MRI STAGING ACCURACY AND DETERMINANTS OF STAGING DISCREPANCY IN NASOPHARYNGEAL CARCINOMA: A MULTICENTRE COMPARATIVE IMAGING STUDY ACROSS CENTRAL AND PROVINCIAL HOSPITALS IN VIETNAM

Pham Thi Bich Dao [12]*, Tran Van Tam [2], Nguyen Dieu My [1], Nguyen Minh Đuc[3], Tong Xuan Thang [1], Pham Tran Anh [1], Nguyen Thi Anh Dao [2], Phung Thi Hoa [4], Nguyen Thi Quynh Trang [4], Phan Xuan Nam [5], Hoang Huu Truong [6], Nguyen Thi Ngoc Anh [7], Pham Thi Phuong Hanh [8]

1.Hanoi Medical University; 2. Hanoi Medical University Hospital;
3. Department of medical service administration - Ministry of health, Viet Nam
4.VNU University of Medicine and Pharmacy
5. Quang Tri General Hospital; 6.Thanh Hoa General Hospital;
7. University of Medicine and Pharmacy, Thai Nguyen University;
8.Ninh Binh Department of Health

https://doi.org/10.60137/tmhvn.v71i79.416

ABSTRACT

Objective: To compare the accuracy of magnetic resonance imaging (MRI) staging for nasopharyngeal carcinoma (NPC) between central and provincial hospitals in Vietnam; to quantify agreement between institutional reports and expert consensus; and to identify institutional, technical, and workflow factors associated with staging discrepancy.

Methods: This multicentre retrospective comparative imaging study included 536 MRI examinations from 536 patients with histologically confirmed NPC treated or evaluated at five Vietnamese hospitals (three central and two provincial). Institutional MRI reports were compared with a blinded consensus re-reading by two senior head-and-neck radiologists. Six staging criteria were assessed: parapharyngeal extension, skull-base or bony invasion, retropharyngeal lymph-node involvement, suspicious subcentimetric cervical lymph nodes, intracranial extension, and cavernous sinus involvement. Agreement was quantified with Cohen's kappa. Case-level logistic regression was used to identify predictors of any staging discrepancy. Criterion-level associations were explored separately to avoid unit-of-analysis error.

Results: Central hospitals showed higher point estimates of sensitivity, specificity, and positive predictive value across all six MRI criteria. The largest inter-tier gaps were observed for suspicious subcentimetric cervical lymph nodes (sensitivity 79.6% vs. 51.3%) and retropharyngeal lymph-node detection (83.4% vs. 57.9%). Agreement between institutional reports and expert consensus was excellent or substantial at central hospitals (overall κ=0.83) but only moderate at provincial hospitals (overall κ=0.52). Overall staging discrepancy affected 23.7% of central cases and 75.2% of provincial cases ($p<0.001$). A treatment-plan change was recommended after expert review in 5.6% of central cases and 29.8% of provincial cases. In the case-level model, provincial hospital tier (adjusted odds ratio [aOR] 3.14), absence of double reading (aOR 2.47), use of 1.5T MRI (aOR 2.03), absence of regular multidisciplinary team meetings (aOR 1.94), and absence of PACS/RIS (aOR 1.72) were the strongest predictors of discrepancy.

Conclusions: MRI staging discrepancy in NPC was substantially more frequent at provincial hospitals, particularly for nodal criteria that require high spatial resolution and subspecialty experience. Double reading, structured reporting, MDT review, PACS/RIS connectivity, and staged MRI infrastructure upgrades are the most actionable priorities. These findings identify the staging criteria most suitable for AI-assisted second-reader tools designed to reduce diagnostic inequality across hospital tiers.

Keywords: nasopharyngeal carcinoma; MRI staging; staging discrepancy; Cohen's kappa; retropharyngeal lymph node; double reading; provincial hospital; Vietnam

---

* Coresponding author: Pham Thi Bich Dao Phone number: 0912 189 853

Email: daoptb0024@hmuh.vn

Date of receipt: 03/07/2026 Date of receipt of feedback: 16/09/2026

Date of receipt of review: 20/09/2026 Date of approval for publication: 25/09/2026

I. INTRODUCTION

Accurate staging is central to treatment selection in nasopharyngeal carcinoma (NPC). Under the AJCC/UICC 8th edition staging system, MRI is essential for defining the extent of the primary tumour and regional nodal disease, including parapharyngeal extension, skull-base involvement, intracranial extension, cavernous sinus invasion, retropharyngeal lymph nodes, and cervical nodal disease [1-4]. Staging errors can alter radiotherapy target volumes, dose prescription, and the indication for systemic therapy [5-7].

Although MRI is the preferred imaging modality for NPC staging, accuracy depends on more than the scanner itself. Field strength, slice thickness, sequence protocol, availability of PACS/RIS, structured reporting, subspecialty radiology experience, double-reading practice, and multidisciplinary team (MDT) review all contribute to the reliability of staging interpretation [3,4,8,9]. These enabling conditions are unevenly distributed across hospital tiers in Vietnam, where central tertiary hospitals generally have greater access to 3T MRI, head-and-neck radiologists, PACS/RIS infrastructure, and regular MDT meetings than provincial hospitals.

Existing Vietnamese data have described broad differences in diagnostic capacity across hospital levels, but the clinical consequence of these differences for MRI staging accuracy in NPC remains insufficiently quantified. In particular, it is unclear which staging criteria generate the largest central-provincial accuracy gaps and which modifiable institutional factors should be prioritised for quality improvement or AI-assisted second-reader development.

This study had three aims: (1) to compare MRI staging performance for six clinically important NPC staging criteria between central and provincial hospitals; (2) to quantify agreement between institutional radiology reports and an expert consensus reference; and (3) to identify case-level institutional, technical, and workflow determinants of staging discrepancy, while examining criterion-level vulnerabilities separately.

II. METHODS

2.1. Study design and setting

This was a multicentre retrospective comparative imaging study with an embedded hospital workflow assessment. The study was conducted at five Vietnamese hospitals: three central or tertiary facilities (Hanoi Medical University Hospital, Vietnam National Cancer Hospital, and Thai Nguyen National Hospital) and two provincial general hospitals (Thanh Hoa General Hospital and Quang Tri General Hospital). Sites were selected to represent contrasting levels of MRI infrastructure, head-and-neck imaging expertise, and quality-assurance practice. Data collection covered MRI examinations and institutional reports generated between December 2025 and April 2026. The manuscript was prepared in accordance with the STROBE reporting framework for observational studies [10].

2.2. Imaging cohort and eligibility criteria

The imaging cohort included 536 MRI examinations from 536 patients with histologically confirmed NPC. Eligible cases had: (1) pretreatment head-and-neck MRI acquired at the participating hospital; (2) histological confirmation of NPC; (3) a locally produced institutional radiology report, either structured or narrative; and (4) clinical staging according to the AJCC/UICC 8th edition TNM classification. Cases were excluded if MRI was technically inadequate because of severe motion artefact, incomplete coverage of the skull base or lower neck, missing key sequences, or unavailable institutional reports.

Hospital-tier, MRI field strength, PACS/RIS availability, double-reading practice, regular MDT review, and use of structured reporting templates were treated as institutional or workflow variables. Physician-level survey characteristics were not analysed as patient-level outcomes in this manuscript, thereby avoiding the mixing of physician and patient denominators.

2.3. Reference standard and MRI staging criteria

The reference standard was a blinded consensus re-reading by two senior head-and-neck radiologists, each with at least 10 years of NPC imaging experience. The two readers independently reviewed the MRI examinations without access to the original institutional report. Disagreements were resolved by consensus discussion.

Six staging criteria were assessed. T-category criteria included parapharyngeal extension, skull-base or other bony invasion, intracranial extension, and cavernous sinus involvement. N-category criteria included retropharyngeal lymph-node involvement and suspicious subcentimetric cervical lymph nodes. Subcentimetric cervical lymph nodes were not considered positive by size alone; suspicious features included abnormal shape, central necrosis, abnormal signal, clustering, extracapsular spread suspicion, or location in a nodal basin typical for NPC spread. This distinction was introduced to align the analysis with practical head-and-neck MRI interpretation and to avoid classifying all lymph nodes ≤10 mm as N-positive.

2.4. Outcomes

The primary outcomes were: (1) sensitivity, specificity, and positive predictive value (PPV) for each of the six MRI staging criteria by hospital tier; (2) agreement between institutional reports and the expert consensus reference, quantified by Cohen's kappa; and (3) any staging discrepancy, defined as disagreement between the institutional report and reference consensus on at least one T- or N-staging criterion.

Secondary outcomes included T-stage discrepancy, N-stage discrepancy, direction of discrepancy (under-staging vs. over-staging), and treatment-plan change recommended after expert review. Treatment-plan change was defined as any consensus recommendation to modify radiotherapy target volume, dose, nodal coverage, or systemic therapy indication based on the reference re-reading.

2.5. Statistical analysis

Continuous variables were summarised as mean and standard deviation or median and interquartile range, as appropriate. Categorical variables were reported as counts and percentages. Inter-tier comparisons used the independent-samples t-test, Mann-Whitney U test, chi-square test, or Fisher's exact test, depending on distribution and cell counts.

Diagnostic performance for each MRI staging criterion was calculated using the expert consensus as the reference standard. Cohen's kappa was interpreted as follows: <0.20 slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial, and >0.80 excellent agreement. Exact 95% confidence intervals should be generated from the final TP/FP/FN/TN matrices before journal submission; the present working manuscript reports the point estimates available in the source summary.

To avoid unit-of-analysis error, predictors of any staging discrepancy were analysed in a case-level logistic regression model. Candidate predictors included hospital tier, MRI field strength, double-reading practice, MDT review, PACS/RIS availability, radiologist experience, structured reporting template use, age, sex, and TNM stage. Criterion-level vulnerability was examined separately as an exploratory criterion-level analysis, ideally using generalized estimating equations or a mixed-effects logistic model with clustering by patient and hospital when raw criterion-level data are available. Two-sided $p<0.05$ was considered statistically significant. Analyses were conducted using SPSS version 26 (IBM Corp., Armonk, NY, USA) and should be replicated in the final statistical package after data-lock.

2.6. Ethics

The study was approved by the Ethics Committee of Hanoi Medical University (approval number 294/GCN-KHCN, 24 April 2026). The approval covered retrospective review and analysis of de-identified imaging, clinical, and workflow data. Physician workflow information, where used, was anonymised and analysed only at institutional level. Informed consent was waived for retrospective records and imaging review under the approved protocol. The study was conducted in accordance with the Declaration of Helsinki.

III. RESULTS

3.1. Imaging cohort and hospital workflow characteristics

The final imaging cohort included 536 MRI examinations from 536 patients with histologically confirmed NPC: 375 from central hospitals and 161 from provincial hospitals. The mean patient age was 46.9 ± 11.8 years and 362 patients (67.5%) were male. Stage III-IV disease was more frequent in provincial cases than in central cases (75.2% vs. 61.6%). Major infrastructure and workflow differences were also observed: 3T MRI was available in 74.1% of central cases but only 23.6% of provincial cases; PACS/RIS availability was 83.2% versus 46.0%; double reading was practised in 58.1% versus 26.1%;

and regular MDT meetings were reported in 60.0% versus 19.9%.

Table 1. MRI case and hospital workflow characteristics

Characteristic	Overall (n=536)	Central (n=375)	Provincial (n=161)
Patient age (years), mean ± SD	46.9 ± 11.8	47.3 ± 11.5	46.1 ± 12.4
Male sex, n (%)	362 (67.5)	256 (68.3)	106 (65.8)
TNM stage I, n (%)	38 (7.1)	31 (8.3)	7 (4.3)
TNM stage II, n (%)	146 (27.2)	113 (30.1)	33 (20.5)
TNM stage III, n (%)	218 (40.7)	148 (39.5)	70 (43.5)
TNM stage IV, n (%)	134 (25.0)	83 (22.1)	51 (31.7)
3T MRI, n (%)	316 (59.0)	278 (74.1)	38 (23.6)
1.5T MRI, n (%)	220 (41.0)	97 (25.9)	123 (76.4)
PACS/RIS available, n (%)	386 (72.0)	312 (83.2)	74 (46.0)
Double reading practised, n (%)	260 (48.5)	218 (58.1)	42 (26.1)
Regular MDT meetings, n (%)	257 (47.9)	225 (60.0)	32 (19.9)

Note: Percentages are calculated within the imaging cohort. This revised table separates patient/MRI-case variables from physician-level survey variables and avoids using physician denominators as patient denominators. MDT, multidisciplinary team; MRI, magnetic resonance imaging; NPC, nasopharyngeal carcinoma; PACS/RIS, picture archiving and communication system/radiology information system; SD, standard deviation; TNM, tumour-node-metastasis.

3.2. MRI staging accuracy by criterion and hospital tier

Central hospitals demonstrated higher point estimates of sensitivity, specificity, and PPV across all six MRI staging criteria. The largest accuracy gaps were observed for suspicious subcentimetric cervical lymph-node assessment and retropharyngeal lymph-node detection. Intracranial extension showed the smallest inter-tier gap, consistent with its more conspicuous imaging appearance compared with subtle nodal disease.

Table 2. MRI staging performance by criterion and hospital tier

MRI staging criterion	Sn Central (%)	Sn Provincial (%)	Sp Central (%)	Sp Provincial (%)	PPV Central (%)	PPV Provincial (%)
Skull-base / bony invasion	87.3	64.1	91.4	78.3	88.6	71.2

MRI staging criterion	Sn Central (%)	Sn Provincial (%)	Sp Central (%)	Sp Provincial (%)	PPV Central (%)	PPV Provincial (%)
Parapharyngeal extension	92.1	71.8	88.6	73.5	91.4	68.9
Retropharyngeal lymph node	83.4	57.9	86.2	69.4	84.8	61.3
Suspicious cervical node ≤10 mm	79.6	51.3	84.7	63.8	78.4	54.7
Intracranial extension	94.8	82.1	97.3	88.9	96.1	84.3
Cavernous sinus involvement	88.2	63.7	93.6	79.4	89.5	65.8

Note: Reference standard: blinded consensus re-reading by two senior head-and-neck radiologists. The source summary provided point estimates only; final submission should add TP/FP/FN/TN and Wilson 95% confidence intervals for each criterion. PPV, positive predictive value; Sn, sensitivity; Sp, specificity.

3.3. Agreement between institutional reports and expert consensus

Agreement between institutional reports and the expert consensus reference was excellent or substantial at central hospitals for all criteria, but only fair to moderate at provincial hospitals for the most difficult nodal criteria. Overall kappa was 0.83 at central hospitals and 0.52 at provincial hospitals.

Table 3. Agreement between institutional reports and expert consensus reference

MRI staging criterion	κ Central	Agreement Central	κ Provincial	Agreement Provincial	p
Skull-base / bony invasion	0.82	Excellent	0.51	Moderate	<0.001
Parapharyngeal extension	0.87	Excellent	0.58	Moderate	<0.001
Retropharyngeal lymph node	0.79	Substantial	0.44	Moderate	<0.001
Suspicious cervical node ≤10 mm	0.74	Substantial	0.38	Fair	<0.001
Intracranial	0.91	Excellent	0.72	Substantial	0.003

MRI staging criterion	κ Central	Agreement Central	κ Provincial	Agreement Provincial	p
extension					
Cavernous sinus involvement	0.84	Excellent	0.49	Moderate	<0.001
Overall (any criterion)	0.83	Excellent	0.52	Moderate	<0.001

Note: Agreement categories: <0.20 slight; 0.21-0.40 fair; 0.41-0.60 moderate; 0.61-0.80 substantial; >0.80 excellent. Final submission should include 95% confidence intervals for each kappa coefficient.

3.4. Staging discrepancy and treatment-planning implications

Overall staging discrepancy was present in 89/375 central cases (23.7%) and 121/161 provincial cases (75.2%; p<0.001). N-stage discrepancy was more frequent than T-stage discrepancy, and under-staging was more frequent than over-staging. After expert consensus review, treatment-plan modification was recommended in 21/375 central cases (5.6%) and 48/161 provincial cases (29.8%). Among discrepant cases, the corresponding proportions were 23.6% and 39.7%, respectively.

Table 4. Staging discrepancy rates by type, direction, and clinical consequence

Discrepancy type	Central, n (%)	Provincial, n (%)	Overall, n (%)	p
T-stage discrepancy (any)	48/375 (12.8)	71/161 (44.1)	119/536 (22.2)	<0.001
T under-staging	31/375 (8.3)	54/161 (33.5)	85/536 (15.9)	<0.001
T over-staging	17/375 (4.5)	17/161 (10.6)	34/536 (6.3)	0.004
N-stage discrepancy (any)	62/375 (16.5)	89/161 (55.3)	151/536 (28.2)	<0.001
N under-staging	44/375 (11.7)	72/161 (44.7)	116/536 (21.6)	<0.001
N over-staging	18/375 (4.8)	17/161 (10.6)	35/536 (6.5)	0.006
Overall discrepancy (T or N)	89/375 (23.7)	121/161 (75.2)	210/536 (39.2)	<0.001
Treatment-plan change, all cases	21/375 (5.6)	48/161 (29.8)	69/536 (12.9)	<0.001
Treatment-plan change among discrepant cases	21/89 (23.6)	48/121 (39.7)	69/210 (32.9)	0.015

Note: Discrepancy was defined as disagreement between institutional report and expert consensus on any of the six staging

criteria. Treatment-plan change was defined as a recommended modification of radiotherapy target, dose, nodal coverage, or systemic therapy indication after expert review.

3.5. Predictors of staging discrepancy

In the case-level multivariable model, provincial hospital tier was the strongest independent predictor of any staging discrepancy, followed by absence of double reading, use of 1.5T rather than 3T MRI, absence of regular MDT meetings, and absence of PACS/RIS. These findings identify a hierarchy of modifiable system-level targets. Criterion-level vulnerability was evaluated separately because criterion-level predictors are not interchangeable with case-level predictors.

Table 5. Independent predictors of MRI staging discrepancy

Variable	aOR	95% CI	p	Domain
Case-level model: provincial hospital tier	3.14	2.18 - 4.52	<0.001	Institutional
Case-level model: absence of double reading	2.47	1.71 - 3.57	<0.001	Workflow
Case-level model: 1.5T MRI vs. 3T	2.03	1.42 - 2.90	<0.001	Technical
Case-level model: absence of regular MDT meetings	1.94	1.35 - 2.79	<0.001	Institutional
Case-level model: absence of PACS/RIS	1.72	1.18 - 2.51	0.005	Technical
Case-level model: radiologist experience <5 years	1.58	1.08 - 2.31	0.019	Workforce
Case-level model: no structured reporting template	1.51	1.04 - 2.19	0.030	Workflow
Exploratory criterion-level: N-staging criterion vs. T-staging	1.87	1.41 - 2.49	<0.001	Criterion

Variable	aOR	95% CI	p	Domain
Exploratory criterion-level: small-node criterion	2.31	1.62-3.29	<0.001	Criterion

Note: The revised table separates case-level and criterion-level estimates to avoid unit-of-analysis error. The final criterion-level model should be run on the 536 × 6 criterion dataset using generalized estimating equations or a mixed-effects logistic model with clustering by patient and hospital. aOR, adjusted odds ratio; CI, confidence interval; MDT, multidisciplinary team; MRI, magnetic resonance imaging; PACS/RIS, picture archiving and communication system/radiology information system.

IV. DISCUSSION

This multicentre comparative imaging study shows that MRI staging discrepancy in NPC is substantially more frequent at provincial hospitals than at central hospitals in Vietnam. The central-provincial gap was most pronounced for nodal staging criteria, particularly retropharyngeal lymph-node assessment and suspicious subcentimetric cervical lymph nodes. These criteria require high spatial resolution, consistent imaging protocols, and head-and-neck subspecialty experience, explaining why they are especially vulnerable in settings where 1.5T MRI predominates and structured quality-assurance workflows are limited.

The first major finding is the high discrepancy burden in provincial cases. Overall discrepancy affected 75.2% of provincial cases compared with 23.7% of central cases, and N-stage under-staging was the dominant error pattern. This is clinically important because nodal staging directly influences nodal target-volume delineation, dose coverage, and systemic therapy decisions. Because this study did not measure survival or recurrence outcomes, the finding should be interpreted as evidence of potential treatment-planning risk rather than proof of downstream oncologic harm.

Second, the criterion-level findings identify the specific anatomical targets where quality improvement and AI-assisted decision support are most likely to help. Small cervical lymph-node assessment and retropharyngeal node detection showed both the largest accuracy gaps and the lowest agreement values in provincial hospitals. Future AI tools should therefore prioritise criterion-level detection, measurement, and structured reporting of these nodal findings rather than producing only a single overall stage label.

Third, the multivariable model highlights modifiable institutional and technical determinants. Double reading and MDT review are immediately actionable quality-assurance interventions that do not require major equipment investment. PACS/RIS availability and 3T MRI upgrades require capital planning, but they address the technical limitations underlying the observed gaps in skull-base, retropharyngeal, and small-node assessment. Structured reporting templates may provide an additional safety layer by forcing systematic documentation of each staging criterion.

The findings have direct implications for the development of an AI-assisted NPC diagnosis and staging system in Vietnam. The most

appropriate role for AI in this context is not to replace radiologists, but to function as a second reader and workflow support tool: highlighting suspicious small nodes, prompting systematic evaluation of retropharyngeal nodes and skull-base structures, generating criterion-level structured report fields, and flagging cases for double reading or MDT review. Such a design would align AI development with the most clinically consequential errors identified in this study.

Several limitations should be acknowledged. First, the reference standard was expert imaging consensus rather than pathological verification of each staging criterion, which is rarely feasible in routine NPC staging. Second, the retrospective design cannot determine whether staging discrepancies would have been corrected during later clinical review. Third, because several predictors were measured at the hospital or workflow level, final inferential models should adjust for clustering by hospital. Fourth, the source summary did not provide TP/FP/FN/TN counts for each criterion; exact confidence intervals should therefore be generated from the raw diagnostic matrix before submission. Finally, the five participating hospitals may not represent the full range of MRI capacity across Vietnam.

V. CONCLUSIONS

MRI staging discrepancy for NPC was substantially more frequent at provincial hospitals than at central hospitals in Vietnam, with nodal under-staging as the dominant error pattern. The largest performance gaps involved retropharyngeal and suspicious subcentimetric cervical lymph-node assessment. Provincial hospital tier, absence of double reading, 1.5T MRI field strength, absence of regular MDT meetings, and PACS/RIS absence were the strongest case-level predictors of discrepancy. Double-reading protocols, structured reporting, MDT review, PACS/RIS implementation, and staged MRI infrastructure upgrades should be prioritised. AI-assisted tools should focus on criterion-level second reading for nodal and skull-base staging, integrated into structured reporting workflows rather than deployed as black-box stage classifiers.

AUTHORS' CONTRIBUTIONS

Dao Pham Thi Bich: conceptualisation, study design, supervision, manuscript review, and final approval. Tam Tran Van and Dao Nguyen Thi Anh: clinical data collection and records review. My Nguyen Dieu, Thang Tong Xuan, and Anh Pham Tran: data cleaning, coding, and statistical analysis. Hoa Phung Thi, Trang Nguyen Thi Quynh: imaging accuracy protocol and statistical methods. Nam Phan Xuan and Truong Hoang Huu: site coordination and provincial data collection. My guyen Dieu, Anh Nguyen Thi Ngoc and Hanh Pham Thi Phuong: literature review, reference-reading protocol, and manuscript dNrafting. All authors read, revised, and approved the final version.

DECLARATION OF COMPETING INTERESTS

The authors declare no competing interests related to this study.

FUNDING

This study was conducted as part of preparatory research activities for the project "Research and development of an AI-assisted NPC diagnosis system using multimodal medical imaging and clinical data in Vietnam." No external funding was received for article processing charges or manuscript publication.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Approved by the Ethics Committee of Hanoi Medical University (294/GCN-KHCN, 24 April 2026). The approval covered retrospective review of de-identified imaging and clinical records. Informed consent was waived under the approved protocol.

ACKNOWLEDGEMENTS

The authors thank the radiology, ENT, and oncology teams at all five participating hospitals for supporting data collection; the senior reference radiologists for independent re-reading; and the research coordination team at Hanoi Medical University for data management and quality control.

REFERENCES

1. Amin MB, Edge SB, Greene FL, et al. AJCC Cancer Staging Manual. 8th ed. Cham: Springer; 2017.
2. Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal carcinoma. Lancet. 2019;394:64-80.
3. King AD, Bhatia KSS. Magnetic resonance imaging staging of nasopharyngeal carcinoma in the head and neck. World J Radiol. 2010;2:159-165.
4. Abdel Razek AAK, King A. MRI and CT of nasopharyngeal carcinoma. AJR Am J Roentgenol. 2012;198:11-18.
5. Liu LZ, Zhang GY, Xie CM, et al. Magnetic resonance imaging of retropharyngeal lymph node metastasis in nasopharyngeal carcinoma. Radiography. 2006;12:310-318.
6. Li YZ, Xie CM, Wu YP, et al. Nasopharyngeal carcinoma patients with retropharyngeal lymph node metastases: a minimum axial diameter of 6 mm is a more accurate prognostic predictor than 5 mm. AJR Am J Roentgenol. 2015;204:20-26.
7. Lee AWM, Ma BBY, Ng WT, Chan ATC. Management of nasopharyngeal carcinoma: current practice and future perspective. J Clin Oncol. 2015;33:3356-3364.
8. Ng WT, Lee MC, Hung WM, et al. Clinical outcomes and patterns of failure after intensity-modulated radiotherapy for nasopharyngeal carcinoma. Int J Radiat Oncol Biol Phys. 2011;79:420-428.
9. Mao YP, Tang LL, Chen L, et al. Prognostic factors and failure patterns in non-metastatic nasopharyngeal carcinoma after intensity-modulated radiotherapy. Chin J Cancer. 2016;35:103.
10. Pillay B, Wootten AC, Crowe H, et al. The impact of multidisciplinary team meetings on patient assessment, management and outcomes in oncology settings: a systematic review of the literature. Cancer Treat Rev. 2016;42:56-72.
11. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement. Lancet. 2007;370:1453-1457.
12. Jin GQ, Su DK, Xie D, et al. The diagnostic value of 1.5-T diffusion-weighted MR imaging in detecting 5 to 10 mm metastatic lymph nodes of patients with nasopharyngeal carcinoma. Medicine (Baltimore). 2016;95:e4286.
13. Ke L, Deng Y, Xia W, et al. Development of a self-constrained 3D DenseNet model in automatic detection and segmentation of nasopharyngeal carcinoma using magnetic resonance images. Oral Oncol. 2020;110:104862.

14. Huang X, Chen H, Yan JD, et al. Study on structured method of Chinese MRI report of nasopharyngeal carcinoma. BMC Med Inform Decis Mak. 2021;21(Suppl 2):214.

15. Liu S, Liu Y, Gong Q, et al. A coordinate registration-based structured magnetic resonance imaging reporting method for nasopharyngeal carcinoma: a preliminary study. Quant Imaging Med Surg. 2025;15:2774-2787.

AUTHOR DATA-CHECK NOTES BEFORE SUBMISSION

- Confirm that the imaging cohort is exactly 536 MRI examinations from 536 patients, and that central/provincial denominators are 375 and 161.
- Add TP/FP/FN/TN and Wilson 95% confidence intervals to Table 2 when raw criterion-level matrices are available.
- Run the final case-level model with hospital-level clustering or mixed effects if possible, because several predictors are institutional variables.
- Run the final criterion-level model using GEE or mixed-effects logistic regression on the 536 × 6 criterion dataset.
- Verify that ethics approval language accurately reflects whether all data were retrospective and de-identified before analysis.
- Remove this data-check section before journal submission.