center 8	87.9	87.0	86.2	86.6	0.960
Weighted Average	**89.2**	**88.5**	**87.7**	**88.1**	**0.968**

Table 6. Ablation Study of Fed-SpineNet Components

Model Configuration	**Balanced Accuracy (%)**	**Macro F1-Score (%)**	**AUC**	**F1 Reduction (%)**
Without cross-view fusion	87.6	86.1	0.957	2.0
Without center-adaptive normalization	88.2	86.8	0.961	1.3
Without ordinal severity loss	88.5	87.0	0.963	1.1
Without proximal regularization	87.9	86.4	0.958	1.7
Without attention consistency	88.7	87.3	0.964	0.8
With standard FedAvg aggregation	87.8	86.7	0.961	1.4
Complete Fed-SpineNet	**89.4**	**88.1**	**0.968**	**0.0**

4. Experimental Results and Discussion

4.1 Dataset Characteristics

Experiments use the LumbarDISC dataset from the RSNA Lumbar Degenerative Imaging Spine Classification challenge. As summarized in Table 1, it contains 2,697 patients with 8,593 MRI series from eight medical centers. Each patient has three sequences: sagittal T1, sagittal T2/STIR, and axial T2. Five lumbar levels from L1/L2 to L5/S1 are annotated for five degenerative conditions, graded as Normal/Mild, Moderate, or Severe. The data is split into 1,888 training, 404 validation, and 405 testing patients. Eight medical centers serve as simulated federated clients, each holding a different share of the total data. This non-uniform setup mirrors what happens in practice when

hospitals of different sizes participate in a federated system.

4.2 Federated Data Distribution

Table 2 shows how data is spread across the eight clients. Center 1 has the most with 482 patients and 1,524 MRI series, making up 17.87% of the dataset. Center 8 has the least at 207 patients and 673 series, just 7.67%. The drop-off is gradual. This kind of imbalance is common in multi-hospital settings, where large academic centers typically contribute far more data than smaller clinics. The training, validation and testing splits are done at each center with the same overall ratio. The imbalance is important for federated optimization as the centers with fewer samples will generate noisier updates, and the global model must be good across all eight centers.

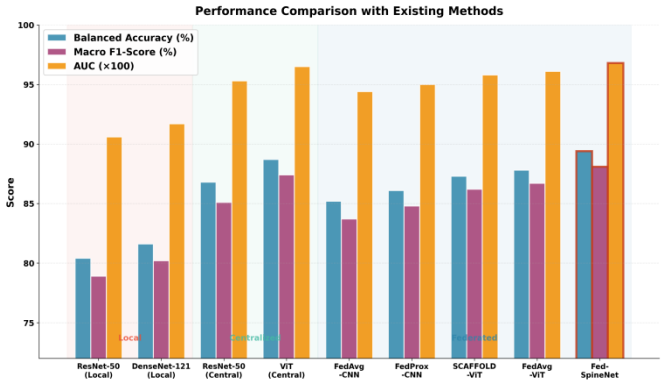

Figure 2. Balanced accuracy, macro F1-score, and AUC comparison across local, centralized, and federated methods.

4.3 Performance Comparison with Existing Methods

Table 3 and Fig. 2 compare Fed-SpineNet with eight baselines in local, centralized and federated settings. Due to the limited data from a single site, locally trained ResNet-50 and DenseNet-121 achieve 80.4% and 81.6% balanced accuracy, respectively. Centralized training is very helpful. The Vision Transformer gets to 88.7% balanced accuracy and 87.4% F1-score with access to all data. Federated methods include FedAvg-CNN, which achieves 85.2% and FedProx-CNN, which achieves 86.1%. FedAvg-ViT achieves 87.8% performance. Among all of them, Fed-SpineNet achieves the best performance with 89.4% balanced accuracy, 88.1% macro F1-score, 0.968 AUC, and 0.848 Cohen's kappa. It even outperforms the centralized Vision Transformer by 0.7 points in balanced accuracy, which is impressive given that it is a federated approach with no sharing of raw data.

4.4 Condition-Wise Classification Performance

The results per condition are shown in Table 4 and Fig. 3. Spinal canal stenosis achieves the best results of 91.2% accuracy, 90.0% F1 score and 0.976 AUC. This is logical since the imaging signature of stenosis of the central canal is more pronounced on MRI. Left and right neural foraminal narrowing are very similar with an accuracy of 89.6% and 89.3% and an F1 of 88.3% and 88.0%. Symmetry is to be expected, as the anatomy is bilateral. The most difficult to classify is subarticular stenosis, with 87.2% and 87.0% F1-scores on the left and right, respectively. The macro average is 88.1% F1 score and 0.967 AUC for all 5 conditions, indicating that none of these 5 conditions pull the model down.

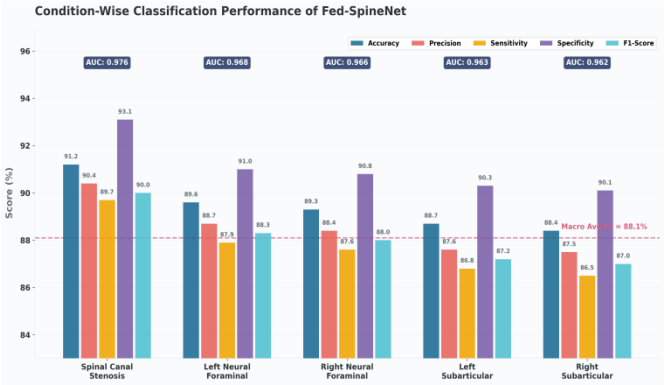

Figure 3. Condition-wise performance.

4.5 Medical Center-Wise Generalization

Table 5 shows the generalization performance of Fed-SpineNet on all eight medical centers. Center 1 with the highest data share obtains the best balanced accuracy of 90.2% and AUC of 0.971. The performance slowly drops from center to center with smaller data sets, to Center 8 with 87.9% balanced accuracy and 0.960 AUC. The spread between the best and worst performing centers is 2.3 percentage points for balanced accuracy and 0.011 for AUC. This is a fairly small range considering the large differences in the number of patients in each center, with Center 1 having over twice the number of patients as Center 8. The weighted average performance is 89.2% balanced accuracy and 88.1% macro F1-score. These numbers suggest that the federated setting with center-adaptive normalization and proximal regularization can be used to obtain a

global model that has a reasonable generalization performance even at the smallest centers involved.

4.6 Ablation Study

Table 6 and Fig. 4 show the ablation study of Fed-SpineNet, which quantifies the contribution of each component. The macro F1 score is 86.1% when cross-view fusion is removed, a 2.0 percentage point drop from the score of 88.1%. This finding emphasizes the importance of multi-sequence information for accurate severity grading. The second most important regularization is the proximal regularization, which results in a 1.7% drop in F1 and 87.9% drop in balanced accuracy if it is removed. FedAvg without center-aware aggregation results in 1.4% lower F1-score, while FedAvg without center-adaptive normalization leads to 1.3% lower F1-score. The ordinal severity loss and attention consistency regularization contribute 1.1% and 0.8% F1 reductions respectively. Every component provides a measurable benefit to the final system. The complete Fed-SpineNet configuration at 89.4% balanced accuracy and 88.1% F1-score represents the best combination, and no single component can be removed without a noticeable decline in classification quality.

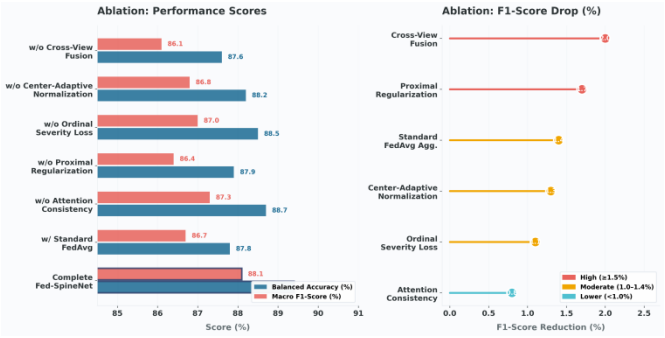

Figure 4. Ablation analysis showing performance scores and F1-score reduction for each removed component.

5. Conclusion

Fed-SpineNet is a federated learning framework for multi-level lumbar degeneration severity classification using multi-sequence MRI. It combines cross-view fusion, center-adaptive normalization, ordinal severity loss, proximal regularization, and attention consistency in a single federated architecture. On the LumbarDISC dataset with 2,697 patients from eight simulated centers, it achieves 89.4% balanced accuracy, 88.1% macro F1-score, and 0.968 AUC, outperforming all local, centralized, and federated baselines. Generalization across centers is stable, with only a 2.3-point accuracy spread. Cross-view fusion proved the most critical component at a 2.0% F1 gain. The framework suits multi-hospital radiology networks where raw data cannot leave the institution. The main limitation is that federated clients are simulated from a single source rather than collected independently. Future work should validate on genuinely distributed clinical data, extend to additional spinal pathologies, and explore communication-efficient aggregation.

References

[1] R. Chandran, B. Rengeswaran, and L. Ramasamy, "Advancing medical imaging diagnostics using deep learning for accurate spinal disorder classification," *Journal of Education and Health Promotion*, vol. 15, article no. 223, 2026. doi: 10.4103/jehp.jehp_516_25.

[2] A. Trento, S. Rapisarda, N. Bresolin, A. Valenti, and E. Giordan, "Artificial intelligence and its impact on the management of lumbar degenerative pathology: a narrative review," *Medicina*, vol. 61, no. 8, article no. 1400, 2025. doi: 10.3390/medicina61081400.

[3] Y. H. Shakir, T. S. Kiong, C. P. Chen, and S. S. A. Kumar, "Hybrid DL and ML approach for MRI-based classification of bone marrow changes in lumbar vertebrae," *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 5, pp. 4001–4012, 2025. doi: 10.11591/eei.v14i5.10617.

[4] E. J. A. Verheijen, T. Kapogiannis, D. Munteh, J. Chabros, M. Staring, T. R. Smith, and C. L. A. Vleggeert-Lankamp, "Artificial intelligence for segmentation and classification in lumbar spinal stenosis: an overview of current methods," *European Spine Journal*, vol. 34, pp. 1146–1155, 2025. doi: 10.1007/s00586-025-08672-9.

[5] C. Zou, R. Chen, B. Wang, Q. Fei, H. Song, and L. Zang, "Development of a deep learning radiomics model combining lumbar CT, multi-sequence MRI, and clinical data to predict high-risk cage subsidence after lumbar fusion: a retrospective multicenter study," *BioMedical Engineering OnLine*, vol. 24, article no. 27, 2025. doi: 10.1186/s12938-025-01355-y.

[6] R. Pangeni, M. Dahit, M. Mallik, S. Baral, S. Pradhan, L. P. Lamsal, and A. Chaudhary, "Epidemiological analysis of lumbar spine degenerative disease diagnosed by MRI: level-wise and sex-based distribution in the Gandaki province population," *Journal of Gandaki Medical College-Nepal*, vol. 18, no. 2, pp. 116–121, 2025. doi: 10.3126/jgmc-n.v18i2.83554.

[7] Z. Xie *et al.*, "Enhanced diagnosis of axial spondyloarthritis using machine learning with sacroiliac

joint MRI: a multicenter study," *Insights into Imaging*, vol. 16, article no. 91, 2025. doi: 10.1186/s13244-025-01967-x.
[8] M. T. Ibrahim, E. Milliron, and E. Yu, "Artificial intelligence in spinal imaging: a narrative review," *Artificial Intelligence Surgery*, vol. 5, no. 1, pp. 139–149, 2025. doi: 10.20517/ais.2024.41.
[9] Y. Wang *et al.*, "Analysis of the correlation between ABO blood type and three bone mineral density measurements in patients with lumbar degenerative diseases," *European Spine Journal*, vol. 34, pp. 1731–1740, 2025. doi: 10.1007/s00586-025-08796-y.
[10] K. Y. Gete *et al.*, "Diagnostic accuracy of deep learning for automated detection of spinal degenerative disease on MRI: a systematic review and meta-analysis," *Journal of Imaging Informatics in Medicine*, online-first, 2026. doi: 10.1007/s10278-026-01897-0.
[11] P. Giaccone *et al.*, "Prevention and management of degenerative lumbar spine disorders through artificial intelligence-based decision support systems: a systematic review," *BMC Musculoskeletal Disorders*, vol. 26, article 126, 2025. doi: 10.1186/s12891-025-08356-x.
[12] P. Chen *et al.*, "Deep learning-based assessment of paraspinal muscle degeneration and its relationships to muscle function and disability outcomes in chronic low back pain: a prospective study," *European Radiology*, vol. 36, pp. 4197–4207, 2026. doi: 10.1007/s00330-025-12171-2.
[13] C. Zou, T. Wang, B. Wang, Q. Fei, H. Song, and L. Zang, "Developing a deep learning radiomics model combining lumbar CT, multi-sequence MRI, and clinical data to predict high-risk adjacent segment degeneration following lumbar fusion: a retrospective multicenter study," *Global Spine Journal*, vol. 16, no. 1, pp. 263–277, 2026. doi: 10.1177/21925682251342531.
[14] L. Pina Vegas *et al.*, "Ten-year follow-up of degenerative spinal lesions on radiographs and MRI in axial spondyloarthritis: results of the DESIR cohort," *European Radiology*, vol. 35, pp. 5381–5391, 2025. doi: 10.1007/s00330-025-11432-4.
[15] F. Beyer, P. Eysel, and J. Bredow, "The diagnosis and treatment of degenerative changes of the lumbar spine," *Deutsches Ärzteblatt International*, vol. 122, no. 9, pp. 249–256, 2025. doi: 10.3238/arztebl.m2025.0056.