European Journal of Radiology 200 (2026) 112868

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Node-RADS v1.0 on chest CT for lung cancer: Reproducibility and diagnostic performance

Lorenzo Cereser [a,b,*], Tin Nadarević [c], Consuelo Ciancimino [a], Eugenia Versienti [a], Antonia Pia Pace [a], Giuseppe Como [b], Alessandro De Pellegrin [d], Francesco Cortiula [e,f], Giuseppe Aprile [e], Chiara Zuiani [a,b], Rossano Girometti [a,b]

[a] Institute of Radiology, Department of Medicine (DMED), University of Udine, Italy
[b] Institute of Radiology, University Hospital S. Maria Della Misericordia, Azienda Sanitaria Universitaria Friuli Centrale (ASUFC), Udine, Italy
[c] Department of Diagnostic and Interventional Radiology, Clinical Hospital Center Rijeka, University of Rijeka, Rijeka, Croatia
[d] Department of Pathology, University Hospital S. Maria Della Misericordia, Azienda Sanitaria-Universitaria Friuli Centrale (ASUFC), Udine, Italy
[e] Department of Oncology, University Hospital S. Maria Della Misericordia, Azienda Sanitaria Universitaria Friuli Centrale (ASUFC), Udine, Italy
[f] Department of Radiation Oncology (Maastro), GROW School for Oncology and Reproduction, Maastricht University Medical Centre, Italy

ARTICLE INFO

Keywords:
Non-small cell lung cancer
Computed tomography
Lymph nodes
Node-RADS
Observer variability

ABSTRACT

Purpose: To evaluate intra- and inter-reader agreement of Node-RADS v1.0 for mediastinal lymph node assessment on chest CT in patients with stage I–III non–small cell lung cancer (NSCLC) and to assess its diagnostic performance across radiologists with different levels of expertise.

Methods: This retrospective, single-center study included 46 patients (mean age, 71 ± 8 years; 38 adenocarcinomas, 8 squamous cell carcinomas) with 158 pathologically confirmed mediastinal lymph nodes (22 malignant, 136 benign). Four radiologists (two experts, two juniors) independently assigned Node-RADS scores and descriptors ("size" and "configuration") in two sessions, three weeks apart. Agreement was calculated using Gwet's AC2 statistics. Diagnostic performance was assessed by ROC analysis; sensitivity, specificity, and predictive values were calculated at a Node-RADS score ≥ 3 threshold.

Results: Inter-reader agreement for Node-RADS scores was almost perfect for experts (AC2 = 0.97; 95%CI: 0.96–0.99) and juniors (AC2 = 0.95; 95%CI: 0.93–0.97). Intra-reader agreement AC2 values ranged from 0.95 to 0.99. Descriptor agreement was similarly high (AC2 ≥ 0.85). ROC AUCs ranged from 0.71 to 0.76 for experts and 0.68–0.84 for juniors. At the ≥ 3 threshold, specificity and negative predictive value were consistently ≥ 90%, while sensitivity remained limited (<64%).

Conclusions: Node-RADS v1.0 demonstrated excellent reproducibility among radiologists with different levels of expertise for mediastinal lymph node assessment on chest CT in NSCLC. The consistently high specificity and negative predictive value support its role as a standardized framework for structured lymph node reporting and training, while the limited sensitivity underscores the need for complementary diagnostic tools.

1. Introduction

Lung cancer (LC) is the leading cause of cancer-related death worldwide. The term LC refers to a heterogeneous group of malignancies comprising over 50 histopathological subtypes, with non-small cell lung cancer (NSCLC) accounting for approximately 80–85% of cases [1]. LC

Abbreviations: LC, lung cancer; NSCLC, non-small cell lung cancer; AJCC, American Joint Committee on Cancer; UICC, Union for International Cancer Control; TNM, Tumor-Node-Metastasis; CT, computed tomography; 18F-FDG PET/CT, 18F-Fluorodeoxyglucose positron emission tomography/CT; EBUS-TBNA, endobronchial ultrasound-guided transbronchial needle aspiration; Node-RADS v1.0, Node Reporting and Data System; GRRAS, Guidelines for Reporting Reliability and Agreement Studies; IASLC, International Association for the Study of Lung Cancer; R1–R4, radiologists 1–4; MDTs, multidisciplinary teams; PACS, Picture Archiving and Communication System; IQR, interquartile range; PA, percent agreement; k, Cohen's kappa; CI, confidence intervals; ROC, receiver operating characteristic; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value.

* Corresponding author at: Institute of Radiology, Department of Medicine (DMED), University of Udine, University Hospital S. Maria della Misericordia, Azienda Sanitaria Universitaria Friuli Centrale (ASUFC), p.le Santa Maria della Misericordia, 15 – 33100 Udine, Italy.
E-mail address: lorenzo.cereser@uniud.it (L. Cereser).

https://doi.org/10.1016/j.ejrad.2026.112868
Received 31 October 2025; Received in revised form 8 February 2026; Accepted 12 April 2026
Available online 15 April 2026

staging follows the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC) developed Tumor-Node-Metastasis (TNM) classification system. The 9th edition of the TNM staging system is currently used in clinical practice to estimate prognosis and inform personalized treatment strategies [2].

Imaging plays a key role in TNM staging of LC, with contrast-enhanced chest computed tomography (CT) representing the reference modality for morphological cross-sectional imaging. CT is usually complemented by 18F-Fluorodeoxyglucose positron emission tomography/CT (18F-FDG PET/CT) and endoscopic or minimally invasive surgical procedures (e.g., bronchoscopy, endoscopic ultrasound, or mediastinoscopy) for histopathological confirmation [3]. Regarding the N component of the TNM system, the assessment of mediastinal lymph node involvement on CT has traditionally been based solely on size criteria. Lymph nodes exceeding 10 mm short-axis diameter in the axial plane are considered suspicious. As expected, this criterion has low diagnostic accuracy, with a sensitivity of 55% and specificity of 81% [4]. Therefore, the evaluation of the N parameter must be complemented by additional investigations, including 18F-FDG PET/CT, endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA), and mediastinoscopy [5]. These additional tests are required when the disease appears localized at CT, to confirm eligibility for curative treatments.

In 2021, a multidisciplinary working group introduced the first version of the Node Reporting and Data System (Node-RADS v1.0) [6]. Node-RADS aims to standardize CT reporting of lymph node involvement by combining two morphological descriptors, i.e., size and configuration, into a 5-point score reflecting the likelihood of malignancy. Scores range from 1 ("very low probability") to 5 ("very high probability") of malignant involvement.

Although Node-RADS has potential applications in multiple malignancies [7], only one study has investigated its use in LC [8]. That study demonstrated that Node-RADS could differentiate between N0 and N1–3 stages ($p < 0.001$), although the inter-reader agreement was only moderate. Given this background, performing a reader agreement analysis for Node-RADS scoring in the CT-based mediastinal lymph node evaluation is essential before its broader clinical adoption.

The primary study aim was to evaluate intra- and inter-reader agreement in assigning Node-RADS v1.0 scores for mediastinal lymph node assessment in patients with LC undergoing chest CT for staging. Secondary aims were to assess agreement for individual Node-RADS descriptors (size and configuration) and explore the diagnostic performance of the system for predicting pathologic nodal involvement.

2. Materials and methods

Our Institutional Review Board approved the study (Prot. IRB: 019/2025) and waived the requirement for informed consent from patients, given the retrospective design.

Overall, the study was conducted following the Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [9] and coordinated by a radiologist who was not involved in image interpretation. This radiologist handled case selection, dataset preparation, nodal matching with pathology, random selection and reordering of cases for the second reading session, and worksheet preparation, with access to pathological data limited to reference standard assignment. The study design is illustrated in Fig. 1.

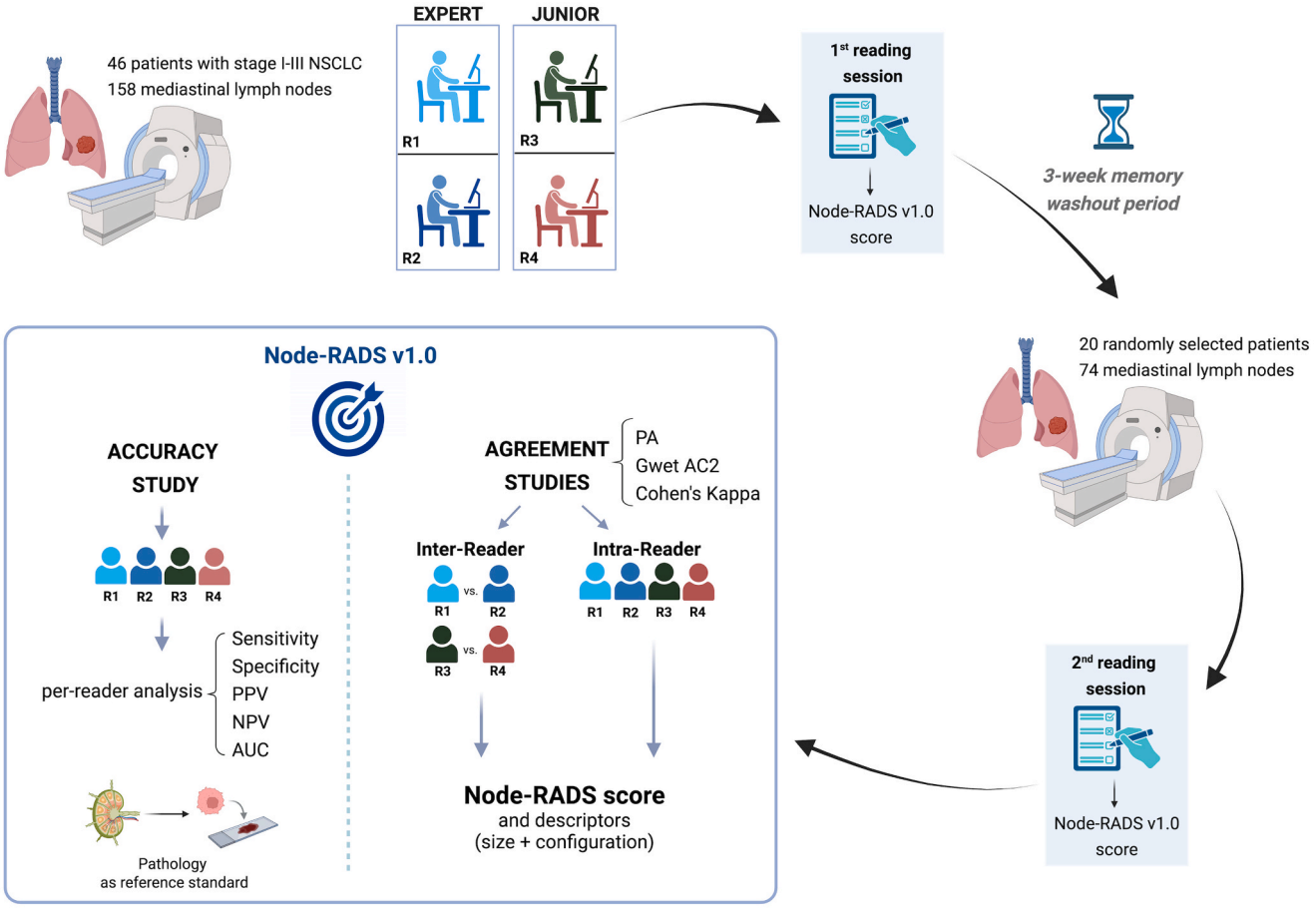


Fig. 1. The study design *Notes*: NSCLC, non-small cell lung cancer. R1, reader 1; R2, reader 2; R3, reader 3; R4, reader 4. PA, percent agreement. PPV, positive predictive value. NPV, negative predictive value. AUC, area under the curve Figure created with BioRender.com.

2.1. Study population

The study cohort was derived from an institutional oncology database and included 106 consecutive patients with pathologically confirmed stage I-III NSCLC evaluated at our tertiary referral university hospital between 2022 and 2023. All cases were classified according to the 8th edition of the TNM staging system [10] and were treatment-naïve with respect to any neoadjuvant therapy.

Sixty patients were excluded for the following reasons: absence of histological or cytological confirmation of mediastinal lymph nodes obtained via EBUS-TBNA or surgery (n = 28), unavailability of baseline contrast-enhanced chest CT (n = 18), baseline CT performed without intravenous contrast (n = 7), CT performed more than two months before histological or cytological examination (n = 5), and missing clinical data (n = 2).

After application of these criteria, the final study cohort comprised 46 patients (20 men and 26 women; median age, 71 years; range, 52–83 years). NSCLC histology was adenocarcinoma in 38 patients (82.6%) and squamous cell carcinoma in the remaining 8 patients (17.4%). The distribution of patients across NSCLC stages was as follows: stage IA, 8/46 (17%); stage IB, 11/46 (23.9%); stage IIA, 2/46 (4.3%); stage IIB, 11/46 (23.9%); stage IIIA, 5/46 (11%); and stage IIIB, 9/46 (19.6%). Patient selection is summarized in Supplementary Fig. 1.

2.2. Reference standard

On a per-lymph node basis, pathological assessment was performed using surgical specimen histology for 133 of 158 lymph nodes (84%). For the remaining 25 lymph nodes (16%), findings from EBUS-TBNA were used in patients who did not undergo surgery. The interval between CT examination and cytological or histological sampling showed a mean of 27 days and a median of 33 days.

2.3. CT examination technique

All CT examinations were performed using a 64-row multidetector CT scanner (Discovery CT750 HD or Optima CT660, GE Healthcare, Milwaukee, WI, USA) with patients in supine position at full inspiration during the venous phase (70–90 s after contrast injection). An iodinated contrast agent (Iomeprol or Iobitridol, Bracco, Milan, Italy; iodine concentration, 350–400 mg/mL) was administered according to the institutional protocol. Key acquisition parameters included submillimetric collimation and thin-slice reconstruction; detailed technical settings are provided in Supplementary Table 1.

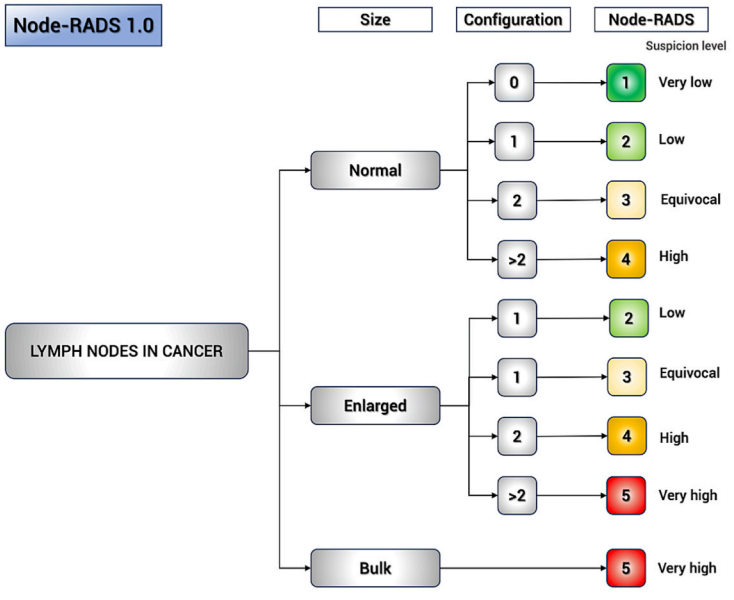
2.4. Image analysis

For each case, the study coordinator annotated, in a dedicated worksheet, all mediastinal lymph node stations for which cytological or histological data were available. Lymph node stations were named according to the International Association for the Study of Lung Cancer (IASLC) lymph node map [11].

Four radiologists (R1–R4) with varying experience levels in chest CT imaging and involvement in multidisciplinary teams (MDTs) for oncology were recruited. Specifically, R1 and R2 had more than 10 years of experience and regular participation in MDTs (hereinafter referred to as *expert readers*); R3 and R4 were senior radiology residents with approximately 300 interpreted chest CT examinations and initial MDT attendance (hereinafter referred to as *junior readers*). Notably, R2 was from an external institution, whereas R1, R3, and R4 were from the coordinating center.

Each reader received an evaluation form indicating the mediastinal stations to be reviewed for each case, along with a brief checklist describing the criteria for Node-RADS scoring (Fig. 2). When multiple lymph nodes were present in the same assigned station, readers evaluated the most suspicious node based on size and/or shape, i.e., one lymph node per station. Nodal size was assessed as the short-axis diameter measured on axial images; multiplanar reconstructions were used to assist lymph node identification and assessment of configuration features, in accordance with the Node-RADS framework [6].

The readers completed two independent reading sessions, separated by a three-week washout interval. In the first session, each reader



SIZE *Choose one category*		
Normal *Short axis*	**Enlarged** *Short axis*	**Bulk** *Any axis*
• General <10 mm • Inguinal <15 mm • Facial, parotid, retropharyngeal, anterior jugular <5 mm • Cardio-phrenic, retrocrural, obturator, mesorectal <5 mm	• Larger than normal but no bulk • ≥2 mm increase to prior imaging if available	• ≥30 mm

CONFIGURATION *Choose one feature per category, sum scores*		
Texture	**Border**	**Shape**
• 0: Homogeneous • 1: Heterogeneous • 2: Focal necrosis • 3: Gross necrosis or any new necrosis • 3: Entity specific findings	• 0: Smooth • 1: Irregular or ill-defined	• 0: Any shape with preserved fatty hilum • 0: Kidney-bean-like or oval without fatty hilum • 1: Spherical without fatty hilum

Fig. 2. The checklist describing the criteria for Node-RADS scoring (modified from reference [6]).

assigned a Node-RADS score to the most suspicious lymph node in every designated mediastinal station across all CT examinations. In the second session, after the study coordinator randomly selected and reordered 20 out of 46 CT cases, the same readers reassigned scores to the same lymph nodes (n = 74).

All CT examinations were reviewed using a dedicated Picture Archiving and Communication System (PACS) workstation (Suitestensa, Ebit Srl, Esaote Group Company, Genoa, Italy).

2.5. Statistical analysis

Continuous variables are presented as mean ± standard deviation or median and interquartile range (IQR), as appropriate, after testing normality using the Kolmogorov-Smirnov test. Categorical variables are presented as counts and percentages.

Node-RADS score assignments were summarized descriptively for each reader. Intra- and inter-reader agreement in Node-RADS scoring was calculated between expert (R1–R2) and junior (R3–R4) readers using percent agreement (PA), Cohen's κ, and Gwet's AC2 with quadratic weights, with 95% confidence intervals (CI). Because of the low prevalence of pathology-proven malignant nodes (≈14%), Gwet's AC2 was prioritized over κ due to its robustness to prevalence imbalance and skewed marginals [12,13]. Corresponding κ values were also reported for completeness. Agreement values were interpreted according to Landis and Koch [14].

Secondary analyses evaluated agreement for the individual Node-RADS descriptors "size" and "configuration" using the same methods. The "size" descriptor was categorized as normal, enlarged, or bulky, and "configuration" derived from combined texture, borders, and shape criteria [6].

Diagnostic performance of Node-RADS in predicting lymph node malignancy was assessed by receiver operating characteristic (ROC) curve analysis, with area under the curve (AUC) values compared among readers using the DeLong test. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated per reader using a threshold of ≥ 3 as indicative of malignancy. Pairwise comparisons of sensitivity and specificity between readers were performed using logistic regression models, with p-value (p) < 0.05 considered statistically significant.

Two exploratory analyses were performed in the expert readers: (i) evaluation of diagnostic performance using a lower Node-RADS threshold (≥2) and (ii) comparison of Node-RADS ≥ 3 with a stand-alone size-based criterion (short-axis ≥ 10 mm). Diagnostic performance using the lower Node-RADS threshold (≥2) was evaluated to illustrate the sensitivity–specificity trade-off and to allow contextual comparison with the study by Meyer et al. [8]. Diagnostic performance of Node-RADS ≥ 3 was compared with a stand-alone size-based criterion, defined as a lymph node short-axis diameter ≥ 10 mm, using the same node-level pathological reference standard.

All analyses were conducted using Stata/BE, version 19.5 (StataCorp LLC, College Station, TX, USA).

3. Results

3.1. Lymph node dataset description

The agreement analyses were performed on a total of 158 mediastinal lymph nodes with a pathological reference standard, including 22 (13.9%) malignant and 136 (86.1%) benign nodes. Of the 46 patients included in the study, 18 patients (39%) presented at least one pathologically confirmed malignant lymph node.

Pathologically evaluated lymph nodes were distributed across mediastinal stations as follows (with bilateral stations considered together): station 2, 13 nodes (8%); station 4, 25 nodes (16%); station 5, 6 nodes (4%); station 6, 1 node (1%); station 7, 31 nodes (20%); station 8, 12 nodes (8%); station 9, 10 nodes (6%); station 10, 29 nodes (18%); station 11, 24 nodes (15%); and station 12, 7 nodes (4%).

The distribution of lymph node short-axis size, based on 158 values each representing the mean of the R1-R2 measurements, was as follows: median, 6 mm; IQR, 5–8 mm; and range, 1–47 mm.

Inter- and intra-reader agreement of Node-RADS v.1 scores.

Table 1 reports each reader's distribution of Node-RADS v.1 scores.

In the first reading session, the inter-reader agreement for Node-RADS v.1 scores was almost perfect for both the expert readers (*R1-R2*, PA = 0.98, Gwet's AC2 = 0.97, 95% CI: 0.96–0.99, k = 0.68, 95% CI: 0.51–0.86) and the junior readers (*R3-R4*, PA = 0.97, Gwet's AC2 = 0.95, 95% CI: 0.93–0.97, k = 0.63, 95% CI: 0.48–0.79).

The intra-reader agreement results for Node-RADS v.1 are shown in Table 2. The agreement was almost perfect for expert and junior readers, with Gwet's AC2 ranging from 0.95 to 0.99.

Examples cases with different Node-RADS v.1 score attributions are presented in Figs. 3 and 4.

Inter- and intra-reader agreement of Node-RADS v.1 descriptors.

In the first reading session, the inter-reader agreement for lymph node size and configuration was almost perfect for both expert and junior readers, with Gwet's AC2 ranging from 0.89 to 0.98 and 0.95 to 0.97, respectively. Table 3 summarizes all inter-reader agreement results for the Node-RADS descriptors.

The intra-reader agreement was almost perfect for expert and junior readers, with Gwet's AC2 ranging from 0.85 to 1.00 for lymph node size and from 0.97 to 1.00 for lymph node configuration. Detailed results for each reader are reported in Supplementary Table 2.

3.2. Diagnostic performance of the node-RADS score

Table 4 summarizes the diagnostic performance metrics for each reader.

The AUC values from ROC curve analyses ranged from 0.71 to 0.76 among expert readers and from 0.68 to 0.84 among junior readers. The only statistically significant difference observed was between the two junior readers ($p = 0.002$).

After dichotomizing Node-RADS scores at a threshold of 3, specificity and NPV were consistently at or above 90% for both expert and junior readers, whereas sensitivity and PPV remained below 64% across all readers. No statistically significant differences were observed between readers in terms of sensitivity or specificity (all $p > 0.05$).

When applying a lower Node-RADS threshold (≥2), sensitivity among expert readers ranged from 59.1% to 63.6%, with a corresponding specificity range of 75.7% to 88.1%.

In an exploratory comparison with conventional size-based nodal assessment, Node-RADS ≥ 3 showed higher sensitivity among expert readers (56–63%) with slightly lower but still high specificity (91%) compared with a stand-alone size criterion (short-axis ≥ 10 mm), which yielded sensitivities of 39–44% and specificities of 95–98%.

Table 1
Distribution of Node-RADS v.1 scores according to the readers in the first reading session.

	Expert readers		Junior readers	
	R1	R2	R3	R4
Node-RADS v.1 score 1	129 (81.7%)	111 (70.2%)	122 (77.2%)	122 (77.2%)
Node-RADS v.1 score 2	12 (7.6%)	29 (18.3%)	18 (11.4%)	12 (7.64%)
Node-RADS v.1 score 3	9 (5.7%)	11 (7.0%)	5 (3.2%)	11 (7.0%)
Node-RADS v.1 score 4	4 (2.5%)	2 (1.3%)	4 (2.5%)	6 (3.8%)
Node-RADS v.1 score 5	4 (2.5%)	5 (3.2%)	9 (5.7%)	7 (4.4%)

Footnotes: R1, reader 1; R2, reader 2; R3, reader 3; R4, reader 4.

Table 2
Intra-reader agreement results for Node-RADS v.1.

		PA	Gwet's AC2	95% CI	Interpretation	Cohen's Kappa	95% CI	Interpretation
Expert readers	R1	0.99	0.99	0.98 – 1.00	almost perfect	0.95	0.87 – 1.00	almost perfect
	R2	0.99	0.98	0.96 – 1.00	almost perfect	0.84	0.68 – 1.00	almost perfect
Junior readers	R3	0.96	0.95	0.90 – 1.00	almost perfect	0.73	0.48 – 0.99	substantial
	R4	1.00	0.99	0.99 – 1.00	almost perfect	0.97	0.94 – 1.00	almost perfect

Footnotes: R1, reader 1; R2, reader 2; R3, reader 3; R4, reader 4; PA, percent agreement; 95% CI, 95% confidence interval.

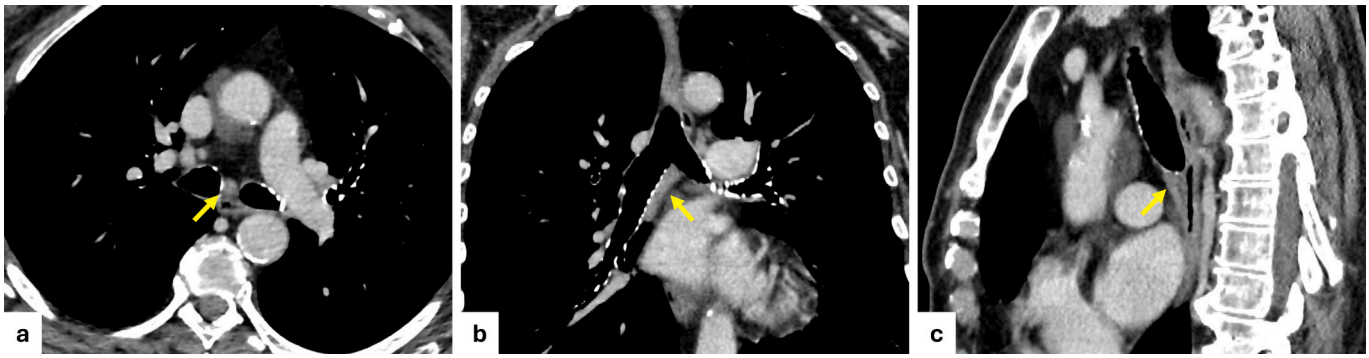


Fig. 3. Example of disagreement between expert readers (R1 and R2) in the assessment of a subcarinal lymph node (station 7), shown on axial **(a)**, coronal **(b)**, and sagittal **(c)** CT images (*arrows*). R1 assigned a Node-RADS score of 1 (size: normal; configuration: 0), whereas R2 assigned a score of 4 (size: enlarged; configuration: 2, due to heterogeneous texture and irregular borders). Histological and cytological analyses confirmed the node was benign.

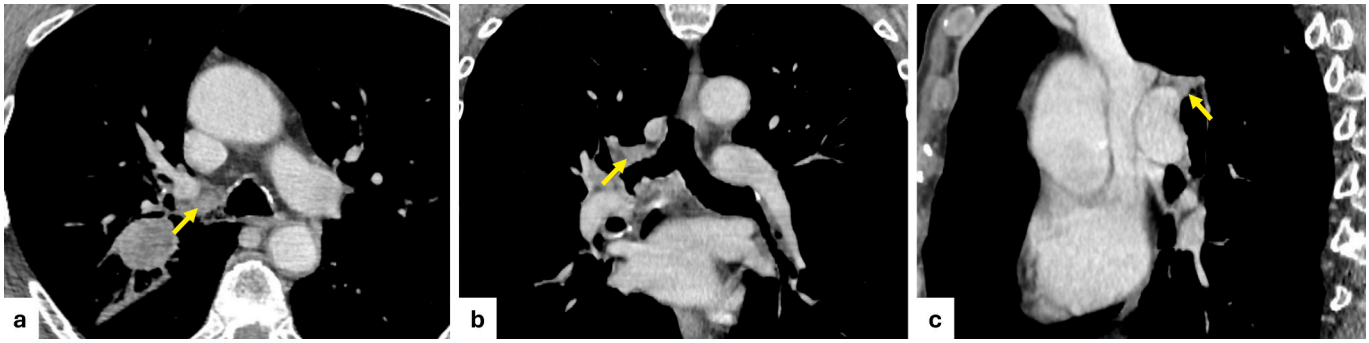


Fig. 4. Example of agreement among readers in the assessment of a right hilar lymph node (station 10R), shown on axial **(a)**, coronal **(b)**, and sagittal **(c)** CT images (*arrows*). All four readers, regardless of expertise level, assigned a Node-RADS score of 4 or 5 (size: enlarged; configuration: ≥2). Cytological analysis confirmed malignancy.

Table 3
Inter-reader agreement results for the Node-RADS v.1 descriptors.

		PA	Gwet's AC2	95% CI	Interpretation	Cohen's Kappa	95% CI	Interpretation
SIZE								
Expert readers	R1	0.99	0.98	0.98 – 0.99	almost perfect	0.60	0.38 – 0.82	moderate
	R2							
Junior readers	R3	0.95	0.89	0.86 – 0.91	almost perfect	0.22	0.06 – 0.38	fair
	R4							
CONFIGURATION								
Expert readers	R1	0.98	0.97	0.96 – 0.99	almost perfect	0.68	0.51 – 0.86	substantial
	R2							
Junior readers	R3	0.97	0.95	0.93 – 0.97	almost perfect	0.63	0.48 – 0.79	substantial
	R4							

Footnotes: R1, reader 1; R2, reader 2; R3, reader 3; R4, reader 4; PA, percent agreement; 95% CI, 95% confidence interval.

4. Discussion

This study evaluated inter- and intra-reader agreement and diagnostic performance of the Node-RADS v1.0 score and its descriptors on chest CT for mediastinal lymph node assessment in patients with stage I–III NSCLC. Node-RADS v1.0 demonstrated almost perfect inter- and intra-reader agreement among both expert and junior readers (Gwet's AC2 ≥ 0.95), indicating excellent reproducibility independent of reader experience. Agreement for the individual descriptors, namely lymph node size and configuration, was likewise high (AC2 range, 0.85–1.00),

Table 4
Diagnostic performance metrics of the Node-RADS v.1 score.

		Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	AUC 0 – 1 (95% CI)
Expert readers	**R1**	63 (35 – 85)	92 (86 – 96)	46 (24 – 68)	96 (91 – 98)	0.76 (0.64 – 0.87)
	R2	56 (31 – 79)	91 (86 – 96)	46 (24 – 68)	94 (89 – 97)	0.71 (0.59 – 0.84)
Junior readers	**R3**	58 (37 – 78)	94 (89 – 97)	64 (41 – 83)	93 (87 – 96)	0.68 (0.56 – 0.80)
	R4	44 (22 – 69)	90 (84 – 94)	36 (17 – 59)	93 (87 – 96)	0.84 (0.74 – 0.94)

Footnotes: R1, reader 1; R2, reader 2; R3, reader 3; R4, reader 4; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve; 95% CI, 95% confidence interval.

supporting the robustness and internal consistency of the scoring system. Overall, these findings suggest that Node-RADS v1.0 can provide referring physicians with a clear and structured representation of malignancy likelihood, while also serving as an educational tool that formalizes the cognitive process of nodal evaluation.

To our knowledge, only one prior study has investigated inter-reader agreement for Node-RADS v.1.0 [8], reporting moderate agreement (k = 0.48) between two relatively inexperienced radiologists (with up to three years of experience in CT imaging). Direct comparison, however, is limited by methodological differences. In our analyses, we prioritized Gwet's AC2 to mitigate the underestimation bias associated with unbalanced datasets [12,13], in line with prior evaluations of various ordinal scoring systems in medical imaging [15–18]. This approach resulted in an almost perfect agreement across all analyses. For comparison with Meyer et al. [8], weighted κ values also confirmed substantial to almost perfect agreement across readers.

The exclusion of stage IV NSCLC cases was an intentional methodological choice aimed at reproducing the clinical scenario in which CT-based nodal assessment can influence management decisions, such as surgical planning or curative-intent therapy. Including stage IV cases would have disproportionately increased high Node-RADS categories (4–5) and introduced a case mix not representative of the intended clinical context. Although this restriction may have reduced disease variability and contributed to the high overall agreement, it ensures that reproducibility was evaluated within the framework where Node-RADS is most likely to be applied in practice. The relatively low prevalence of malignant nodes in our series (approximately 14%) may also have inflated agreement metrics, a recognized limitation of reproducibility studies with skewed distributions.

The high inter- and intra-reader agreement we observed for Node-RADS size and configuration determinants lies within a context of variable results reported in previous literature. For example, while some CT-based Node-RADS studies found fair to moderate inter-reader agreement for morphological features [19], others reported substantial agreement [20]. Standardized CT protocols, optimizing technical parameters such as image thickness and windowing, may have contributed to counterbalancing the inherent subjectivity in assessing nodal size and morphology [21].

The diagnostic performance of the Node-RADS v.1 was moderate overall, with AUC values ranging from 0.68 to 0.84. At a threshold of ≥ 3, specificity and NPV consistently exceeded 90%, whereas sensitivity and PPV remained below 64%. This pattern mirrors prior findings in NSCLC [8] and in a recent meta-analysis of Node-RADS across tumor types [7], underscoring the known trade-off of morphology-based criteria: high specificity at the expense of limited sensitivity for microscopic disease. Comparable performance has been reported for 18F-FDG PET/CT, which yielded pooled sensitivity and specificity of 0.68 and 0.93, respectively [22]. The consistently high NPV at a threshold of ≥ 3 (93–96%) reinforces Node-RADS v1.0 as a reliable framework for excluding malignancy in low-risk lymph nodes, potentially reducing the need for invasive staging procedures. Similar benefit was previously demonstrated in other malignancies such as cervical [23], breast [24], and bladder [25] cancers.

When alternative thresholds were explored, lowering the Node-RADS threshold to ≥ 2 provided only a limited sensitivity gain (59–64%) at the cost of a substantial reduction in specificity (76–88%) compared with the ≥ 3 threshold (sensitivity 56–63%, specificity 91–92%). Meyer et al. reported a sensitivity of 74% and a specificity of 93% using a threshold of 2 in a different context, likely reflecting differences in study design and patient-level analysis [8]; in our cohort, this trade-off supports the use of a ≥ 3 threshold as a more balanced and clinically pragmatic choice. Beyond threshold selection, in an exploratory node-level comparison with stand-alone size-based nodal assessment, Node-RADS ≥ 3 consistently improved sensitivity across expert readers at the expense of a limited reduction in specificity. These findings support the added diagnostic value of integrating nodal configuration beyond size alone. Future studies specifically designed to evaluate the per-patient clinical impact of Node-RADS, including its integration with PET-CT, are warranted.

The main limitation of our study is its single-center design, with all examinations acquired on the same CT scanner. Although the inclusion of an external reader likely mitigated the risk of overestimating agreement due to shared institutional practices, our findings warrant validation in a multi-institutional setting. The relatively small sample size (46 patients) and low proportion of pathologically confirmed malignant lymph nodes (22/158, 14%), reflecting a pathology-enriched cohort, may have affected the robustness of diagnostic performance estimates, particularly sensitivity, and may have contributed to an overestimation of agreement metrics. Consequently, the generalizability of our results may be limited in unselected clinical populations or in settings with a higher burden of nodal metastases.

Finally, we did not integrate the reader-based lymph node analysis with artificial intelligence (AI)-driven radiomics or clinical data. Future multimodal approaches combining structured visual assessment with complementary quantitative and clinical information may enhance the performance of the Node-RADS scoring system in NSCLC, as suggested by promising results in gastric [26] and rectal [27] cancer. Incorporating hybrid models that merge tumor characteristics with mediastinal features, such as pericardial fat radiomic markers, might also improve the detection of occult nodal involvement [28].

In conclusion, Node-RADS v1.0 demonstrated excellent reproducibility among readers with different expertise levels for mediastinal lymph node evaluation on chest CT in stage I–III NSCLC. The diagnostic performance observed in this study aligns with existing literature, further supporting Node-RADS as a standardized framework for reliable lymph node reporting in oncologic imaging. Beyond its clinical applicability, the system also holds potential educational value, providing a structured approach that can facilitate training, harmonize interpretation across expertise levels, and promote consistency in reporting practice.

CRediT authorship contribution statement

Lorenzo Cereser: Writing – original draft, Validation, Methodology, Formal analysis, Conceptualization. **Tin Nadarević:** Writing – review & editing, Methodology, Investigation. **Consuelo Ciancimino:** Writing – original draft, Visualization, Data curation. **Eugenia Versienti:** Writing – review & editing, Investigation. **Antonia Pia Pace:** Writing – review & editing, Investigation. **Giuseppe Como:** Writing – review & editing, Investigation. **Alessandro De Pellegrin:** Writing – original draft, Data curation. **Francesco Cortiula:** Writing – original draft, Data curation.

Giuseppe Aprile: Writing – review & editing, Supervision. **Chiara Zuiani:** Writing – review & editing, Project administration. **Rossano Girometti:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejrad.2026.112868.

References

[1] G.J. Riely, D.E. Wood, D.S. Ettinger, D.L. Aisner, W. Akerley, J.R. Bauman, et al., Non-small cell lung cancer, version 4.2024, NCCN clinical practice guidelines in oncology, J. Natl. Compr. Canc. Netw. 22 (2024) 249–274, https://doi.org/10.6004/jnccn.2204.0023.

[2] M. Klug, Z. Kirshenboim, M.T. Truong, V. Sorin, E. Ofek, R. Agrawal, et al., Proposed ninth edition TNM staging system for lung cancer: guide for radiologists, Radiographics 44 (2024) e240057. https://doi.org/10.1148/rg.240057.

[3] J. Remon, J.C. Soria, S. Peters, Early and locally advanced non-small-cell lung cancer: update of the ESMO clinical practice guidelines, Ann. Oncol. 32 (2021) 1637–1642, https://doi.org/10.1016/j.annonc.2021.08.1994.

[4] G.A. Silvestri, A.V. Gonzalez, M.A. Jantz, M.L. Margolis, M.K. Gould, L.T. Tanoue, et al., Methods for staging non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines, Chest 143 (2013) e211S–e250S, https://doi.org/10.1378/chest.12-2355.

[5] National Comprehensive Cancer Network, NCCN clinical practice guidelines in oncology: non-small cell lung cancer, Version 7.2025. https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf (accessed 30 Oct 2025).

[6] F.H.J. Elsholtz, P. Asbach, M. Haas, M. Becker, R.G.H. Beets-Tan, H.C. Thoeny, et al., Introducing the node reporting and data system 1.0 (Node-RADS): A concept for standardized assessment of lymph nodes in cancer, Eur. Radiol. 31 (2021) 6116–6124, https://doi.org/10.1007/s00330-020-07572-4.

[7] J. Zhong, S. Mao, H. Chen, Y. Wang, Q. Yin, Q. Cen, et al., Node-RADS: A systematic review and meta-analysis of diagnostic performance, category-wise malignancy rates, and inter-observer reliability, Eur. Radiol. 35 (2025) 2723–2735, https://doi.org/10.1007/s00330-024-11160-1.

[8] H.J. Meyer, B. Schnarkowski, J. Pappisch, T. Kerkhoff, H. Wirtz, A.K. Höhn, et al., CT texture analysis and Node-RADS CT score of mediastinal lymph nodes: diagnostic performance in lung cancer patients, Cancer Imaging 22 (2022) 75, https://doi.org/10.1186/s40644-022-00506-x.

[9] J. Kottner, L. Audigé, S. Brorson, A. Donner, B.J. Gajewski, A. Hróbjartsson, et al., Guidelines for reporting reliability and agreement studies (GRRAS) were proposed, J. Clin. Epidemiol. 64 (2011) 96–106, https://doi.org/10.1016/j.jclinepi.2010.03.002.

[10] O. Lababede, M.A. Meziane, The eighth edition of TNM staging of lung cancer: reference chart and diagrams, Oncologist 23 (2018) 844–848. https://doi.org/10.1634/theoncologist.2017-0659.

[11] A.H. El-Sherief, C.T. Lau, C.C. Wu, R.L. Drake, G.F. Abbott, T.W. Rice, International association for the study of lung cancer (IASLC) lymph node map: Radiologic review with CT illustration, Radiographics 34 (2014) 1680–1691, https://doi.org/10.1148/rg.346130097.

[12] A. Benomar, E. Zarour, L. Létourneau-Guillon, J. Raymond, Measuring interrater reliability, Radiology 309 (2023) e230492, https://doi.org/10.1148/radiol.230492.

[13] K.L. Gwet, Computing inter-rater reliability and its variance in the presence of high agreement, Br. J. Math. Stat. Psychol. 61 (2008) 29–48, https://doi.org/10.1348/000711006X126600.

[14] J.R. Landis, G.G. Koch, The measurement of observer agreement for categorical data, Biometrics 33 (1977) 159–174.

[15] A. Ponsiglione, L. Cereser, E. Spina, L. Mannacio, D. Negroni, L. Russo, et al., PI-QUAL version 2: A multi-reader reproducibility study on multiparametric MRI from a tertiary referral center, Eur. J. Radiol. 181 (2024) 111716, https://doi.org/10.1016/j.ejrad.2024.111716.

[16] A. Cozzi, K. Pinker, A. Hidber, T. Zhang, L. Bonomo, R. Lo Gullo, et al., BI-RADS category assignments by GPT-3.5, GPT-4, and Google Bard: A multilanguage study, Radiology 311 (2024) e232133, https://doi.org/10.1148/radiol.232133.

[17] G. Brembilla, G. Basile, M. Cosenza, F. Giganti, A. Del Prete, T. Russo, et al., Neoadjuvant chemotherapy VI-RADS scores for assessing muscle-invasive bladder cancer response to neoadjuvant immunotherapy with multiparametric MRI, Radiology 313 (2024) e233020, https://doi.org/10.1148/radiol.233020.

[18] R. Jóźwiak, P. Sobecki, T. Lorenc, Intraobserver and interobserver agreement between six radiologists describing mpMRI features of prostate cancer using a PI-RADS 2.1 structured reporting scheme, Life (Basel) 13 (2023) 580, https://doi.org/10.3390/life13020580.

[19] F.N. Loch, K. Beyer, M.E. Kreis, C. Kamphues, W. Rayya, C. Schineis, et al., Diagnostic performance of node reporting and data system (Node-RADS) for regional lymph node staging of gastric cancer by CT, Eur. Radiol. 34 (2024) 3183–3193, https://doi.org/10.1007/s00330-023-10352-5.

[20] Y. Fang, M. Chen, X. Zheng, Y. Yao, K. Huang, S. Chen, et al., Validation of the node reporting and data system (Node-RADS) for standardized CT evaluation of regional lymph nodes in esophageal squamous cell carcinoma patients, Eur. Radiol. 35 (2025) 2999–3009, https://doi.org/10.1007/s00330-024-11234-0.

[21] M. Parillo, C.C. Quattrocchi, Node Reporting and Data System 1.0 (Node-RADS) for the assessment of oncological patients' lymph nodes in clinical imaging, J. Clin. Med. 14 (2025) 263, https://doi.org/10.3390/jcm14010263.

[22] J. Sun, Y. Li, F. Gong, S. Xu, J. Wu, H. Wang, et al., The diagnostic value of PET/CT for the lymph node metastasis in Asian patients with non-small cell lung cancer: a meta-analysis, Hell. J. Nucl. Med. 25 (2022) 196–204, https://doi.org/10.1967/s002449912479.

[23] Q. Wu, J. Lou, J. Liu, L. Dong, Q. Wu, Y. Wu, et al., Performance of node reporting and data system (Node-RADS): A preliminary study in cervical cancer, BMC Med. Imaging 24 (2024) 28, https://doi.org/10.1186/s12880-024-01205-8.

[24] F. Pediconi, R. Maroncelli, M. Pasculli, F. Galati, G. Moffa, A. Marra, et al., Performance of MRI for standardized lymph node assessment in breast cancer: Are we ready for Node-RADS? Eur. Radiol. 34 (2024) 7734–7745, https://doi.org/10.1007/s00330-024-10828-y.

[25] C. Leonardo, R.S. Flammia, S. Lucciola, F. Proietti, M. Pecoraro, B. Bucca, et al., Performance of Node-RADS scoring system for a standardized assessment of regional lymph nodes in bladder cancer patients, Cancers (Basel) 15 (2023) 580, https://doi.org/10.3390/cancers15030580.

[26] C. Jiang, W. Fang, N. Wei, W. Ma, C. Dai, R. Liu, et al., Node reporting and data system combined with computed tomography radiomics can improve the prediction of non-enlarged lymph node metastasis in gastric cancer, J. Comput. Assist. Tomogr. 49 (2025) 215–224, https://doi.org/10.1097/RCT.0000000000001673.

[27] Y. Niu, L. Wen, Y. Yang, Y. Zhang, Y. Fu, Q. Lu, et al., Diagnostic performance of node reporting and data system (Node-RADS) for assessing mesorectal lymph nodes in rectal cancer by CT, BMC Cancer 24 (2024) 716, https://doi.org/10.1186/s12885-024-12487-0.

[28] W.J. Huang, H.B. Xie, P.P. Liu, L. Liu, Z.Y. Liu, Q.J. Wang, et al., Pericardial fat and primary tumor radiomics for predicting occult N2 disease and survival in clinical stage I non-small cell lung cancer: multicenter study with biologic correlation, AJR Am. J. Roentgenol. 225 (2025) e2532861, https://doi.org/10.2214/AJR.25.32861.