

PURPOSE

Determining which specific outputs of multi-pathology AI-algorithms can be reliably integrated into real-world radiology reports remains a major challenge and their actual impact on efficiency is often uncertain. This study describes a real-world validation process for the safe implementation of multi-pathology AI tools and introduces a surrogate indicator of their impact on reporting efficiency.

METHODS AND MATERIALS

We present a detailed case study on the validation and implementation of a multi-pathology lumbar spine MRI algorithm. AI-generated interpretative findings for 26 pathologies were tabulated and evaluated using a representative sample of 50 lumbar spine MRI scans from a single radiology department. Two expert spine radiologists independently evaluated the same scans and resolved discrepant interpretations by consensus, establishing a ground truth:

SAMPLE CASE
In **BLUE** : discrepant interpretations between the 2 experts radiologists resolved by consensus
In **YELLOW**: the Ground Truth
In **RED**: discrepant interpretations between the AI and the Ground Truth

	SAMPLE CASE					SAMPLE CASE					SAMPLE CASE					SAMPLE CASE			
	Radiologist 1	Radiologist 2	Ground Truth	AI		Radiologist 1	Radiologist 2	Ground Truth	AI		Radiologist 1	Radiologist 2	Ground Truth	AI		Radiologist 1	Radiologist 2	Ground Truth	AI
SCOLIOSIS					DISC BULGING					LEFT LIGAMENT FLAVUM HYPERTROPHY					LEFT FORAMINAL NRC				
right	1	1	1	0	SYMMETRICAL					L1/L2	0	1	1	0	L1/L2	0	0	0	0
left	0	0	0	1	L1/L2	0	0	0	0	L2/L3	1	1	1	0	L2/L3	0	0	0	0
STRAIGHTENING	0	0	0	1	L2/L3	1	0	1	1	L3/L4	1	1	1	0	L3/L4	0	0	0	0
HYPERLORDOSIS	0	0	0	0	L3/L4	1	1	1	1	L4/L5	1	1	1	1	L4/L5	0	0	0	0
KYPHOSIS	0	0	0	1	L4/L5	1	1	1	0	L5/S1	0	0	0	0	L5/S1	0	1	0	0
ANTERO-LISTHESIS					L5/S1	1	1	1	0	MILD CENTRAL STENOSIS					DISC HERNIATION WITH NRC				
L1/L2	0	0	0	0	ASYMMETRICAL WITH RIGHT PREDOMINANCE					L1/L2	0	0	0	0	L1/L2	0	0	0	0
L2/L3	0	0	0	0	L1/L2	0	0	0	0	L2/L3	0	1	1	0	L2/L3	1	1	1	1
L3/L4	0	0	0	0	L2/L3	0	0	0	0	L3/L4	0	0	0	0	L3/L4	0	1	0	1
L4/L5	0	0	0	0	L3/L4	0	0	0	0	L4/L5	0	0	0	0	L4/L5	0	1	1	1
L5/S1	0	0	0	0	L4/L5	0	0	0	0	L5/S1	0	0	0	0	L5/S1	0	1	0	1
RETRO-LISTHESIS					L5/S1	0	0	0	0	MODERATE CENTRAL STENOSIS					DISC HERNIATION WITHOUT NRC				
L1/L2	0	0	0	1	ASYMMETRICAL WITH LEFT PREDOMINANCE					L1/L2	0	0	0	0	L1/L2	0	0	0	0
L2/L3	0	0	0	0	L1/L2	0	0	0	0	L2/L3	1	0	0	0	L2/L3	0	0	0	0
L3/L4	0	0	0	0	L2/L3	0	1	0	0	L3/L4	1	0	0	1	L3/L4	0	0	0	0
L4/L5	0	0	0	0	L3/L4	0	0	0	0	L4/L5	1	0	0	1	L4/L5	0	0	0	0
L5/S1	0	0	0	1	L4/L5	0	0	0	0	L5/S1	1	0	0	1	L4/L5	0	0	0	0
DISC DESICCATION					L5/S1	0	0	0	0	SEVERE CENTRAL STENOSIS					L5/S1	0	0	0	0
L1/L2	0	0	0	0	DISC BULGING (REGARDLESS ITS MORPHOLOGY)					L1/L2	0	0	0	0	DISC HERNIATION (REGARDLESS OF ITS IMPACT ON NERVE ROOTS)				
L2/L3	1	1	1	0	L1/L2	0	0	0	0	L2/L3	0	0	0	0	L1/L2	0	0	0	0
L3/L4	1	1	1	1	L2/L3	1	1	1	1	L3/L4	0	0	0	0	L2/L3	1	1	1	1
L4/L5	1	1	1	1	L3/L4	1	1	1	1	L4/L5	0	0	0	0	L3/L4	0	1	0	1
L5/S1	1	1	1	1	L4/L5	1	1	1	0	L5/S1	0	0	0	0	L4/L5	0	1	1	1
LOSS OF DISC HEIGHT					L5/S1	1	1	1	0	RIGHT LATERAL RECESS STENOSIS WITH NRC					L5/S1	0	1	0	1
L1/L2	1	1	1	1	RIGHT FACET JOINT HYPERTROPHY					L1/L2	0	0	0	0	LS VERTEBRAL BODY HYPOPLASIA				
L2/L3	0	1	1	0	L1/L2	0	1	1	0	L2/L3	0	0	0	0	LOSS OF VERTEBRAL BODY HEIGHT				
L3/L4	0	0	0	0	L2/L3	1	1	1	0	L3/L4	1	1	1	1	L1/L2	0	0	0	0
L4/L5	0	0	0	0	L3/L4	1	1	1	0	L4/L5	1	1	1	1	L2/L3	0	0	0	0
L5/S1	0	0	0	0	L4/L5	1	1	1	0	L5/S1	0	0	0	0	L3/L4	0	0	0	0
APOPHYSEAL OSTEOPHYTES					L5/S1	1	1	1	1	LEFT LATERAL RECESS STENOSIS WITH NRC					L4/L5	0	0	0	0
L1/L2	1	1	1	1	LEFT FACET JOINT HYPERTROPHY					L1/L2	0	0	0	0	L5/S1	0	0	0	0
L2/L3	1	1	1	1	L1/L2	0	1	1	0	L2/L3	1	1	1	1	LOCATION OF CONUS MEDULLARIS				
L3/L4	1	1	1	1	L2/L3	1	1	1	0	L3/L4	1	0	1	1	T12	1	1	1	1
L4/L5	0	1	0	1	L3/L4	1	1	1	0	L4/L5	1	1	1	1	L1	0	0	0	0
L5/S1	0	0	0	0	L4/L5	1	1	1	0	L5/S1	0	1	1	1	L2	0	0	0	0
MODIC 1					L5/S1	1	1	1	1	RIGHT FORAMINAL NRC					L3	0	0	0	0
L1/L2	0	1	0	1	RIGHT LIGAMENT FLAVUM HYPERTROPHY					L1/L2	0	0	0	0	L4	0	0	0	0
L2/L3	0	0	0	1	L1/L2	0	1	1	0	L2/L3	0	0	0	0	L5	0	0	0	0
L3/L4	0	0	0	1	L2/L3	1	1	1	0	L3/L4	0	0	0	0	POSTERIOR PARSPINAL MUSCLE ATROPHY				
L4/L5	0	0	0	0	L3/L4	1	1	1	0	L4/L5	0	0	0	0	AORTIC DIAMETER	0	0	0	0
L5/S1	0	0	0	0	L4/L5	1	1	1	1	L5/S1	0	0	0	0					
					L5/S1	1	1	1	1										

AI-interpretive errors were quantified using accuracy, positive predictive value (PPV) and negative predictive value (NPV). Agreement between AI and ground truth was quantified using prevalence-adjusted and bias-adjusted kappa statistics (PABAK). Only AI-findings meeting predefined thresholds of accuracy > 90%, PPV or PNV > 90%, and PABAK > 0.60 were accepted for inclusion into integrated AI-radiologist reports:

	PATHOLOGIES INCLUDED into the AI assisted report	PPV	NPV	Accuracy	PABAK (Ground Trut -AI)	Agreement Interpretation	PATHOLOGIES NOT INCLUDED into the AI assisted report
1	MODERATE CENTRAL STENOSIS	0.33	0.99	0.97	0.94	Near Perfect	11 DISC DESICCATION
2	SEVERE CENTRAL STENOSIS	1.00	1.00	1.00	0.99	Near Perfect	12 LOSS OF DISC HEIGHT
3	RIGHT LATERAL RECESS STENOSIS WITH LIKELY NRC	0.82	0.99	0.98	0.95	Near Perfect	13 LEFT LIGAMENT FLAVUM HYPERTROPHY
	LEFT LATERAL RECESS STENOSIS WITH LIKELY NRC	0.71	0.99	0.96	0.93	Near Perfect	RIGHT LIGAMENT FLAVUM HYPERTROPHY
4	RIGHT FORAMINAL NRC	0.78	0.99	0.98	0.96	Near Perfect	14 SYMMETRICAL
	LEFT FORAMINAL NRC	0.50	0.98	0.96	0.92	Near Perfect	15 ASYMMETRICAL WITH LEFT PREDOMINANCE
5	DISC HERNIATION WITH NRC	0.33	0.99	0.93	0.86	Near Perfect	ASYMMETRICAL WITH RIGHT PREDOMINANCE
6	ANTERO-LISTHESIS	0.88	0.99	0.98	0.97	Near Perfect	16 APOPHYSEAL OSTEOPHYTES
7	RETRO-LISTHESIS	0.35	1.00	0.93	0.86	Near Perfect	17 RIGHT FACET JOINT HYPERTROPHY
8	LOSS OF VERTEBRAL BODY HEIGHT	0.33	1.00	0.99	0.98	Near Perfect	LEFT FACET JOINT HYPERTROPHY
9	POSTERIOR PARSPINAL MUSCLE ATROPHY	0.44	1.00	0.90	0.61	Moderate	18 MILD CENTRAL STENOSIS
10	AORTIC DIAMETER	1.00	1.00	1.00	1.00	Perfect	19 DISC HERNIATION (REGARDLESS OF ITS IMPACT ON NERVE ROO
							20 DISC HERNIATION WITHOUT NRC
							21 STRAIGHTENING
							22 SCOLIOSIS
							23 HYPERLORDOSIS
							24 KYPHOSIS
							25 MODIC 1
							26 L5 VERTEBRAL BODY HYPOPLASIA

After an integrated AI-Radiologist reporting pilot in routine clinical work (7 radiologists,600 reports), the proportion of report text generated by the AI-algorithm was measured as a surrogate indicator of reporting efficiency gain:

AI report output:	AI + RADIOLOGIST final report:
PREVIOUS IMAGING	PREVIOUS IMAGING
Not Available	Not Available
TECHNIQUE	TECHNIQUE
AI_STIR_TSE, AI_T2W_TSE TRA, AI_T1W_TSE SAG, AI_T2W_TSE SAG, SURVEY	AI_STIR_TSE, AI_T2W_TSE TRA, AI_T1W_TSE SAG, AI_T2W_TSE SAG, SURVEY
CONUS MEDULLARIS	CONUS MEDULLARIS
Located at L1.	Located at L1.
ALIGNMENT	ALIGNMENT
Left convex curvature of the lumbar spine in supine position. Straightened lordosis. Grade I retrolisthesis at L5-S1.	Left convex curvature of the lumbar spine in supine position. Straightened lordosis. Grade I retrolisthesis at L5-S1.
FINDINGS	FINDINGS
Normal lumbar vertebral body height.	Normal lumbar vertebral body height. Desiccation, loss of height, and bulging of the discs together with apophyseal osteophytes and facet joint hypertrophy, predominantly of the lower lumbar spine.
L4-L5: Degenerative significant spinal canal stenosis with almost complete effacement of the CSF and cauda equina crowding. No signs suggestive of cauda equina compression.	L4-L5: Degenerative significant spinal canal stenosis with almost complete effacement of the CSF and cauda equina crowding. No signs suggestive of cauda equina compression.
L5-S1: Degenerative significant spinal canal stenosis with almost complete effacement of the CSF and cauda equina crowding. No signs suggestive of cauda equina compression.	L5-S1: Degenerative significant spinal canal stenosis with almost complete effacement of the CSF and cauda equina crowding. No signs suggestive of cauda equina compression.
POSTERIOR PARSPINAL MUSCLES	POSTERIOR PARSPINAL MUSCLES
No relevant paraspinal muscle change.	No relevant paraspinal muscle change.
CONCLUSION	CONCLUSION
	Degenerative lumbar spine changes. Retrolisthesis at L5-S1. Moderate spinal canal stenosis at L4-5 and L5-S1 with crowding of the cauda equina, but without signs of cauda equina compression.

RESULTS

The interpretive findings of 10 of the 26 pathologies met the predefined accuracy and inter-rater agreement thresholds and were incorporated into integrated AI-Radiologist reports. These included: moderate and severe spinal canal stenosis, lateral recess stenosis with nerve root compression, foraminal stenosis with nerve root compression, disc herniation with nerve root compression, antero- and retrolisthesis, loss of vertebral body height, paraspinal muscle atrophy, abdominal aorta dilatation. The AI-generated component constituted in average approximately 25% of the integrated AI-Radiologist reports.

CONCLUSION

Pathology-tailored accuracy and prevalence-adjusted variability measurements of independent data sets, combined with a real-world integrated AI-radiologist reporting pilot, support the safe clinical implementation of multi-pathology AI-algorithms and enable controlled adoption of integrated AI-Radiologist reports. We suggest the proportion of report content generated by AI may be a valid surrogate indicator of reporting efficiency gains.