RESEARCH



Establishment of a deep-learning-assisted recurrent nasopharyngeal carcinoma detecting simultaneous tactic (DARNDEST) with high cost-effectiveness based on magnetic resonance images: a multicenter study in an endemic area

Yishu Deng^{1,2,4†}, Yingying Huang^{2,3†}, Haijun Wu^{5†}, Dongxia He^{2,3†}, Wenze Qiu⁶, Bingzhong Jing^{2,4}, Xing Lv^{2,7}, Weixiong Xia^{2,7}, Bin Li^{2,4}, Ying Sun^{2,8}, Chaofeng Li^{2,4,9*}, Chuanmiao Xie^{2,3*} and Liangru Ke^{2,3*}

Abstract

Background To investigate the feasibility of detecting local recurrent nasopharyngeal carcinoma (rNPC) using unenhanced magnetic resonance images (MRI) and optimize a layered management strategy for follow-up with a deep learning model.

Methods Deep learning models based on 3D DenseNet or ResNet frames using unique sequence (T1WI, T2WI, or T1WIC) or a combination of T1WI and T2WI sequences (T1_T2) were developed to detect local rNPC. A **d**eep-learning-**a**ssisted **r**ecurrent **N**PC **de**tecting **s**imultaneous **t**actic (DARNDEST) utilized DenseNet was optimized by superimposing the T1WIC model over the T1_T2 model in a specific population. Diagnostic efficacy (accuracy, sensitivity, specificity) and examination cost of a single MR scan were compared among the conventional method, T1_T2 model, and DARNDEST using McNemar's Z test.

Results No significant differences in overall accuracy, sensitivity, and specificity were found between the T1WIC model and T1WI, T2WI, or T1_T2 models in both test sets (all *P* > 0.0167). The DARNDEST had higher accuracy and sensitivity but lower specificity than the T1_T2 model in both the internal (accuracy, 85.91% vs. 84.99%; sensitivity, 90.36% vs. 84.26%; specificity, 82.20% vs. 85.59%) and external (accuracy, 86.14% vs. 84.16%; sensitivity, 90.32% vs. 84.95%; specificity, 82.57% vs. 83.49%) test sets. The cost of a single MR examination using DARNDEST was \$330,724

 $^{\rm t}\rm Yishu$ Deng, Yingying Huang, Haijun Wu and Dongxia He contributed equally to this work.

*Correspondence: Chaofeng Li Licf@sysucc.org.cn Chuanmiao Xie Xiechm@sysucc.org.cn Liangru Ke kelr@sysucc.org.cn Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.gr/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.gr/licenses/by/4.0/.

(internal) and \$328,971 (external) with a hypothetical cohort of 1,000 patients, relative to \$313,250 of the T1_T2 model and \$340,865 of the conventional method.

Conclusions Detecting local rNPC using unenhanced MRI with deep learning is feasible and DARNDEST-driven follow-up management is efficient and economic.

Background

Nasopharyngeal carcinoma (NPC), a malignancy originating from the mucosal epithelium of the nasopharynx, is particularly prevalent in Southeast Asia and North Africa [1]. Despite advancements in radiotherapy and imaging techniques, treatment failure due to tumor recurrence remains a significant concern for NPC patients, with local recurrence occurring in 10% to 20% of cases in endemic regions [2–4]. Early detection and precise localization of recurrent NPC (rNPC) are essential for effective salvage treatment and improved prognosis [4–7].

However, detecting local rNPC poses challenges as it tends to grow submucosally, and deep-seated local rNPCs may be missed during clinical examination [8] or result in inaccessible endoscopic biopsy, thus thwart histopathological diagnosis. Hence, noninvasive and efficient modalities are urgently required for post-treatment assessment in NPC patients. Magnetic resonance imaging (MRI) has proven valuable in NPC management due to its excellent resolution in soft tissue, sensitivity in detecting bone marrow infiltration and advantage of multiparameter imaging [9], but its diagnostic efficacy in detecting local rNPC varies across studies, with sensitivity ranging from 56% to 83% and specificity from 78% to 83% [10, 11]. Radiation-induced anatomic distortion, alteration in signal intensity pattern, inflammatory reactions and scarring pose diagnostic difficulty in differentiating recurrent tumors from the variable appearance of radiation-induced alteration [12, 13].

Given the limitations of single imaging modalities, regular follow-up with repetitive MRI examinations is preferred for detecting local rNPC. However, repeated administration of MR contrast agent might increase the economic burden, mental stress and potential side effects for patients. Notably, MRI signals lack specific for distinguishing between recurrence and post-treatment changes, moreover, the 5-year local relapse-free survival has been increased from 66.6-90.8% to 87.7-95.5% in NPC patients with different stages of disease by introducing intensity modulated radiotherapy (IMRT) and comprehensive treatment strategy [14–16]. Deep learning algorithms have exhibited remarkable performance in lesion detection and segmentation by capturing

intrinsic features of medical images. Deng YS et al demonstrated that contrast-enhanced magnetic resonance imaging (ceMRI) could be substituted in the identification and segmentation of NPC with the aid of deep learning models [17]. Likewise, the convolutional neural network-based algorithm could automatically discriminate between malignant and benign diseases at nasopharynx using T2-weighted fat-suppressed MR images [18]. Recently, MR-based deep learning models have enhanced radiologists' accuracy in detecting regional recurrent NPC [19]. This prompts the question of whether deep learning algorithms could aid in the differentiation between rNPC and radiation alterations, optimizing the follow-up strategy by selectively discarding contrast-enhanced MR examinations in certain subpopulation. This approach alleviates the economic burden on patients and national health systems and minimizes the potential risks associated with gadolinium-based contrast agents (GBCAs) in some extent.

To address that, deep learning models employed selfconstrained 3D DenseNet or ResNet using unique sequence images (T1-weighted image [T1WI], T2-weighted image [T2WI], or post-contrast T1WI [T1WIC]) or their combination were developed to detect local rNPC. Consequently, we have demonstrated that unenhanced MRI is adequate in selective post-treatment NPC patients during follow-up with the aid of deep learning-based models and verified that in three different institutes. Moreover, an optimal deeplearning-assisted recurrent NPC detecting simultaneous tactic (DARNDEST) based on MRI and a layer management strategy was developed, thereby providing an efficient, cost-effective, and lower-risk approach for NPC follow-up.

Methods

Patient enrollment

The participants in this study were previously subjects of research investigating the feasibility of detecting regional rNPC using deep learning-aided MRI [19]. In contrast, this research focused on local rNPC, with an emphasis on the viability of omitting enhanced MR examination and incorporating layer management for follow-up by introducing economic analysis.

We searched the picture archiving and communication system (GE, USA) using the following terms and strategy: examination time = from Dec 01st, 2008 to July 31st, 2020, equipment = MR, examination type = plain and enhancement, region = head and neck, diagnosis = recurrent nasopharyngeal carcinoma or post-treatment or postradiotherapy alteration. The following cases were excluded: a) poor quality images due to artifacts or incomplete images; b) unconfirmed diagnosis; c) heterochronous tumor located at the nasopharynx or adjacent region; and d) surgery delivered to the nasopharynx. Finally, 4349, 420 and 257 eligible patients were enrolled from Sun Yat-sen University Cancer Center (SYSUCC), the First People's Hospital of Foshan (FPHF) and the Affiliated Cancer Hospital of Guangzhou Medical University (ACH-GMU), respectively. To enhance model robustness, external participants were combined into an "external cohort". The participants were randomly split into training, validation and test sets according to ratios of 8:1:1 and 1.4:1:1 in the internal and external cohorts, respectively (Fig. 1).

Scanning parameters

Digital Imaging and Communications in Medicine images (DICOM) of T1-weighted (T1WI), and T2-weighted (T2WI) images preenhancement and T1-weighted images postenhancement (T1WIC) in the axial view were used for development, tuning and evaluation of the models. The images in the internal cohort were non-fat saturated, while those in the external cohort were fat saturated. Detailed information on the scanning machines and parameters was available in a previous study [19].

Radiologists' assessment and Image tagging

Initially, the image tagging was performed based on the imaging report. A junior radiologist (D.H.) with two years of experience, along with two radiation oncologists



Fig. 1 Flowchart of the participant enrollment and development of deep learning models

(H.W. and W.Q.) with over ten years of experience in head and neck malignancies, subsequently reviewed the image tagging in sequence and made necessary corrections according to the following criteria, 1) local recurrence referred to the detection of viable lesions in the nasopharynx or adjacent skull base/intracranial, including histopathology-confirmed nasopharyngeal carcinoma or those whose nasopharyngeal lesions increased in size as detected by time-serial images three months before or after the current MR examination, or who had suspicious nasopharyngeal lesions detected by the current MR examination that subsequently shrunk after antitumor treatment; 2) local recurrence-free referred to the absence of detectable viable lesions in the nasopharynx or adjacent skull base/intracranial on the current examination, with no evidence of recurrence, as defined above, after a follow-up period of one year. Notably, participants with regional recurrence, such as metastatic retropharyngeal or cervical lymph nodes, but no viable lesions in the nasopharynx or adjacent skull base/intracranial, were classified as negative for local recurrence. For cases lacking histopathological evidence and where there were different opinions between the junior radiologist and radiation oncologists, a senior radiologist (C.X.) with over thirty years of experience in head and neck malignancy made the final determination.

A senior radiologist (L.K.) with over ten years of experience in field of head and neck malignancy and a junior radiologist (Y.H.) with five years of experience participated in the test section. The two radiologists made diagnoses based on a single MR scan with complete DICOM images in the test set and were blinded to the histopathology and previous or subsequent examinations.

Model development and tuning

Two commonly used deep learning model architectures, 3D DenseNet and 3D ResNet34 [20], were employed to investigate whether enhanced imaging could be replaced for the detection of local rNPC. PyTorch with 2 GPUs was utilized to train the 3D DenseNet and ResNet models for local rNPC detection using different sequences. To ensure the comparability of the performance of different models, we adopted the same hyperparameter settings and maintained the consistency of the training process. This includes using the same optimizer, loss function, minibatch size, learning rate adjustment strategy, data preprocessing, data augmentation methods, and model training monitoring. MR image preprocessing is described in supplementary methods. The detailed description of the 3D DenseNet model structure and a summary of its parameters are provided in the supplementary methods and supplementary Table 1, respectively. 'Adam' optimizer with binary cross-entropy loss, minibatch size 12, and initial learning rate $1e^{-3}$ was used for training. Data augmentation with RandomAffine [21] and an adaptive learning rate scheduling strategy was applied. Learning rate was reduced by half if validation set loss plateaued for 5 epochs and monitoring threshold set to $1e^{-4}$. Finally, models with best performance in validation set were chosen. The predicted values of T1WI and T2WI were combined with a weight of 0.5 to create T1_T2 model. The cutoff values for the predicted values of each model were determined using the Youden index (Supplementary Table 2).

Optimization of the DARNDEST

Given the superior overall performance of the DenseNet model in detecting local rNPC compared to the ResNet model, the optimization of DARNDEST focused on the DenseNet model. To further improve the sensitivity of detecting local rNPC and identify those who might benefit from contrast-enhanced MRI (ceMRI), patients were divided into three groups as follows: all patients underwent evaluation using the T1_T2 model initially, those with predicted values above the first-level cutoff value (internal: 0.490, external: 0.460) according to the Youden index were categorized as positive. A second-level cutoff value was established by ensuring that the sensitivity was no less than 90% across the entire test set, achieved by adding the T1WIC model to the non-positive group. Non-positive patients were then allocated to suspicious and negative groups based on the secondary cutoff values (internal: 0.197, external: 0.282) (Supplementary Figure 1). DARNDEST was optimized for layered patient management; herein, patients in the positive and suspicious groups were subjected to enhanced MR examination, whereas patients in the negative group merely underwent unenhanced MR examination, thus maximizing the cost-benefit ratio.

Comparison of identification performance in local rNPC among models and statistical analysis

The developed models' performance in detecting local rNPC was assessed using standard evaluation metrics such as accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). And these metrics were compared between any two models using McNemar's Z test. Receiver operating characteristic (ROC) curves were used to calculate the area under the curve (AUC) to assess the diagnostic efficacy

| Test | No of rNPC | TP | FN | FP | ΤN | sensitivity | specificity | accuracy | PPV | NPV |
|----------|------------|-----|----|----|-----|--------------------|--------------------|--------------------|--------------------|--------------------|
| | | | | | | % 95% CI (%) |
| Internal | 197 | | | | | | | | | |
| T1WIC | | 166 | 31 | 35 | 201 | 84.26(79.18,89.35) | 85.17(80.64,89.70) | 84.76(81.37,88.14) | 82.59(77.34,87.83) | 86.64(82.26,91.02) |
| T1WI | | 168 | 29 | 40 | 196 | 85.28(80.33,90.23) | 82.57(78.26,87.84) | 84.06(80.62,87.51) | 80.77(75.41,86.13) | 87.11(82.73,91.49) |
| T2WI | | 164 | 33 | 40 | 196 | 83.25(78.03,88.46) | 85.32(78.26,87.84) | 83.14(79.61,86.67) | 80.39(74.94,85.84) | 85.59(81.04,90.14) |
| T1_T2 | | 166 | 31 | 34 | 202 | 84.26(79.18,89.35) | 85.59(81.11,90.07) | 84.99(81.62,88.35) | 83.00(77.79,88.21) | 86.70(82.33,91.06) |
| Doctor-S | | 154 | 43 | 13 | 223 | 78.17(71.62,83.60) | 94.49(90.55,96.91) | 87.07(83.45,90.01) | 92.22(86.78,95.62) | 83.83(78.73,87.94) |
| Doctor-J | | 166 | 31 | 41 | 195 | 84.26(78.25,88.90) | 82.63(77.05,87.11) | 83.37(79.45,86.69) | 80.19(73.97,85.26) | 86.28(80.94,90.35) |
| External | 93 | | | | | | | | | |
| T1WIC | | 78 | 15 | 19 | 90 | 83.87(76.40,91.35) | 82.57(75.45,89.69) | 83.17(78.01,88.33) | 80.41(72.51,88.31) | 85.71(79.02,92.41) |
| T1WI | | 75 | 18 | 19 | 90 | 80.65(72.62,88.67) | 82.57(75.45,89.69) | 81.68(76.35,87.02) | 79.79(71.67,87.91) | 83.33(76.30,90.36) |
| T2WI | | 76 | 17 | 16 | 93 | 81.72(73.87,89.58) | 85.32(78.68,91.96) | 83.66(78.57,88.76) | 82.61(74.86,90.35) | 84.55(77.79,91.30) |
| T1_T2 | | 79 | 14 | 18 | 91 | 84.95(77.68,92.21) | 83.49(76.52,90.46) | 84.16(79.12,89.19) | 81.44(73.71,89.18) | 86.67(80.16,93.17) |
| Doctor-S | | 77 | 16 | 3 | 106 | 82.80(73.26,89.55) | 97.25(91.57,99.29) | 90.59(85.49,94.09) | 96.25(88.68,99.03) | 86.89(79.28,92.09) |
| Doctor-J | | 72 | 21 | 4 | 105 | 77.42(67.35,85.18) | 96.33(90.32,98.82) | 87.62(82.09,91.68) | 94.74(86.36,98.30) | 83.33(75.42,89.16) |

Table 1 Comparison of performance in identifying local rNPC among DenseNet models developed using different MRI sequence and doctors in test set

Abbreviation: rNPC recurrent nasopharyngeal carcinoma, PPV positive predictive value, NPV negative predictive value, T1WIT1-weighted image, T2WIT2-weighted image, T1WICT1-weighted image post-enhancement, Doctor-S senior radiologist, Doctor-J junior radiologist.

of each model. Differences in AUCs among models were assessed using paired-sample area differences under the ROC curves. All analyses were performed using the Statistical Program for Social Sciences 22.0 (Chicago, USA). The power of the test during comparison among any two models was assessed using Power Analysis and Sample Size 2023 (Kaysville, USA).

Cost analysis

An economic analysis was conducted based on the actual prevalence of the internal and external cohorts, positive yield of local rNPC, and associated confidence interval with a hypothetical cohort of 1,000 NPC patients. The cost of detecting one true-positive local rNPC either using enhanced or unenhanced MRI was calculated for different groups according to DARNDEST. The Chinese currency was converted to US dollars based on the exchange rate and date (rate of US\$1.00=\$7.17, 2024). The prevalence was defined as those patients with local rNPC divided by the total cases. Additional fees for other examinations or treatment were not included in the economic analysis.

Results

Demographic and clinical characteristics of the eligible participants

Among the 1993 patients with recurrence in the internal cohort, 1467 (73.61%) and 35 (1.76%) patients had histopathologically confirmed recurrent lesion at nasopharynx and metastatic lymphadenopathy, respectively. 73.40% (138/188) and 69.11% (85/123) of patients had histopathologically confirmed local rNPC in FPHF and ACHGMU. Notably, Submucosal local rNPC accounted for 85.6%, 4.26%, and 9.76%, while lesions with necrotic characteristic accounted for 23.53%, 13.30%, and 21.95% in SYSUCC, FPHF, and ACHGMU cohorts, respectively. Most patients in the three institutions were diagnosed with advanced local rNPC, with 74.86%, 84.04%, and 82.11% in SYSUCC, FPHF, and ACHGMU, respectively. Detailed demographic and clinical information from the three institutions are listed in Supplementary Tables 3-5.

Comparison of diagnostic performance of different models in identifying local rNPC

The T1WIC model using DenseNet showed no significant differences in overall accuracy, sensitivity, or specificity when compared to the T1WI, T2WI, or T1 T2 models, in both the internal and external cohorts (all P > 0.0167), with an overall accuracy and sensitivity over 83% in the internal cohort and over 81% in the external cohort, and all of the power of the test more than 95%. Radiologists achieved a slightly higher or non-inferior overall accuracy, benefiting from the favorable specificity that was slightly above or equal to that of the DenseNet models, but they experienced a decline in sensitivity, which was critical in detecting recurrence during post-treatment surveillance. Moreover, the PPV and NPV of the T1_T2 model were slightly higher than those of T1WIC without statistical significance in both test sets (Table 1). The T1_T2 model showed similar performance to the T1WIC



Fig. 2 ROC curves of DenseNet models in the test set. (**a**-**b**) The diagnostic efficacy of DenseNet models developed using different MRI sequences in detecting local recurrent nasopharyngeal carcinoma is compared using ROC curves in the internal (**a**) and external (**b**) test sets. The area under the curve and 95% confidence interval of each model are also shown. Abbreviations: ROC, receiver operator characteristic curve; Doctor-S, senior radiologist; Doctor-J, junior radiologist.

model when using the ResNet frame, with slightly higher sensitivity but no statistical significance in both the internal (84.26% vs. 82.74%) and external (84.95% vs. 84.95%) test sets (Supplementary Table 6).

Particularly, the ROC curves revealed no significant differences in the AUC between the T1WIC model and the T1WI, T2WI, or T1_T2 models in both test sets when using DenseNet (0.9124 vs. 0.9098, 0.9145, 0.9296 in the internal test set; 0.9071 vs. 0.8874, 0.8877, 0.9065 in the external test set; all P > 0.0167, Fig. 2) or ResNet (0.8970 vs. 0.8840, 0.8939, 0.9045 in the internal test set; 0.9061 vs. 0.8800, 0.8791, 0.8973 in the external test set; all P > 0.0167, Supplementary Figure 2).

Comparison of errors in diagnosing local rNPC among models and radiologist

Further analysis of errors in diagnosing local rNPC was performed to understand the mechanism of DenseNet models in detecting lesions. In misdiagnosed cases, mistaking radiation-induced fibrosis as recurrence was the first major cause for both the DenseNet models and the junior radiologist in both test sets, whereas mistaking radiation necrosis or wrongly identifying sinusitis or enhanced turbinate as recurrence were the second major causes for the DenseNet models in the internal or external test set, respectively (Fig. 3a, b). However, the senior radiologist, whose specificity was higher than that of deep learning models and junior radiologist (94.49% vs. 82.57-85.17% in internal test set, 97.25% vs. 82.57-96.22% in external test set, Table 1), excelled in distinguishing radiation-induced fibrosis from recurrence in both test sets but encountered challenges in differentiating recurrence from radiation necrosis, leading to lower sensitivity compared to models and junior radiologist in the internal test set (78.17% vs. 83.25-85.28%, Table 1 and Fig. 3b, c).

(See figure on next page.)

Fig. 3 Classification and comparison of errors in diagnoses among deep learning models and doctors. **a** Representative images of major errors in diagnosing local rNPC. The representative cases are radiation necrosis (necrotic lesion indicated by asterisk) wrongly identified by models in a case without recurrence, and missed diagnose due to low confidence (early recurrent lesion indicated by arrowhead at the right parietal wall of the nasopharynx). The columns from left to right in each case are original MR images, heatmap predicted by the DenseNet model and merged images of the former two. **b**-**c** The errors of misdiagnose (**b**) and missed diagnose (**c**) in local rNPC are shown in the stacked bar chart. Herein, focal recurrence refers to cases with rT1 stage disease; atypical recurrent site refers to uncommon site involved by NPC, such as nasolacrimal duct, nasal cavity, etc.; difficult to differentiate refers to misdiagnose between local rNPC and radiation necrosis.



Fig. 3 (See legend on previous page.)

| Cohort | Model/Method | Accuracy (%) | Sensitivity (%) | Specificity (%) | тс | TT(hr) |
|-------------------|--------------|------------------------|------------------------|------------------------|-----------------------|--------|
| | | 95% CI (%) | 95% CI (%) | 95% CI (%) | | |
| Internal | T1_T2 | 84.99 (81.62,88.35) | 84.26 (79.18,89.35) | 85.59 (81.11,90.07) | ¥2,246,000(\$313,250) | 129.2 |
| | DARNDEST | 85.91 (82.19,88.98) | 90.36 (85.13,93.94) | 82.20 (76.59,86.74) | ¥2,371,293(\$330,724) | 190.7 |
| External | T1_T2 | 84.16 (79.12,89.19) | 84.95 (77.68,92.21) | 83.49 (76.52,90.46) | ¥2,246,000(\$313,250) | 129.2 |
| | DARNDEST | 86.14 (80.42,90.44) | 90.32 (81.97,95.20) | 82.57 (73.86,88.92) | ¥2,358,723(\$328,971) | 184.5 |
| Internal/External | Conventional | / | / | / | ¥2,444,000(\$340,865) | 226.4 |

Table 2 Comparison of the diagnostic efficacy, total cost and time in a single MR examination using DARNDEST and T1_T2 model or conventional methods

Abbreviation: TC total cost, TT Total time, DARNDEST deep-learning-assisted recurrent NPC detecting simultaneous tactic.

Missed diagnoses for radiologists in both test sets were primarily due to recurrence at an early stage with focal lesions. Conversely, deep learning models struggled with accurate identification of lesions in the correct location and displayed uncertainty in true recurrent cases, with positional misjudgments as a secondary cause for missed diagnoses. Additionally, recurrence at an atypical site or skull base constituted a critical cause of missed diagnoses for radiologists in the internal or external test set, respectively (Fig. 3a, c).

Evaluation of DARNDEST's diagnostic performance and a comparative analysis of economic and time factors against the T1_T2 model and conventional methods for detecting rNPC

Expectedly, DARNDEST showed higher accuracy and sensitivity compared to the T1_T2 model in the whole test sets, with slightly reduced specificity in both the internal (accuracy, 85.91% vs. 84.99%; sensitivity, 90.36% vs. 84.26%; specificity, 82.20% vs. 85.59%) and external (accuracy, 86.14% vs. 84.16%; sensitivity, 90.32% vs. 84.95%; specificity, 82.57% vs. 83.49%) cohorts (Table 2).

Further economic analysis was compared among the conventional method, T1_T2 model alone and DARND-EST. There would be 462, 171 and 367 patients categorized as positive, suspicious and negative group in the internal cohort according to DARNDEST, of which 383, 62 and 9 patients were diagnosed with local rNPC based on the actual incidence of the internal cohort with the hypothesis of 1000 patients. Meanwhile, 383, 28 and 5 patients would be detected as having local rNPC precisely using DARNDEST. $\S 6,381$ (5,671, 7,231), $\S 87,286$ (49,878, 162,933) and $\S 488,800$ (135,778, 244,000) were needed to detect a patient with local rNPC in each group. Similarly, there would be 480, 89 and 431 patients categorized as positive, suspicious and negative group, of which 391, 35 and 35 patients were diagnosed with local For a single examination in a hypothetical cohort of 1,000 patients, DARNDEST required 190.7 and 184.5 hours in the internal and external cohorts, while the T1_T2 model and conventional method required 129.2 and 226.4 hours, respectively. The total cost of MR examination using DARNDEST was 105.58% and 100.53% of the cost of the T1_T2 model ($\frac{1}{2}$,371,293 vs $\frac{1}{2}$,246,000 and $\frac{1}{2}$,358,723 vs $\frac{1}{2}$,246,000, respectively) in the internal and external cohorts, whereas it was 97.03% and 96.51% ($\frac{1}{2}$,371,293 vs $\frac{1}{2}$,444,000 and $\frac{1}{2}$,358,723 vs $\frac{1}{2}$,2444,000, respectively) of the cost of the conventional method in the internal and external cohorts (Table 2).

Discussion

Our study showcased the viability of unenhanced MRI for detecting local rNPC during post-treatment surveillance using widely adopted deep learning frameworks, DenseNet or ResNet, validated across internal and external cohorts. The combined T1WI and T2WI model exhibited sensitivity comparable to that of radiologists. Error analysis revealed analogous misdiagnosis and missed diagnosis patterns for DenseNet models and radiologists, indicating similar lesion detection mechanisms. Additionally, we introduced an efficient and economical follow-up strategy using DenseNet, DARNDEST, optimizing sensitivity by overlaying the T1WIC model onto the T1_T2 model in positive and suspicious subpopulations while maintaining cost-effectiveness.

Early detection of recurrent NPC is crucial for timely salvage treatment [4]. Routine annual MR examination



Fig 4 Comparison of the economic burden associated with using enhanced MRI in each group based on DARNDEST. The economic analysis was modeled based on the actual prevalence, positive yield of local rNPC, and associated CI derived from this study with a hypothetical cohort of 1,000 patients who underwent follow-up MR examination after radical treatment. For each group in the figure, numbers in bonds denote cases and associated CI categorized as positive, suspicious or negative by DARNDEST. Italicized numbers above the line represent true-positive cases of local rNPC detected by DARNDEST, while those below represent true-positive cases confirmed by histopathology or follow-up MRI. Dark blue text boxes labeled "enhanced/unenhanced" is recommended MR examination methods according to DARNDEST. Data in gridline represent the costs per true-positive case of each group. Cost in China from the Medical Insurance Administration Bureau of Guangzhou, 2024. One dollar = 7.17 RMB. Abbreviation: CV_i or CV_{er} cut-off value of the internal or external set, respectively

of the nasopharynx and neck is essential for NPC surveillance since submucosal or intracranial lesions can be missed during physical examination or endoscopy [4, 8, 22]. Detecting submucosal recurrent lesions and earlystage disease is challenging, and distinguishing recurrent NPC from radiation-induced alterations is complex [4, 12]. Conventional MRI has limited sensitivity and specificity, with 56% sensitivity and 78-83% specificity reported since it only relied on the morphological characteristics [10]. However, our DenseNet models achieved higher sensitivity (83.25-84.26% internal, 80.65-84.95% external) and comparable specificity (82.57-85.59% internal, 82.57-85.32% external). Moreover, the optimized DARNDEST gained an impressively increased sensitivity (internal, 90.36% and external, 90.32%) with mild reduction in specificity (internal, 82.20% and external, 82.57%); still comparable to previous studies and could be supplemented by radiologists.

Additionally, although several studies have demonstrated a lack of added clinical value of using gadolinium contrast agent for routine follow-up scans of benign intracranial tumors (e.g., meningiomas [23], pituitary macroadenoma [24] and vestibular schwannomas [25]), cystic lesions of pancreas [26, 27] and small solid renal masses [28], we have similarly shown that enhanced MRI can be exempted in the initial detection of NPC with the aid of a deep learning model [17]. However, no studies have evaluated unenhanced MRI or abbreviated MRI protocols for imaging follow-up of NPC, nor have they addressed the economic impact of eliminating gadolinium contrast agent. Herein, the economic analysis revealed that using DARNDEST could save significant costs (13.4/15.6 and 76.6/78.2 times) in detecting positive cases of local rNPC in the suspicious or negative groups compared to conventional methods. Moreover, DARNDEST reduced MR examination costs and time by approximately ¥72,707/85,277 (\$10,140/11,894) and 35.7/41.9 hours in a hypothetic cohort of 1,000 patients based on the actual incidence of local rNPC in the internal or external cohorts, respectively. Taken together, this demonstrated for the first time in the era of artificial intelligence, enhanced MR examination may be exempt reasonably in the administration of NPC patients after radical treatment, since not all patients are indicated for

gadolinium contrast agents, especially in those patients with impaired renal function induced by multiple causes, including chemotherapy [29–31]. Particularly, eliminating gadolinium contrast agent for the follow-up of NPC could lead to significant cost and time savings, while improving accessibility and patient tolerance.

Notably, the T1_T2 model exhibited slightly higher or equal sensitivity but lower specificity compared to the radiologists, indicating that unenhanced MRI was sufficient for detecting local lesions in unsupervised situations, whereas enhanced MRI aided in differentiation for radiologists. Particularly, the T1 T2 model performed consistently in both cohorts, while radiologists showed variations, likely due to distinct disease characteristics and scanning parameters between institutions; for example, the proportion of rNPC cases exhibiting necrotic characteristics was higher in the internal cohort than in the external cohort and the postenhancement axial T1-weighted images of the external cohort were fat saturated, different from those of the internal cohort. Obviously, variable data resources have helped to improve model robustness. Moreover, causes of error in misdiagnoses or missed diagnoses among DenseNet models developed using either unenhanced or enhanced MRI were consistent. Specifically, the models tended to misconstrued the asymmetric structure with a mass effect as a recurrent lesion, such as fibrotic scarred tissue induced by radiation, and were prone to miss diagnosis in cases with limited lesions although localized lesion precisely but insufficient confidence, similar to that of junior radiologist. Therefore, we speculated that shape or texture might play a more important role than intensity in the recognition of deep convolution neural networks (DCNNs) in grayscale images. Consistent with previous studies, DCNNs have access to local shape information in the form of local edge relations in object classification [32, 33]. However, further comparative studies focusing on feature extraction in identification using DCNNs are warranted in the future.

The study has some limitations. First, a portion of patients (internal, 24.64%; external, 25.53% and 30.89%) were diagnosed based on time-serial MRI without histopathological confirmation due to challenging biopsy locations. However, debatable cases were excluded initially, and the remaining enrolled cases were confirmed by either detecting size enlargement through follow-up MRIs or observing lesion shrinkage after anti-cancer treatment. Second, the focus was on detecting local rNPC for overall accuracy when developing deep learning models, omitting considerations for metastatic lymph nodes or tumor segmentation. Third, only single-time point axial images were used for model development since the purpose of this study was to determine the

feasibility of detecting local rNPC using unenhanced MRI. Abbreviated protocols should also be explored for follow-up patients in the era of artificial intelligence, where time-serial images and multiple orientation images should be considered in future studies. On the other hand, time-serial MRI examinations during NPC followup could provide additional insights in clinical practice, whereas a single-time point MRI examination may lead to an underestimation of the radiologists' performance in this study. Fourth, clinical data, particularly Epstein-Barr Virus DNA copy numbers, were not incorporated into the models. Future studies should consider comprehensive models that integrate clinical data and timeserial images. Lastly, for cases that might benefit from enhanced MRI, a repetition of the MRI scan with contrast agent could introduce additional costs, which may outweigh the initial savings from eliminating the contrast agent. Therefore, the developed T1_T2 models should be integrated into the scanning system to provide real-time guidance during follow-up scans.

Conclusion

Using unenhanced MRI with deep learning models for detecting local rNPC is feasible. DARNDEST's layer management during follow-up is efficient, economic, timesaving, and minimally toxic.

Abbreviations

| area under curve |
|---|
| contrast-enhanced magnetic resource images |
| deep-learning-assisted recurrent NPC detecting simultaneous |
| tactic |
| digital imaging and communications in medicine |
| deep convolution neural networks |
| negative predictive value |
| positive predictive value |
| recurrent nasopharyngeal carcinoma |
| receiver operating characteristic curve |
| T1-weighted image |
| T2-weighted image |
| post-contrast T1WI |
| |

Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s40644-025-00853-5.

Additional file 1.

Additional file 2.

Acknowledgements

Not applicable.

Authors' contributions

Conceptualization, L.K., and C.L.; Methodology, Y.D., B.J., and B.L.; Investigation, L.K., Y.H., H.W., D.H., W.Q., W.X., and X.L.; Formal analysis, L.K., Y.D., and Y.H.; Writing - Original Draft, L.K., and Y.D.; Writing - Review & Editing, all of the authors; Funding Acquisition, L.K., Y.H., Y.D., and C.L.; Resources, C.X., C.L., and Y.S.; Project administration, L.K., C.X., C.L., and Y.S.

Funding

Study design, data acquisition, quality control of data and algorithms, statistical analysis was supported by the National Natural Science Foundation of China [No. 31900461, 82402382, 82172862] and Guangdong Medical Science and Technology Research Fund Project (No. A2024037). Manuscript preparation and editing was supported by the Guangdong Provincial Health Commission [No. A2023078].

Data availability

De-identified data supporting this study may be shared based on reasonable written request to the corresponding author. Access to de-identified data will require a Data Access Agreement and IRB clearance, which will be considered by the institutions who provided the data for this research.

Declarations

Ethics approval and consent to participate

The study was approved by the institutional review board of the authors (No. G2022-149-01).

Consent for publication

Informed consent was exempted due to retrospective nature for this study.

Competing interests

The authors declare no competing interests.

Author details

¹School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, Guangdong, China. ²State Key Laboratory of Oncology in South China, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou 510060, China. ³Department of Radiology, Sun Yat-Sen University Cancer Center, Guangzhou 510060, China. ⁴Information Technology Center, Sun Yat-Sen University Cancer Center, Guangzhou 510060, China. ⁵Department of Radiation Oncology, The First People's Hospital of Foshan, Foshan 528000, China. ⁶Department of Radiation Oncology, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou 510095, China. ⁷Department of Nasopharyngeal Carcinoma, Sun Yat-Sen University Cancer Center, Guangzhou 510060, China. ⁸Department of Radiation Oncology, Sun Yat-Sen University Cancer Center, Guangzhou 510060, China. ⁹Precision Medicine Center, Sun Yat-Sen University, Guangzhou 510060, China.

Received: 10 September 2024 Accepted: 26 February 2025 Published online: 24 March 2025

References

- Tang LL, Chen WQ, Xue WQ, He YQ, Zheng RS, Zeng YX, et al. Global trends in incidence and mortality of nasopharyngeal carcinoma. Cancer Lett. 2016;374(1):22-30. Epub 2016/02/02. https://doi.org/10.1016/j. canlet.2016.01.040. PubMed PMID: 26828135.
- Li YQ, Tian YM, Tan SH, Liu MZ, Kusumawidjaja G, Ong EHW, et al. Prognostic Model for Stratification of Radioresistant Nasopharynx Carcinoma to Curative Salvage Radiotherapy. J Clin Oncol. 2018;36(9):891-9. Epub 2018/02/08. https://doi.org/10.1200/JCO.2017.75.5165. PubMed PMID: 29412781.
- Ng WT, Wong ECY, Cheung AKW, Chow JCH, Poon DMC, Lai JWY, et al. Patterns of care and treatment outcomes for local recurrence of NPC after definite IMRT-A study by the HKNPCSG. Head Neck. 2019;41(10):3661-9. Epub 2019/07/28. https://doi.org/10.1002/hed.25892. PubMed PMID: 31350940.
- Lee AWM, Ng WT, Chan JYW, Corry J, Mäkitie A, Mendenhall WM, et al. Management of locally recurrent nasopharyngeal carcinoma. Cancer Treat Rev. 2019;79:101890. Epub 2019/08/31. https://doi.org/10.1016/j. ctrv.2019.101890. PubMed PMID: 31470314.
- Dionisi F, Croci S, Giacomelli I, Cianchetti M, Caldara A, Bertolin M, et al. Clinical results of proton therapy reirradiation for recurrent nasopharyngeal carcinoma. Acta Oncol. 2019;58(9):1238-45. Epub 2019/06/04.

https://doi.org/10.1080/0284186x.2019.1622772. PubMed PMID: 31155998.

- Wong EHC, Liew YT, Loong SP, Prepageran N. Five-year Survival Data on the Role of Endoscopic Endonasal Nasopharyngectomy in Advanced Recurrent rT3 and rT4 Nasopharyngeal Carcinoma. Ann Otol Rhinol Laryngol. 2020;129(3):287-93. Epub 2019/11/09. https://doi.org/10.1177/ 0003489419887410. PubMed PMID: 31701754.
- Thamboo A, Patel VS, Hwang PH. 5-year outcomes of salvage endoscopic nasopharyngectomy for recurrent nasopharyngeal carcinoma. J Otolaryngol Head Neck Surg. 2021;50(1):12 (PubMed PMID: 33597031).
- Ng SH, Chang JT, Ko SF, Wan YL, Tang LM, Chen WC. MRI in recurrent nasopharyngeal carcinoma. Neuroradiology. 1999;41(11):855-62. Epub 1999/12/22. https://doi.org/10.1007/s002340050857. PubMed PMID: 10602863.
- Chong VF, Fan YF, Khoo JB. Nasopharyngeal carcinoma with intracranial spread: CT and MR characteristics. J Comput Assist Tomogr. 1996;20(4):563–9 (Epub 1996/07/01 PubMed PMID: 8708057).
- Chong VF, Fan YF. Detection of recurrent nasopharyngeal carcinoma: MR imaging versus CT. Radiology. 1997;202(2):463-70. Epub 1997/02/01. https://doi.org/10.1148/radiology.202.2.9015075. PubMed PMID: 9015075.
- Li Z, Li Y, Li N, Shen L. Positron emission tomography/computed tomography outperforms MRI in the diagnosis of local recurrence and residue of nasopharyngeal carcinoma: An update evidence from 44 studies. Cancer Med. 2019;8(1):67–79 (PubMed PMID: 30578604).
- Lai V, Khong PL. Updates on MR imaging and 1⁸F-FDG PET/CT imaging in nasopharyngeal carcinoma. Oral Oncol. 2014;50(6):539-48. Epub 2013 Jun 14. https://doi.org/10.1016/j.oraloncology.2013.05.005.
- Lin GW, Wang LX, Ji M, Qian HZ. The use of MR imaging to detect residual versus recurrent nasopharyngeal carcinoma following treatment with radiation therapy. Eur J Radiol. 2013;82(12):2240-6. Epub 2013/10/10. https://doi.org/10.1016/j.ejrad.2013.09.014. PubMed PMID: 24103355.
- Zhang MX, Li J, Shen GP, Zou X, Xu JJ, Jiang R, et al. Intensity-modulated radiotherapy prolongs the survival of patients with nasopharyngeal carcinoma compared with conventional two-dimensional radiotherapy: A 10-year experience with a large cohort and long follow-up. Eur J Cancer. 2015;51(17):2587-95. Epub 2015/09/01. https://doi.org/10.1016/j.ejca. 2015.08.006. PubMed PMID: 26318726.
- Moon SH, Cho KH, Lee CG, Keum KC, Kim YS, Wu HG, et al. IMRT vs. 2D-radiotherapy or 3D-conformal radiotherapy of nasopharyngeal carcinoma : Survival outcome in a Korean multi-institutional retrospective study (KROG 11-06). Strahlenther Onkol. 2016;192(6):377-85. Epub 2016/03/15. https://doi.org/10.1007/s00066-016-0959-y. PubMed PMID: 26972085.
- Ma Z, Umezawa R, Yamamoto T, Ishikawa Y, Takahashi N, Takeda K, et al. IMRT improves local control in patients with nasopharyngeal carcinoma compared with conventional radiotherapy: propensity score-matched analysis. Jpn J Clin Oncol. 2021;51(9):1444-51. Epub 2021/07/13. https:// doi.org/10.1093/jjco/hyab100. PubMed PMID: 34250545.
- Deng Y, Li C, Lv X, Xia W, Shen L, Jing B, et al. The contrast-enhanced MRI can be substituted by unenhanced MRI in identifying and automatically segmenting primary nasopharyngeal carcinoma with the aid of deep learning models: An exploratory study in large-scale population of endemic area. Comput Methods Programs Biomed. 2022;217:106702. Epub 2022/03/02. https://doi.org/10.1016/j.cmpb.2022.106702. PubMed PMID: 35228147.
- Wong LM, King AD, Ai QYH, Lam WKJ, Poon DMC, Ma BBY, et al. Convolutional neural network for discriminating nasopharyngeal carcinoma and benign hyperplasia on MRI. Eur Radiol. 2021;31(6):3856-63. Epub 2020/11/27. https://doi.org/10.1007/s00330-020-07451-y. PubMed PMID: 33241522.
- Deng Y, Huang Y, Jing B, Wu H, Qiu W, Chen H, et al. Deep learning-based recurrence detector on magnetic resonance scans in nasopharyngeal carcinoma: A multicenter study. Eur J Radiol. 2023;168:111084. Epub 2023/09/18. https://doi.org/10.1016/j.ejrad.2023.111084. PubMed PMID: 37722143.
- He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV, USA, 2016, pp. 770-778, https://doi.org/10.1109/ CVPR.2016.90.

- Pérez-García F, Sparks R, Ourselin S. TorchIO: A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Comput Methods Programs Biomed. 2021;208:106236 (PubMed PMID: 34311413).
- Sham JS, Wei WI, Kwan WH, Chan CW, Kwong WK, Choy D. Nasopharyngeal carcinoma. Pattern of tumor regression after radiotherapy. Cancer. 1990;65(2):216-20. Epub 1990/01/15. https://doi.org/10.1002/1097-0142(19900115)65:2<216::aid-cncr2820650206>3.0.co;2-z. PubMed PMID: 2295044.
- Boto J, Guatta R, Fitsiori A, Hofmeister J, Meling TR, Vargas MI. Is Contrast Medium Really Needed for Follow-up MRI of Untreated Intracranial Meningiomas? AJNR Am J Neuroradiol. 2021;42(8):1421–8 (PubMed PMID: 34117017).
- Alali AA, Hanagandi PB, Maralani PJ. Do We Need Gadolinium-Based Contrast Agents for Routine MRI Surveillance of Unoperated Pituitary Macroadenoma? AJNR Am J Neuroradiol. 2022;43(7):1024–8 (PubMed PMID: 35738673).
- Buch K, Juliano A, Stankovic KM, Curtin HD, Cunnane MB. Noncontrast vestibular schwannoma surveillance imaging including an MR cisternographic sequence: is there a need for postcontrast imaging? J Neurosurg. 2018;131(2):549-54. Epub 2018/08/11. https://doi.org/10.3171/2018.3. Jns1866. PubMed PMID: 30095338.
- Macari M, Lee T, Kim S, Jacobs S, Megibow AJ, Hajdu C, et al. Is gadolinium necessary for MRI follow-up evaluation of cystic lesions in the pancreas? Preliminary results. AJR Am J Roentgenol. 2009;192(1):159-64. Epub 2008/12/23. https://doi.org/10.2214/ajr.08.1068. PubMed PMID: 19098196.
- Nougaret S, Reinhold C, Chong J, Escal L, Mercier G, Fabre JM, et al. Incidental pancreatic cysts: natural history and diagnostic accuracy of a limited serial pancreatic cyst MRI protocol. Eur Radiol. 2014;24(5):1020-9. Epub 2014/02/27. https://doi.org/10.1007/s00330-014-3112-2. PubMed PMID: 24569848.
- Mawi H, Narine R, Schieda N. Adequacy of Unenhanced MRI for Surveillance of Small (Clinical T1a) Solid Renal Masses. AJR Am J Roentgenol. 2021;216(4):960-6. Epub 2021/02/18. https://doi.org/10.2214/ajr.20.23458. PubMed PMID: 33594909.
- Boehm I, Hungerbühler M, Heverhagen JT. Insight into the dynamic of gadolinium based contrast agent (GBCA) hypersensitivity: Acquisition, persistence and disappearance. Magnetic resonance imaging. 2018;49:1-3. Epub 2017/12/17. https://doi.org/10.1016/j.mri.2017.12.013. PubMed PMID: 29247752.
- Kanda T, Ishii K, Kawaguchi H, Kitajima K, Takenaka D. High signal intensity in the dentate nucleus and globus pallidus on unenhanced T1-weighted MR images: relationship with increasing cumulative dose of a gadolinium-based contrast material. Radiology. 2014;270(3):834-41. Epub 2014/01/31. https://doi.org/10.1148/radiol.13131669. PubMed PMID: 24475844.
- Marckmann P, Skov L, Rossen K, Dupont A, Damholt MB, Heaf JG, et al. Nephrogenic systemic fibrosis: suspected causative role of gadodiamide used for contrast-enhanced magnetic resonance imaging. J Am Soc Nephrol. 2006;17(9):2359-62. Epub 2006/08/04. https://doi.org/10.1681/ ASN.2006060601. PubMed PMID: 16885403.
- Baker N, Lu H, Erlikhman G, Kellman PJ. Deep convolutional networks do not classify based on global object shape. PLoS Comput Biol. 2018;14(12):e1006613 (PubMed PMID: 30532273).
- Jacob G, Pramod RT, Katti H, Arun SP. Qualitative similarities and differences in visual object representations between brains and deep networks. Nature Communications. 2021;12(1):1872. https://doi.org/10. 1038/s41467-021-22078-3.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.