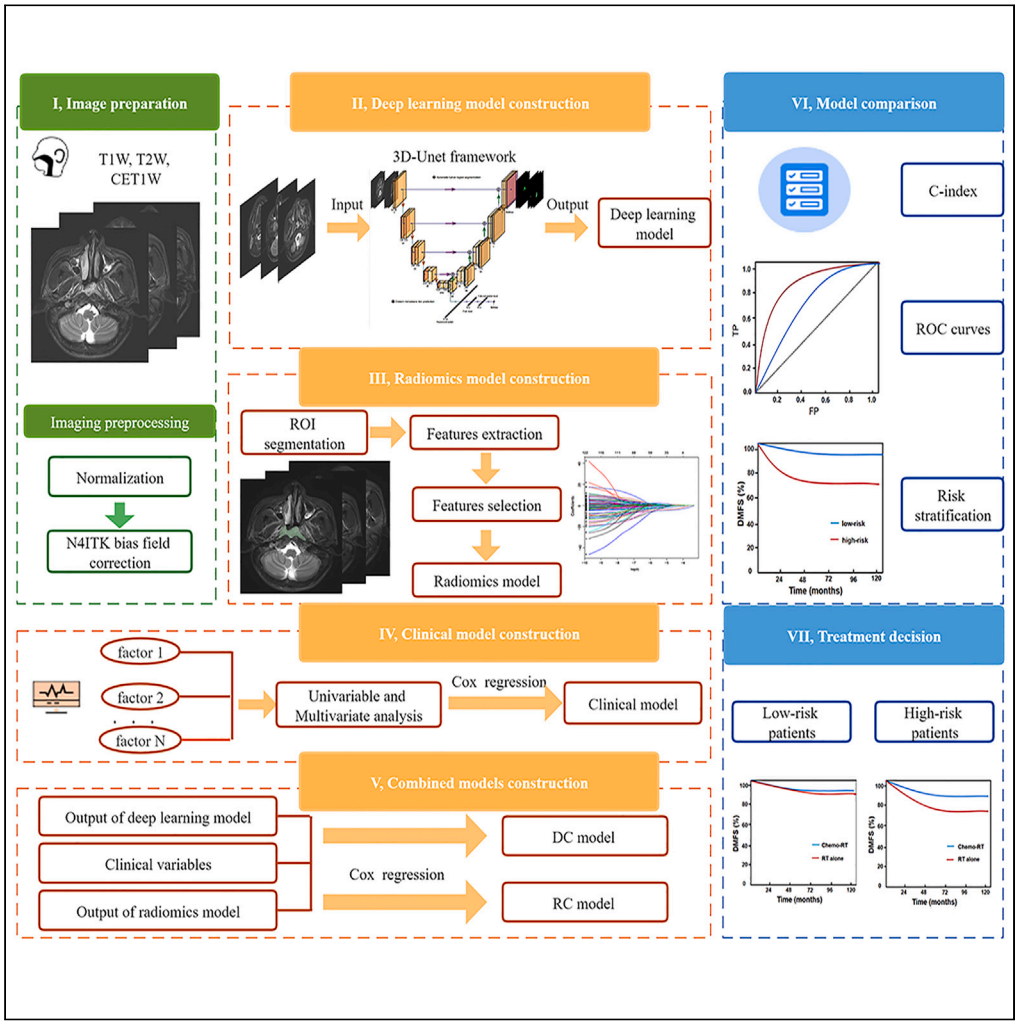
iScience

CellPress
OPEN ACCESS

Article

MRI-based deep learning model predicts distant metastasis and chemotherapy benefit in stage II nasopharyngeal carcinoma



Yu-Jun Hu, Lin Zhang, You-Ping Xiao, ..., Jian-Ji Pan, Jin-Gao Li, Yun-Fei Xia

panjianji@126.com (J.-J.P.)
lijingao@hotmail.com (J.-G.L.)
xiayf@sysucc.org.cn (Y.-F.X.)

Highlights

An MRI-based deep learning model was constructed for stage II NPC using 3D-Unet

The deep learning model is powerful in assessing distant metastasis risk

The deep learning model has the potential to be a recommender of treatment decisions

Hu et al., iScience 26, 106932
June 16, 2023
https://doi.org/10.1016/j.isci.2023.106932

iScience

CellPress
OPEN ACCESS

Article

MRI-based deep learning model predicts distant metastasis and chemotherapy benefit in stage II nasopharyngeal carcinoma

Yu-Jun Hu,[1,2,11] Lin Zhang,[3,4,11] You-Ping Xiao,[5,11] Tian-Zhu Lu,[3,4,6] Qiao-Juan Guo,[7,8] Shao-Jun Lin,[7,8] Lan Liu,[9] Yun-Bin Chen,[5] Zi-Lu Huang,[1,2] Ya Liu,[1,2] Yong Su,[3,4] Li-Zhi Liu,[10] Xiao-Chang Gong,[3,4,6] Jian-Ji Pan,[7,8,*] Jin-Gao Li,[3,4,6,*] and Yun-Fei Xia[1,2,12,*]

SUMMARY

Chemotherapy remains controversial for stage II nasopharyngeal carcinoma because of its considerable prognostic heterogeneity. We aimed to develop an MRI-based deep learning model for predicting distant metastasis and assessing chemotherapy efficacy in stage II nasopharyngeal carcinoma. This multicenter retrospective study enrolled 1072 patients from three Chinese centers for training (Center 1, n = 575) and external validation (Centers 2 and 3, n = 497). The deep learning model significantly predicted the risk of distant metastases for stage II nasopharyngeal carcinoma and was validated in the external validation cohort. In addition, the deep learning model outperformed the clinical and radiomics models in terms of predictive performance. Furthermore, the deep learning model facilitates the identification of high-risk patients who could benefit from chemotherapy, providing useful additional information for individualized treatment decisions.

INTRODUCTION

The benefit of chemotherapy for stage II nasopharyngeal carcinoma (NPC) in the intensity-modulated radiation therapy (IMRT) era has been controversial because the 5-year overall survival (OS) and distant metastasis-free survival (DMFS) have reached over 90% in recent years.[1–3] However, some studies[4,5] have shown that survival outcomes vary significantly among stage II patients, especially for distant metastasis (DM). Stratified treatment is proposed in the latest version of the National Comprehensive Cancer Network guidelines to better manage patients with stage II NPC.[6] For patients with T2N0 disease, radiotherapy alone is routinely recommended, whereas concurrent chemoradiotherapy is administered in the presence of high-risk factors [e.g., bulky tumor volume or high Epstein-Barr virus deoxyribonucleic acid (EBV DNA)].[6] Besides concurrent chemoradiotherapy, induction or adjuvant chemotherapy is recommended for patients with T1-2N1 disease with adverse features.[6]

EBV DNA is a robust prognostic marker with potential clinical applications in NPC and is routinely detected.[7] However, interlaboratory detection of EBV DNA varies considerably for the same test using identical procedures, and the primer/probe sets are not uniform.[8] On the other hand, the lack of an objective, standardized method to describe and measure bulky tumor volumes limits its application in clinical practice.[9] An important finding from several studies[10–12] has revealed that radiologic extranodal extension (rENE) is a powerful imaging biomarker for risk stratification in NPC. Nevertheless, due to the use of non-standardized diagnostic criteria, the incidence of rENE varies from 7.7% to 75.6%.[13–15] Thus, identifying additional biomarkers that can help predict prognosis and developing individualized treatment strategies are critical.

Radiomics is an emerging field that aims to extract high-throughput quantitative features from medical images and reveal the underlying pathophysiology of diseases.[16] The introduction of deep learning (DL) enables radiomics to extract information quickly and precisely from biomedical images by using deep convolutional neural networks in a fully automated manner.[17] Emerging evidence suggests that DL has remarkably improved the diagnostic, prognostic, and therapeutic responses of various cancer types.[18–20]

[1]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou, China

[2]Department of Radiation Oncology, Sun Yat-Sen University Cancer Center, Guangzhou, China

[3]Department of Radiation Oncology, Jiangxi Cancer Hospital of Nanchang University, Nanchang, Jiangxi, China

[4]NHC Key Laboratory of Personalized Diagnosis and Treatment of Nasopharyngeal Carcinoma (Jiangxi Cancer Hospital of Nanchang University), Jiangxi, China

[5]Department of Radiology, Fujian Medical University Cancer Hospital & Fujian Cancer Hospital, Fuzhou, China

[6]Jiangxi Key Laboratory of Translational Cancer Research, Jiangxi Cancer Hospital of Nanchang University, Jiangxi, China

[7]Department of Radiation Oncology, Fujian Medical University Cancer Hospital & Fujian Cancer Hospital, Fuzhou, China

[8]Fujian Key Laboratory of Translational Cancer Medicine, Fuzhou, China

[9]Department of Radiology, Jiangxi Cancer Hospital of Nanchang University, Nanchang, China

[10]Department of Radiology, Sun Yat-Sen University Cancer Center, Guangzhou, China

[11]These authors contributed equally

Continued

To date, most DL applications in NPC have focused on predicting prognosis and exploring the benefits of chemotherapy in locally advanced NPC,[21–23] predicting recurrence,[24] and tumor region segmentation.[25,26] However, no study has reported the value of a MRI-based deep learning model for stage II NPC.

In this study, we aimed to develop and externally validate a multiparametric MRI-based DL model for predicting distant metastases in patients with stage II NPC. Furthermore, based on the DL model, we sought to identify patients most likely to benefit from chemotherapy.

RESULTS

Baseline characteristics

A total of 1072 patients from three Chinese hospitals were recruited (Figure 1). The clinicopathological characteristics of patients in the training cohort [n = 575; 395 (68.7%) men] and external validation cohort [n = 497; 326 (65.6%) men] are summarized in Table 1. The median age was 44 years (interquartile range, 38−51 years) in the training cohort and 48 years (interquartile range, 40−56 years) in the validation cohort.

Construction of models

To construct the clinical model, T category, rENE, and sex, which showed statistically significant differences in the univariable and the multivariable analysis of DM (Table 2), were included in the Cox proportional hazards regression analysis. For the radiomics model, a total of 1395 features were extracted, with each MRI sequence generating 465 features. There were three types of extracted radiomics features from each sequence: 216 first-order statistical features, 42 shape-based features, and 207 texture features. After LASSO regression, four, nine, six, and four features were selected from the axial T1W, T2W, CET1W, and the combined sequence, respectively, which were strongly associated with DM. The radiomics models of each sequence and the combined sequence were created separately. Table 3 summarizes the mean Dice similarity coefficient for the three sequences of the 3D-Unet framework segmentation structure, which ranged from 0.82 to 0.85. The DL models for each sequence and the combined sequence were determined by the output scores of the 3D-Unet framework.

Comparison of the predictive accuracy of models

We calculated the C-indexes and plotted 5-year TD-ROC curves (Figure 2) to compare the accuracy of the models in predicting DM. The DL model and radiomics model based on the combined sequence yielded higher C-indexes than the single MR sequence in the training cohort (DL $_{T1W}$ vs. DL $_{T2W}$ vs. DL $_{CET1W}$ vs. DL $_{combined}$: 0.79 vs. 0.72 vs. 0.78 vs. 0.85; radiomics $_{T1W}$ vs. radiomics $_{T2W}$ vs. radiomics $_{CET1W}$ vs. radiomics $_{combined}$: 0.57 vs. 0.56 vs. 0.64 vs. 0.66) and the validation cohort (DL $_{T1W}$ vs. DL $_{T2W}$ vs. DL $_{CET1W}$ vs. DL $_{combined}$: 0.80 vs. 0.78 vs. 0.71 vs. 0.84; radiomics $_{T1W}$ vs. radiomics $_{T2W}$ vs. radiomics $_{CET1W}$ vs. radiomics $_{combined}$: 0.61 vs. 0.69 vs. 0.68 vs. 0.71), respectively. Then, we utilized Cox proportional hazards regression analysis to integrate crucial clinical features, including T category, rENE, and sex, into the output scores obtained from the DL $_{combined}$ and radiomics $_{combined}$ models for developing DC and RC models in the training cohort.

As shown in Table 4, the C-index and area under curve (AUC) of the radiomics $_{combined}$ model were lower than those of clinical, DL $_{combined}$, RC, and DC models in both training and validation cohorts. The DC model yields the highest C-index and AUC value compared to other models in the training (C-index: 0.89, 95% CI:0.84–0.94; AUC: 0.90, 95% CI: 0.85–0.96) and validation (C-index: 0.87, 95% CI:0.80–0.93; AUC: 0.85, 95% CI: 0.78–0.93) cohorts. Besides, the DC model had a significantly higher C-index and AUC value than the DL $_{combined}$ model in the training cohort (C-index: 0.89 vs. 0.85, p = 0.02; AUC: 0.90 vs. 0.85, p = 0.007); however, consistent results were not obtained in the external validation cohort (C-index: 0.87 vs. 0.84, p = 0.76; AUC: 0.85 vs. 0.84, p = 0.75). The C-index and AUC of the RC model were not significantly improved compared to those in the clinical model in both training (C-index: 0.78 vs. 0.77, p = 0.65; AUC: 0.80 vs. 0.78, p = 0.53) and validation (C-index: 0.78 vs. 0.73, p = 0.28; AUC: 0.74 vs. 0.72, p = 0.65) cohorts.

Comparison of the discrimination ability of the models

Patients were then divided into high- and low-risk groups according to the optimal cutoff value of each model (DL $_{combined}$ model: 0.501, clinical model: −0.262, radiomics $_{combined}$ model: −5.245, RC model: −1.42, DC model: 5.11). Kaplan−Meier curves of DMFS and OS were plotted to identify the association between the models and prognosis. In each model, high-risk patients had a significantly worse DMFS

[12]Lead contact

*Correspondence:
panjianji@126.com (J.-J.P.),
lijingao@hotmail.com
(J.-G.L.),
xiayf@sysucc.org.cn (Y.-F.X.)

https://doi.org/10.1016/j.isci.2023.106932

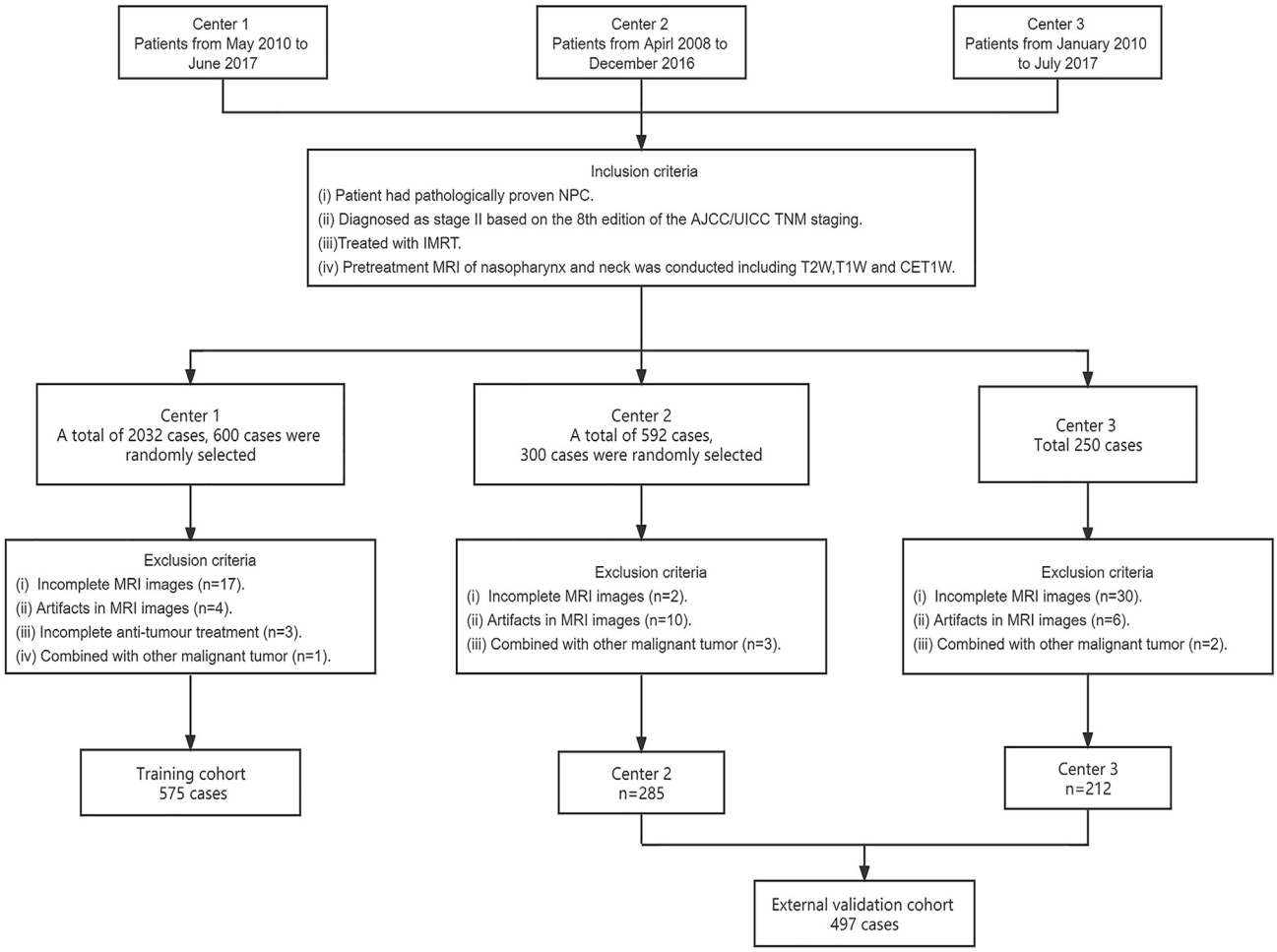


Figure 1. Flowchart of the selection process of patients from three centers

Abbreviation: NPC, nasopharyngeal carcinoma; AJCC/UICC, American Joint Committee on Cancer/Union for International Cancer Control; IMRT, intensity-modulated radiation therapy; MRI, magnetic resonance imaging.

compared to that for low-risk patients in the training cohort (DL $_{combined}$ model: hazard ratio (HR), 16.05; 95% CI, 7.68–33.57; $p < 0.001$; radiomics $_{combined}$ model: HR, 3.71; 95% CI, 2.00–6.88; $p < 0.001$; clinical model: HR, 10.98; 95% CI, 3.92–30.77; $p < 0.001$; RC model: HR: 2.73; 95% CI, 2.05–3.63; $p < 0.001$; DC model: HR, 20.73; 95% CI: 10.89–39.44; $p < 0.001$; Figures 3A–3E) and the validation cohort (DL $_{combined}$ model: HR, 14.49; 95% CI, 7.20–29.14; $p < 0.001$; radiomics $_{combined}$ model: HR, 4.65; 95% CI, 2.47–8.77; $p < 0.001$; clinical model: HR, 3.62; 95% CI, 1.72–7.62; $p < 0.001$; RC model: HR, 2.57; 95% CI: 1.88–3.52; $p < 0.001$; DC model: HR, 16.11; 95% CI: 8.56–30.32; $p < 0.001$; Figures 3F–3J). A similar trend was observed for OS in the training and validation cohorts (Figures S1; for all, $p < 0.05$).

Subgroups analysis based on the DL $_{combined}$ model

Considering that there was no significant improvement in the predictive accuracy of the DC and RC models, subsequent subgroup analyses and analyses of chemotherapy benefits were performed using the DL $_{combined}$ model. Subgroup analyses were conducted within the stratification factors (age, sex, rENE, EBV DNA, and clinical stage) in the training and validation cohorts to assess the performance of the DL $_{combined}$ model in predicting DM. The forest plots (Figure 4) show that the DL $_{combined}$ model had promising predictive ability in most subgroups, except for the T2N0 subgroup in the training (HR: 3.73, 95% CI: 0.30–45.93, $p = 0.27$) and validation (HR: 5.07, 95% CI: 0.41–62.62, $p = 0.18$) cohorts. However, the *p* values for interaction in the clinical stage and other subgroups were all $p > 0.05$, indicating no significant prediction difference of the DL $_{combined}$ model among subgroups in both training and validation cohorts (Figures 4A and 4B). Of note, in the T2N0 subset, the prognosis did not improve in patients who

Table 1. Baseline characteristics of 1072 patients in training and validation cohort

Characteristics	Training cohort (n = 575)	Validation cohort (n = 497)	*p* value
Age, median (IQR), years	44 (38–51)	48 (40–56)	<0.001
Sex assigned at birth, No. (%)			0.28
Male	395 (68.7)	326 (65.6)	
Female	180 (31.3)	171 (34.4)	
Histopathology, No. (%)			<0.001
WHO I	0 (0)	55 (11.1)	
WHO II-III	575 (100)	442 (88.9)	
T[a], No. (%)			0.15
T1	301 (52.3)	238 (47.9)	
T2	274 (47.7)	259 (52.1)	
N[a], No. (%)			0.32
N0	64 (11.1)	46 (9.3)	
N1	511 (88.9)	451 (90.7)	
LDH[b], No. (%)			<0.001
Normal	554 (96.3)	467 (94.0)	
Abnormal	21 (3.7)	9 (1.8)	
Unknown	0 (0)	21 (4.2)	
EBV DNA[c], No. (%)			<0.001
Undetectable	311 (54.1)	310 (62.4)	
Detectable	238 (41.4)	122 (24.5)	
Unknown	26 (4.5)	65 (13.1)	
rENE, No. (%)			0.13
Without	436 (75.8)	399 (80.3)	
Coalescent nodes	120 (20.9)	87 (17.5)	
Adjacent structures infiltration	19 (3.3)	11 (2.2)	
Chemotherapy, No. (%)			<0.001
Without	175 (30.4)	102 (20.5)	
DDP/NDP-based	367 (63.9)	347 (69.8)	
Non-DDP/NDP-based	33 (5.7)	48 (9.7)	
5-year DMFS (%) (95% CI)	93.5 (91.4–95.5)	92.9 (90.6–95.2)	0.70
5-year OS (%) (95% CI)	92.2 (90.0–94.5)	94.5 (92.4–96.6)	0.09
FU time (month)[d]	80.0 (69.0–102.0)	77.0 (63.0–97.0)	NA

Abbreviations: IQR: interquartile range; WHO, World Health Organization; LDH, lactate dehydrogenase; EBV DNA, Epstein-Barr virus deoxyribonucleic acid; rENE, radiologic extranodal extension; DDP, cisplatin; NDP, nedaplatin; DMFS, distant metastasis-free survival; OS, overall survival; FU, follow-up.

[a]According to the eighth edition of the American Joint Committee on Cancer/Union for International Cancer Control cancer staging manual.

[b]Abnormal, center 1: >245 U/L, center 2 and center 3: >250U/L.

[c]Detectable thresholds, center 1: <1000copy/mL, center 2: <500copy/mL, center 3: <500copy/mL.

[d]Data are represented as median (IQR).

received chemoradiotherapy compared to those who received radiotherapy alone in terms of 5-year DMFS (97.8% vs. 98.3%, p = 0.85) and OS (93.4% vs. 92.7%, p = 0.98) (Figures 5A and 5B).

Chemotherapy benefits in patients with T1-2N1 disease based on the DL $_{combined}$ model

We conducted an exploratory analysis to investigate the value of chemotherapy in low- and high-risk patients based on the DL $_{combined}$ model. The DL $_{combined}$ model was unable to stratify risk in the patients with T2N0 disease; therefore, 886 patients with T1-2N1 disease who received cisplatin/nedaplatin-based chemotherapy in the training and validation cohorts were included in the subsequent analysis. The

Table 2. Univariate and multivariable analysis of DM in the training cohort for clinical model construction

	Univariate analysis			Multivariate analysis		
Covariate	HR	95% CI	*p* value	HR	95% CI	*p* value
Age, years	1.00	0.97–1.03	0.88	0.99	0.96–1.03	0.68
Sex assigned at birth						
Male	Reference			Reference		
Female	0.29	0.11–0.74	0.009	0.35	0.13–0.90	0.03
T[a] category						
T1	Reference			Reference		
T2	2.27	1.19–4.30	0.01	2.54	1.27–5.08	0.008
N[a] category						
N0	Reference			Reference		
N1	2.61	0.63–10.80	0.19	1.82	0.38–8.65	0.45
LDH[b]						
Normal	Reference			Reference		
Abnormal	2.25	0.70–7.29	0.18	2.20	0.67–7.24	0.19
EBV DNA[c]						
Undetectable	Reference			Reference		
Detectable	3.24	1.65–6.37	0.001	1.74	0.83–3.64	0.14
rENE						
Without	Reference			Reference		
G1	3.82	1.95–7.49	<0.001	2.98	1.43–6.20	0.004
G2	15.53	6.67–36.14	<0.001	9.51	3.62–24.96	<0.001
Chemotherapy						
No	Reference			Reference		
Yes	1.90	0.88–4.10	0.10	1.01	0.44–2.35	0.97

G1: Coalescent nodes; G2: Adjacent structures infiltration.
Abbreviations: DMFS, distant metastasis-free survival; OS, overall survival; HR, hazard ratio; CI, confidence interval; LDH, Lactate dehydrogenase; EBV DNA, Epstein-Barr virus deoxyribonucleic acid; rENE, radiologic extranodal extension.
[a]According to the eighth edition of the American Joint Committee on Cancer/Union for International Cancer Control cancer staging manual.
[b]Abnormal, center 1: >245 U/L, center 2 and center 3: >250U/L.
[c]Detectable thresholds, center 1: <1000copy/mL, center 2: <500copy/mL, center 3: <500copy/mL.

cumulative cisplatin/nedaplatin dose (CCND) data were collected. We adopted 160 mg/m^2 as the threshold for the CCND because it was the median value. In the high-risk group, patients who received a CCND >160 mg/m^2 had a better DMFS but not OS compared to those who received a CCND of ≤ 160 mg/m^2 or radiotherapy alone (5-year DMFS: 83.5% vs. 60.0% vs.72.1%, p = 0.02; OS: 88.7% vs. 72.7% vs. 78.1%, p = 0.06; Figures 5E and 5F). However, in the low-risk group, the 5-year DMFS and OS rates were similar among patients receiving radiotherapy alone, receiving a CCND ≤160 mg/m^2, and receiving a CCND >160 mg/m^2 (DMFS: 98.8% vs. 98.0% vs. 98.1%, p = 0.57; OS: 97.1% vs. 97.0% vs. 96.6%, p = 0.98; Figures 5C and 5D).

DISCUSSION

In this multicenter study, we developed a DL $_{combined}$ model to predict DM in patients with stage II NPC. The DL $_{combined}$ model showed encouraging results in both the training and validation cohorts, accurately classifying patients into high- and low-risk groups. High-risk patients with T1-2N1 disease could benefit from chemotherapy with a CCND >160 mg/m^2 in terms of reducing DM, whereas low-risk patients with T1-2N1 disease and patients with T2N0 disease did no benefit from chemotherapy.

ResNet[27] and CNN[22] networks have been used in studies focusing on prognosis prediction. In contrast to the networks used in these studies, the 3D-Unet model in the current study implements automatic

Table 3. Performance of 3D-Unet framework in primary tumor and metastatic lymph nodes segmentation

Mean DSC	Training cohort	Validation cohort
CET1W	0.84 ± 0.03	0.84 ± 0.03
T1W	0.84 ± 0.04	0.85 ± 0.01
T2W	0.83 ± 0.02	0.82 ± 0.02

Abbreviations: DSC, Dice similarity coefficient; T1W, axial T1-weighted; T2W, axial T2-weighted; CET1W, contrast-enhanced axial T1-weighted.
Data are represented as mean ± SD.

segmentation of the tumor and metastatic lymph node, using skip connections between the encoder and decoder to merge low- and high-level features, which ultimately allows finer image details to be retained for predicting DM.[28] Based on a promising automatic segmentation, the DL model achieved encouraging results in predicting DM.

The DL and radiomics models from combined sequences had better prognostic performance than did either sequence alone. The use of three axial sequences may have allowed for the comprehensive capture of microscopic and macroscopic features, reflecting the biological behavior of tumors. Shen et al.[29] and Zhang et al.[30] both proposed a radiomics model based on only several handcrafted features for locally advanced NPC (the C-index ranged from 0.686 to 0.758). In addition, radiomics relies on the precise delineation of the tumor boundary by oncologists, which suffers from an interrater bias and the potential to miss important peritumoral microenvironment information, besides being time-consuming. In comparison, DL can learn complex and non-linear functions between the inputs and output labels by combining numerous nodes and layers while tuning parameters to maximize prediction accuracy.[17] Then, the DL framework could maximally and comprehensively extract higher-order features for more accurate predictions.

Accurate prediction of prognosis is crucial for the management of patients with cancer. The DL model in the current study successfully differentiated patients into high- and low-risk subgroups, with significant differences in DM risk between the two groups. For low-risk patients with T1-2N1 disease, radiotherapy alone may be sufficient because no benefit can be derived from chemotherapy. In a recent clinical trial, Tang et al.[31] similarly reported that radiotherapy alone is practicable for intermediate-risk patients with NPC (stage II and T3N0 without adverse features such as rENE or EBV DNA load ≥4000 copies/mL). Nevertheless, the assessment of ENE currently relies on imaging and lacks consistent diagnostic criteria, thereby leading to varying subjective results among radiologists.[10–14] In addition, no internationally accepted standardized test procedure exists for EBV DNA testing. Given these aspects, the clinical generalizability and applicability of the trial results by Tang et al.[31] are limited. Unlike unstable clinical factors, our DL model

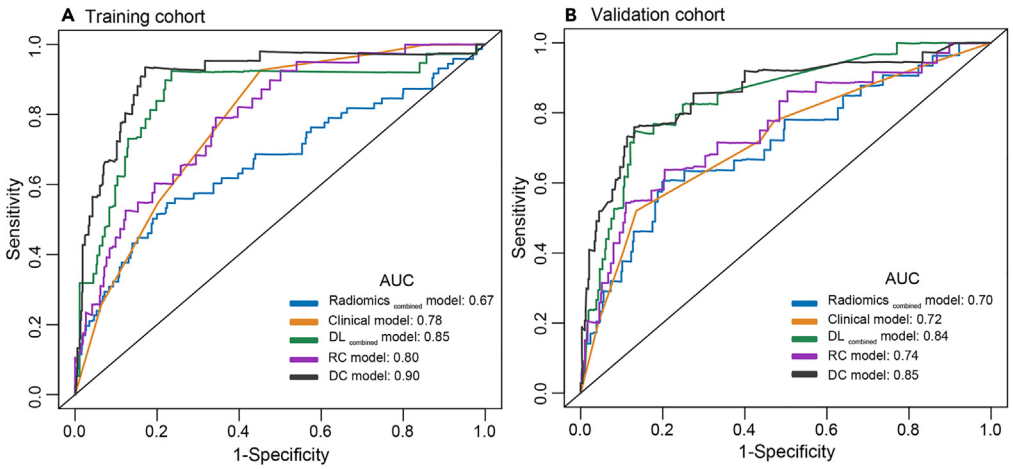


Figure 2. TD–ROC curves of the models

(A) and (B) present the TD-ROC curves of models for DMFS in the training and validation cohorts. Abbreviations: TD-ROC, time-dependent receiver operating characteristic; DMFS, distant metastasis-free survival; AUC, area under curve.

Table 4. Performance of models in predicting DMFS

	Training cohort			Validation cohort		
Models	C-index	95% CI	*p* value	C-index	95% CI	*p* value
Radiomics $_{combined}$ model	0.66	0.59–0.73	Ref	0.71	0.62–0.80	Ref
Clinical model	0.77	0.71–0.83	0.005	0.73	0.64–0.82	0.97
DL $_{combined}$ model	0.85	0.78–0.92	<0.001	0.84	0.78–0.90	0.02
RC model	0.78	0.71–0.84	<0.001	0.78	0.70–0.86	0.43
DC model	0.89	0.84–0.94	<0.001	0.87	0.80–0.93	0.009
DC model vs. DL $_{combined}$ model	–	–	0.02	–	–	0.76
RC model vs. Clinical model	–	–	0.65	–	–	0.28
	AUC	95% CI	*p* value	AUC	95% CI	*p* value
Radiomics $_{combined}$ model	0.67	0.59–0.74	Ref	0.70	0.60–0.80	Ref
Clinical model	0.78	0.73–0.85	0.01	0.72	0.62–0.81	0.78
DL $_{combined}$ model	0.85	0.78–0.93	<0.001	0.84	0.77–0.91	0.02
RC model	0.80	0.73–0.87	<0.001	0.74	0.65–0.84	0.36
DC model	0.90	0.85–0.96	<0.001	0.85	0.78–0.93	0.01
DC model vs. DL $_{combined}$ model	–	–	0.007	–	–	0.75
RC model vs. Clinical model	–	–	0.53	–	–	0.65

Note that, DL $_{combined}$ model and Radiomics $_{combined}$ model were conducted based on the three MR sequences (T1W, T2W, and CET1W). DC model, a model combining deep learning and clinical variables. RC model, a model combining radiomics and clinical variables.

Abbreviations: NPC, nasopharyngeal carcinoma; C-index, Harrell's concordance index; CI: confidence interval; DL, deep learning; Ref, reference; AUC, area under curve.

provides a more objective and accurate prognostic stratification approach to bridge this gap. In addition, the DL model is based on MR images, which are routinely obtained before treatment and are noninvasive. Therefore, the DL model can be used as an effective and practical tool to predict DMFS for patients with stage II NPC.

Notably, the DC and RC models that were constructed by combining the clinical parameters with the DL $_{combined}$ model and radiomics $_{combined}$ model, respectively, did not significantly improve the predictive performance. To be widely used and accepted, predictive models should be simple and practical while accurately predicting prognosis.[32] However, introducing a new clinical parameter into the DL $_{combined}$ model adds extra uncertainty and complexity in practice. In brief, our proposed DL $_{combined}$ model provides the advantages of simplicity and accurate prediction of DM.

In addition to providing risk stratification, the DL model, more importantly, effectively differentiates populations that could benefit from chemotherapy. Our results showed that chemotherapy with a CCND >160 mg/m^2 could improve DMFS for the high-risk T1-2N1 population. This finding lays a foundation for subsequent clinical trials to explore optimal treatment options (e.g., induction chemotherapy, concurrent chemotherapy) for high-risk patients. The current guidelines of the Chinese Society of Clinical Oncology[33] recommend a cumulative cisplatin dose of 200 mg/m^2 for patients receiving concurrent chemoradiotherapy. However, the evidence is based on post hoc analyses of phase III trials on locally advanced NPC.[34–36] The recommended dose for locally advanced NPC may be inappropriate for stage II patients. The chemotherapy dose of 160 mg/m^2 could be an important reference for personalized therapeutic strategy among patients with NPC, and it represents two cycles of chemotherapy (80 mg/m^2 per cycle) in NPC endemic areas.

Subgroup analysis showed that the prognosis prediction of DL $_{combined}$ model was evident in the T1N1 and T2N1 subgroups but was uncertain in the T2N0 subgroup. This may be because patients with T2N0 disease had superior survival outcomes (5-year DMFS: 96.2%, OS: 93.3%), and it is difficult to detect significant differences since the endpoints were small probability events. Notably, patients with T2N0 disease fail to benefit from chemotherapy compared to RT alone. For such patients, RT alone may be sufficient in the IMRT era.

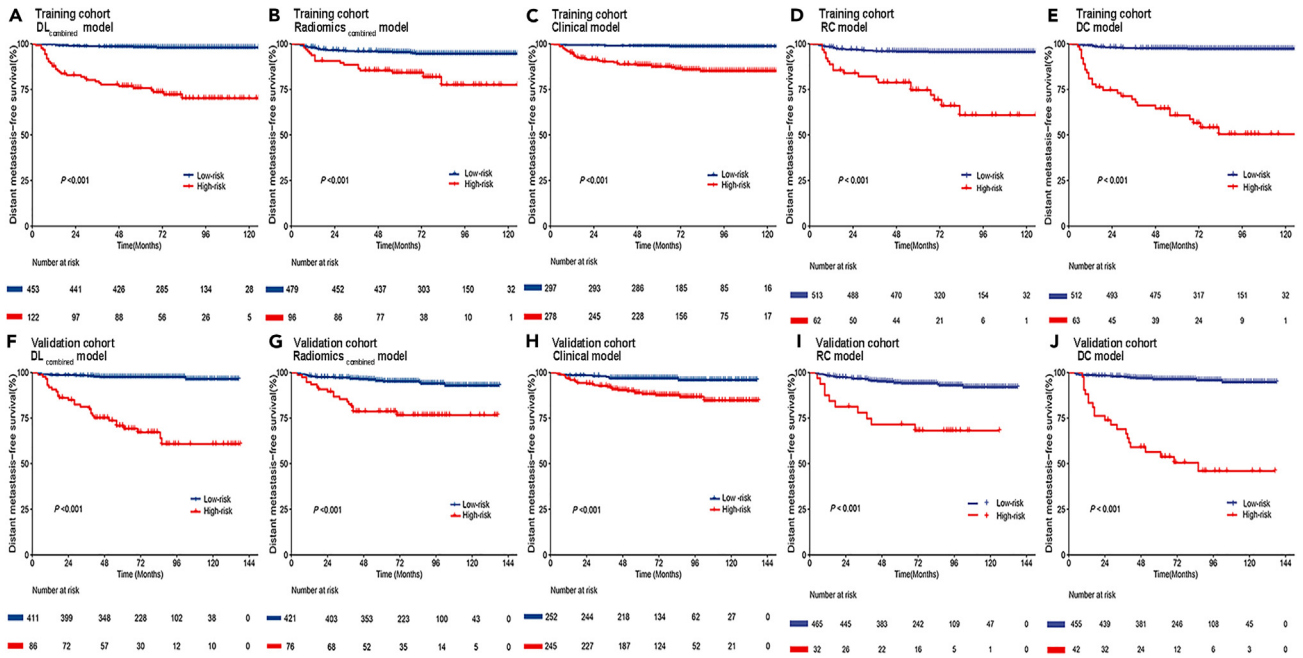


Figure 3. Kaplan–Meier curves of DMFS for models in the training and validation cohorts

(A) and (F) present the DMFS curves of the DL $_{combined}$ model in the training and validation cohorts. (B) and (G) present the DMFS curves of the radiomics $_{combined}$ model in the training and validation cohorts. (C) and (H) present the DMFS curves of the clinical model in the training and validation cohorts. (D) and (I) present the DMFS curves of the RC model in the training and validation cohorts. (E) and (J) present the DMFS curves of the DC model in the training and validation cohorts. Note that, the DL $_{combined}$ model and Radiomics $_{combined}$ model were conducted based on the three sequences (T1W, T2W, and CET1W). DC model, a model combining deep learning and clinical variables. RC model, a model combining radiomics and clinical variables.
Abbreviations: DL, deep learning; DMFS, distant metastasis-free survival.
See also Figure S1.

In conclusion, the deep learning model based on multiparametric MRI provides a strong prediction of DM for patients with T1-2N1 NPC, thus aiding in decision-making regarding individualized treatment strategies.

Limitations of the study

First, this was a retrospective study; therefore, the inherent introduction of selective bias was unavoidable. Second, there was heterogeneity between the training and validation groups in the clinical parameters of age, histopathology, EBV DNA, and LDH. Third, the DL model was constructed based on patients from an endemic area; the applicability of the DL model to patients from nonendemic areas remains unknown. Finally, there is a lack of interpretability of results predicted by the DL model. Prospective trials are warranted to investigate the clinical applicability and interpretability of the DL model for stage II NPC.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Patient cohorts
- METHOD DETAILS
  - Treatment and follow up
  - MRI information and preprocessing
  - Regions of interest segmentation
  - Construction of the DL model

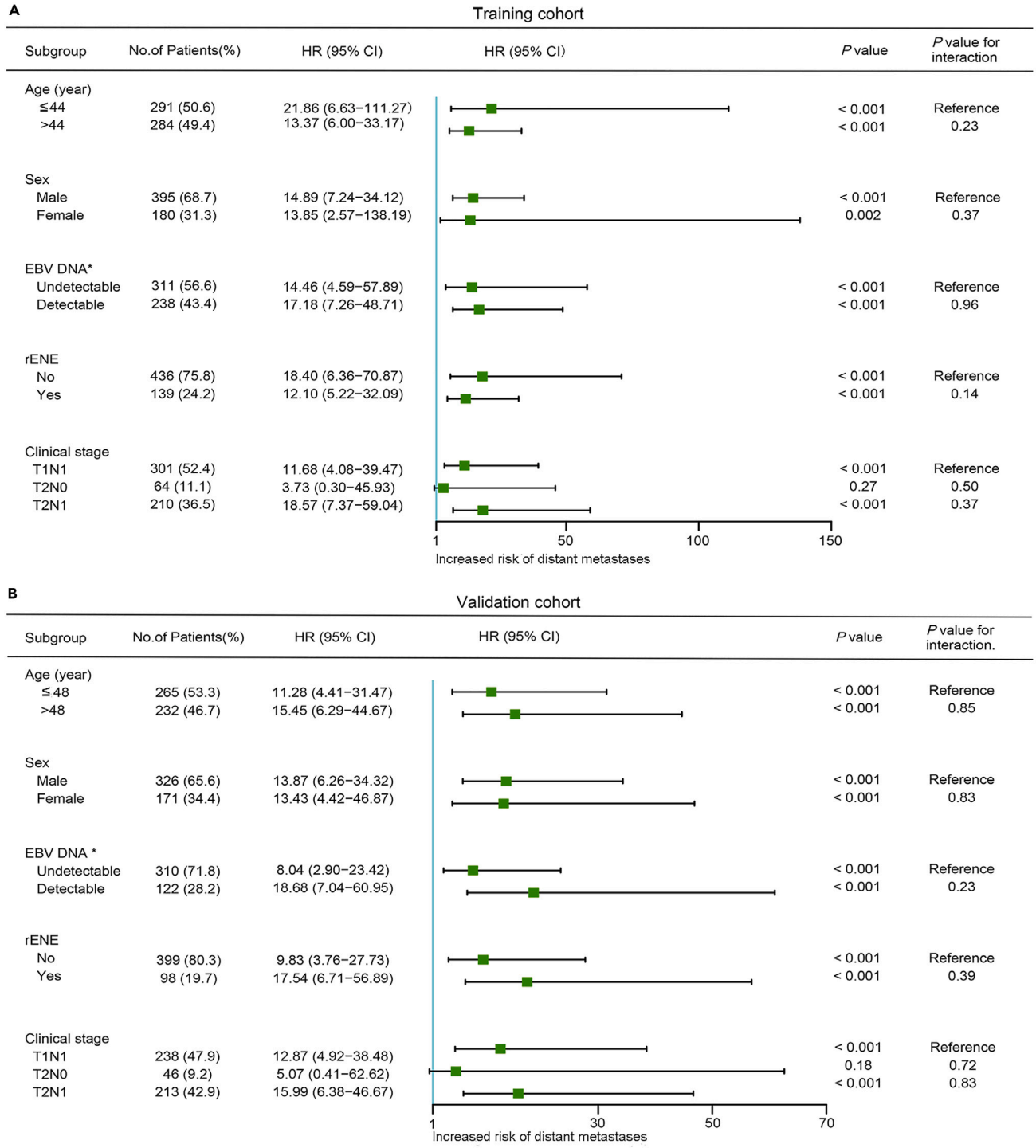


Figure 4. Predictive performance of the DL $_{combined}$ model for DMFS within subgroups

(A) Training cohort; (B) Validation cohort. *: complete case analysis. Abbreviations: DL combined model, deep learning model based on the three sequences (T1W, T2W, and CET1W); DMFS, distant metastases-free survival; HR, hazard ratio; CI, confidence interval; EBV, Epstein-Barr virus deoxyribonucleic acid; rENE, radiologic extranodal extension.

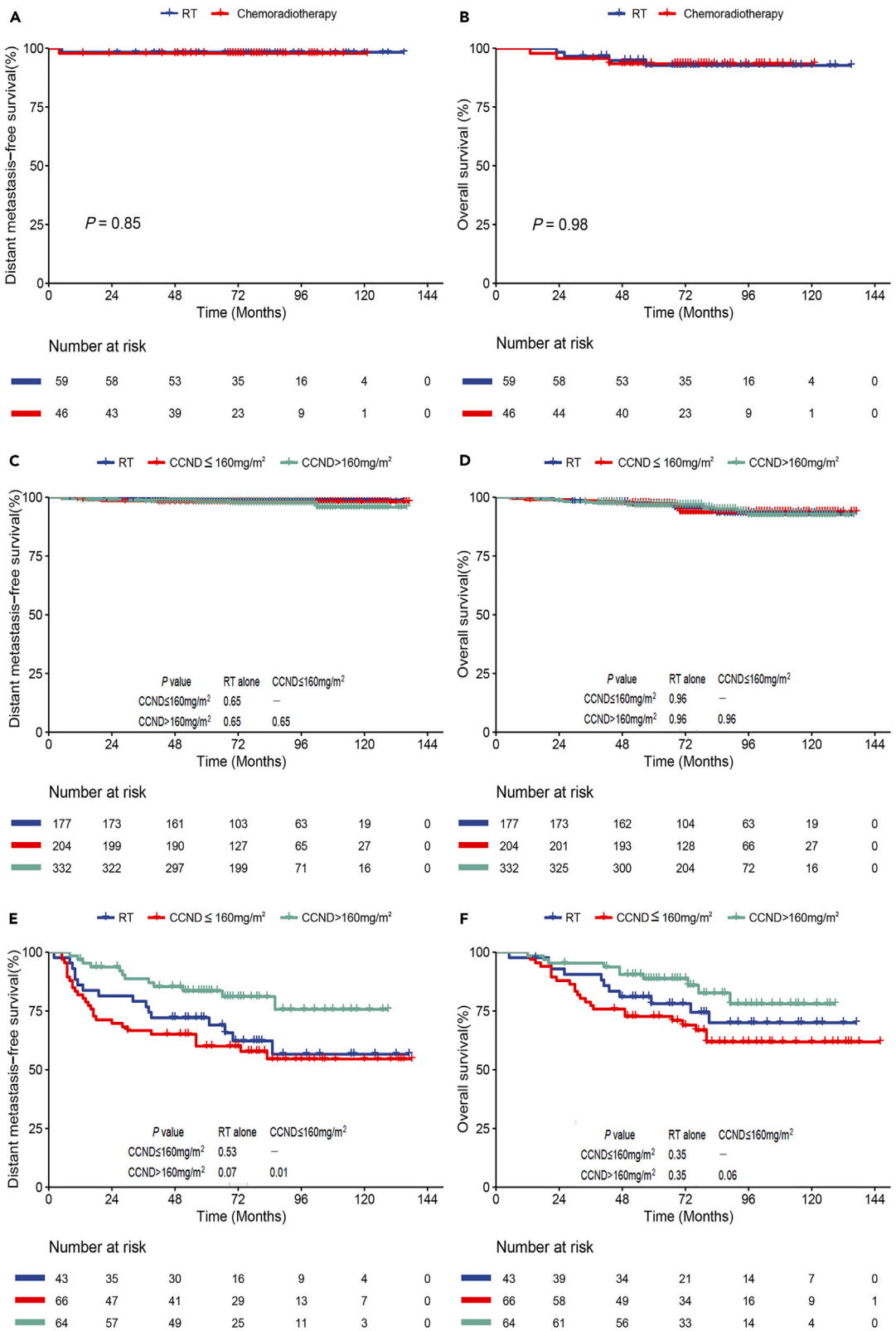

A
RT
Chemoradiotherapy
Distant metastasis−free survival(%)
P = 0.85
Time (Months)
Number at risk
59 58 53 35 16 4 0
46 43 39 23 9 1 0
B
RT
Chemoradiotherapy
Overall survival (%)
P = 0.98
Time (Months)
Number at risk
59 58 53 35 16 4 0
46 44 40 23 9 1 0
C
RT
CCND ≤ 160mg/m²
CCND>160mg/m²
Distant metastasis−free survival(%)
P value RT alone CCND≤160mg/m²
CCND≤160mg/m² 0.65 –
CCND>160mg/m² 0.65 0.65
Time (Months)
Number at risk
177 173 161 103 63 19 0
204 199 190 127 65 27 0
332 322 297 199 71 16 0
D
RT
CCND ≤ 160mg/m²
CCND>160mg/m²
Overall survival(%)
P value RT alone CCND≤160mg/m²
CCND≤160mg/m² 0.96 –
CCND>160mg/m² 0.96 0.96
Time (Months)
Number at risk
177 173 162 104 63 19 0
204 201 193 128 66 27 0
332 325 300 204 72 16 0
E
RT
CCND ≤ 160mg/m²
CCND>160mg/m²
Distant metastasis−free survival(%)
P value RT alone CCND≤160mg/m²
CCND≤160mg/m² 0.53 –
CCND>160mg/m² 0.07 0.01
Time (Months)
Number at risk
43 35 30 16 9 4 0
66 47 41 29 13 7 0
64 57 49 25 11 3 0
F
RT
CCND ≤ 160mg/m²
CCND>160mg/m²
Overall survival(%)
P value RT alone CCND≤160mg/m²
CCND≤160mg/m² 0.35 –
CCND>160mg/m² 0.35 0.06
Time (Months)
Number at risk
43 39 34 21 14 7 0
66 58 49 34 16 9 1
64 61 56 33 14 4 0

Figure 5. Kaplan–Meier analyses of chemotherapy benefit in patients with stage II NPC

(A and B) DMFS and OS curves for patients receiving RT and chemoradiotherapy in the T2N0 subgroups.

(C and D) DMFS and OS curves for low-risk patients receiving RT alone, chemotherapy with CCND $\leq$160 mg/m^2, and CCND >160 mg/m^2 based on DL $_{combined}$ model.

(E and F) DMFS and OS curves for high-risk patients receiving RT alone, chemotherapy with CCND $\leq$160 mg/m^2, and CCND >160 mg/m^2 based on DL $_{combined}$ model. Abbreviations: NPC, nasopharyngeal carcinoma; DMFS, distant metastases-free survival; OS, overall survival; RT, radiotherapy; DL combined model, deep learning model based on the three sequences (T1W, T2W and CET1W).

- Construction of clinical model
- Construction of the radiomics model
- Construction of the combined models
- Performance assessment of the models

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106932.

ACKNOWLEDGMENTS

This work was supported by Science and Technology Planning Project of Guangdong Province (grant 2021A0505110010), National Natural Science Foundation of China (grant 81872464), NHC Key Laboratory of Personalized Diagnosis and Treatment of Nasopharyngeal Carcinoma (grant 2020-PT320-004), The science and technology projects of Health Commission of Jiangxi Province (grant 202210051), The Open Fund for Scientific Research of Jiangxi Cancer Hospital (grant 2021J13), Regional Innovation System Construction cross-regional Research and Development Project of Science and Technology of Jiangxi Province (grant 20221ZDH04056), and National High Technology Research and development Program (grant 2006AA02Z4B4). We sincerely thank Peng Zhang of East China Normal University for his assistance in deep learning model construction.

AUTHOR CONTRIBUTIONS

Conceptualization, Y.F.X., J.G.L., and J.J.P.; Methodology, Y.J.H., L.Z., Y.B.C., and Y.P.X.; Software, Y.J.H., L.Z.L., and P.Z.; Formal Analysis, T.Z.L. and Q.J.G.; Investigation, L.L., Z.L.H., Y.S., and Y.L.; Writing – Original Draft, Y.J.H., L.Z., Y.P.X., and T.Z.L.; Writing – Review & Editing, S.J.L., X.C.G., Y.F.X., J.G.L., and J.J.P.; Visualization, Y.J.H., L.Z., and Y.P.X.; Supervision, Y.F.X., J.G.L., and J.J.P.; Funding Acquisition, Y.F.X. and J.G.L.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: October 6, 2022
Revised: April 11, 2023
Accepted: May 16, 2023
Published: May 19, 2023

REFERENCES

1. Pan, X.B., Li, L., Qu, S., Chen, L., Liang, S.X., and Zhu, X.D. (2020). The efficacy of chemotherapy in survival of stage II nasopharyngeal carcinoma. Oral Oncol. *101*, 104520. https://doi.org/10.1016/j.oraloncology.2019.104520.

2. Wong, K.C.W., Hui, E.P., Lo, K.W., Lam, W.K.J., Johnson, D., Li, L., Tao, Q., Chan, K.C.A., To, K.F., King, A.D., et al. (2021). Nasopharyngeal carcinoma: an evolving paradigm. Nat. Rev. Clin. Oncol. *18*, 679–695. https://doi.org/10.1038/s41571-021-00524-x.

3. Wu, P., Zhao, Y., Xiang, L., and Yang, L. (2020). Management of chemotherapy for stage II nasopharyngeal carcinoma in the intensity-modulated radiotherapy era: a review. Cancer Manag. Res. *12*, 957–963. https://doi.org/10.2147/cmar.S239729.

4. Guo, Q., Lu, T., Lin, S., Zong, J., Chen, Z., Cui, X., Zhang, Y., and Pan, J. (2016). Long-term survival of nasopharyngeal carcinoma patients with Stage II in intensity-modulated radiation therapy era. Jpn. J. Clin. Oncol. *46*, 241–247. https://doi.org/10.1093/jjco/hyv192.

5. Su, S.F., Han, F., Zhao, C., Chen, C.Y., Xiao, W.W., Li, J.X., and Lu, T.X. (2012). Long-term outcomes of early-stage nasopharyngeal carcinoma patients treated with intensity-modulated radiotherapy alone. Int. J. Radiat. Oncol. Biol. Phys. *82*, 327–333. https://doi.org/10.1016/j.ijrobp.2010.09.011.

6. National Comprehensive Cancer Network. Head and Neck Cancers. version 1.2023.

https://www.nccn.org/guidelines/guidelines-detail?category=1&id=1437.

7. Lin, J.C., Wang, W.Y., Chen, K.Y., Wei, Y.H., Liang, W.M., Jan, J.S., and Jiang, R.S. (2004). Quantification of plasma Epstein-Barr virus DNA in patients with advanced nasopharyngeal carcinoma. N. Engl. J. Med. *350*, 2461–2470. https://doi.org/10.1056/NEJMoa032260.

8. Le, Q.T., Zhang, Q., Cao, H., Cheng, A.J., Pinsky, B.A., Hong, R.L., Chang, J.T., Wang, C.W., Tsao, K.C., Lo, Y.D., et al. (2013). An international collaboration to harmonize the quantitative plasma Epstein-Barr virus DNA assay for future biomarker-guided trials in nasopharyngeal carcinoma. Clin. Cancer Res. *19*, 2208–2215. https://doi.org/10.1158/1078-0432.Ccr-12-3702.

9. Chen, Q.Y., Guo, S.Y., Tang, L.Q., Lu, T.Y., Chen, B.L., Zhong, Q.Y., Zou, M.S., Tang, Q.N., Chen, W.H., Guo, S.S., et al. (2018). Combination of tumor volume and epstein-barr virus DNA improved prognostic stratification of stage II nasopharyngeal carcinoma in the intensity modulated radiotherapy era: a large-scale cohort study. Cancer Res. Treat. *50*, 861–871. https://doi.org/10.4143/crt.2017.237.

10. Chin, O., Yu, E., O'Sullivan, B., Su, J., Tellier, A., Siu, L., Waldron, J., Kim, J., Hansen, A., Hope, A., et al. (2021). Prognostic importance of radiologic extranodal extension in nasopharyngeal carcinoma treated in a Canadian cohort. Radiother. Oncol. *165*, 94–102. https://doi.org/10.1016/j.radonc.2021.10.018.

11. Lu, T., Hu, Y., Xiao, Y., Guo, Q., Huang, S.H., O'Sullivan, B., Fang, Y., Zong, J., Chen, Y., Lin, S., et al. (2019). Prognostic value of radiologic extranodal extension and its potential role in future N classification for nasopharyngeal carcinoma. Oral Oncol. *99*, 104438. https://doi.org/10.1016/j.oraloncology.2019.09.030.

12. Mao, Y., Wang, S., Lydiatt, W., Shah, J.P., Colevas, A.D., Lee, A.W.M., O'Sullivan, B., Guo, R., Luo, W., Chen, Y., et al. (2021). Unambiguous advanced radiologic extranodal extension determined by MRI predicts worse outcomes in nasopharyngeal carcinoma: potential improvement for future editions of N category systems. Radiother. Oncol. *157*, 114–121. https://doi.org/10.1016/j.radonc.2021.01.015.

13. Guo, Q., Pan, J., Zong, J., Zheng, W., Zhang, C., Tang, L., Chen, B., Cui, X., Xiao, Y., Chen, Y., and Lin, S. (2015). Suggestions for lymph node classification of UICC/AJCC staging system: a retrospective study based on 1197 nasopharyngeal carcinoma patients treated with intensity-modulated radiation therapy. Medicine *94*, e808. https://doi.org/10.1097/md.0000000000000808.

14. Ma, H., Qiu, Y., Li, H., Xie, F., Ruan, G., Liu, L., Cui, C., and Dong, A. (2021). Prognostic value of nodal matting on MRI in nasopharyngeal carcinoma patients. J. Magn. Reson. Imag. *53*, 152–164. https://doi.org/10.1002/jmri.27339.

15. Wan, Y., Tian, L., Zhang, G., Xin, H., Li, H., Dong, A., Liang, Y., Jing, B., Zhou, J., Cui, C., et al. (2019). The value of detailed MR imaging report of primary tumor and lymph nodes on prognostic nomograms for nasopharyngeal carcinoma after intensity-modulated radiotherapy. Radiother. Oncol. *131*, 35–44. https://doi.org/10.1016/j.radonc.2018.11.001.

16. Lambin, P., Leijenaar, R.T.H., Deist, T.M., Peerlings, J., de Jong, E.E.C., van Timmeren, J., Sanduleanu, S., Larue, R.T., Even, A.J.G., Jochems, A., et al. (2017). Radiomics: the bridge between medical imaging and personalized medicine. Nat. Rev. Clin. Oncol. *14*, 749–762. https://doi.org/10.1038/nrclinonc.2017.141.

17. LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444. https://doi.org/10.1038/nature14539.

18. Dong, D., Fang, M.J., Tang, L., Shan, X.H., Gao, J.B., Giganti, F., Wang, R.P., Chen, X., Wang, X.X., Palumbo, D., et al. (2020). Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. Ann. Oncol. *31*, 912–920. https://doi.org/10.1016/j.annonc.2020.04.003.

19. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature *542*, 115–118. https://doi.org/10.1038/nature21056.

20. Tran, K.A., Kondrashova, O., Bradley, A., Williams, E.D., Pearson, J.V., and Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Med. *13*, 152. https://doi.org/10.1186/s13073-021-00968-x.

21. Peng, H., Dong, D., Fang, M.J., Li, L., Tang, L.L., Chen, L., Li, W.F., Mao, Y.P., Fan, W., Liu, L.Z., et al. (2019). Prognostic value of deep learning PET/CT-Based radiomics: potential role for future individual induction chemotherapy in advanced nasopharyngeal carcinoma. Clin. Cancer Res. *25*, 4271–4279. https://doi.org/10.1158/1078-0432.Ccr-18-3065.

22. Qiang, M., Li, C., Sun, Y., Sun, Y., Ke, L., Xie, C., Zhang, T., Zou, Y., Qiu, W., Gao, M., et al. (2021). A prognostic predictive system based on deep learning for locoregionally advanced nasopharyngeal carcinoma. J. Natl. Cancer Inst. *113*, 606–615. https://doi.org/10.1093/jnci/djaa149.

23. Zhong, L., Dong, D., Fang, X., Zhang, F., Zhang, N., Zhang, L., Fang, M., Jiang, W., Liang, S., Li, C., et al. (2021). A deep learning-based radiomic nomogram for prognosis and treatment decision in advanced nasopharyngeal carcinoma: a multicentre study. EBioMedicine *70*, 103522. https://doi.org/10.1016/j.ebiom.2021.103522.

24. Zhao, X., Liang, Y.J., Zhang, X., Wen, D.X., Fan, W., Tang, L.Q., Dong, D., Tian, J., and Mai, H.Q. (2022). Deep learning signatures reveal multiscale intratumor heterogeneity associated with biological functions and survival in recurrent nasopharyngeal carcinoma. Eur. J. Nucl. Med. Mol. Imag. *49*, 2972–2982. https://doi.org/10.1007/s00259-022-05793-x.

25. Deng, Y., Li, C., Lv, X., Xia, W., Shen, L., Jing, B., Li, B., Guo, X., Sun, Y., Xie, C., and Ke, L. (2022). 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 Progr. Biomed. *217*, 106702. https://doi.org/10.1016/j.cmpb.2022.106702.

26. Luo, X., Liao, W., He, Y., Tang, F., Wu, M., Shen, Y., Huang, H., Song, T., Li, K., Zhang, S., et al. (2023). Deep learning-based accurate delineation of primary gross tumor volume of nasopharyngeal carcinoma on heterogeneous magnetic resonance imaging: a large-scale and multi-center study. Radiother. Oncol. *180*, 109480. https://doi.org/10.1016/j.radonc.2023.109480.

27. Zhang, L., Wu, X., Liu, J., Zhang, B., Mo, X., Chen, Q., Fang, J., Wang, F., Li, M., Chen, Z., et al. (2021). MRI-based deep-learning model for distant metastasis-free survival in locoregionally advanced nasopharyngeal carcinoma. J. Magn. Reson. Imag. *53*, 167–178. https://doi.org/10.1002/jmri.27308.

28. Özgün Çiçek, A.A., Lienkamp, S.S., Brox, T., and Ronneberger, O. (2016). 3D U-net: learning dense volumetric segmentation from sparse annotation. In Medical Image Computing and Computer-Assisted Intervention, pp. 424–432.

29. Shen, H., Yin, J., Niu, R., Lian, Y., Huang, Y., Tu, C., Liu, D., Wang, X., Lan, X., Yuan, X., and Zhang, J. (2022). MRI-based radiomics to compare the survival benefit of induction chemotherapy plus concurrent chemoradiotherapy versus concurrent chemoradiotherapy plus adjuvant chemotherapy in locoregionally advanced nasopharyngeal carcinoma: a multicenter study. Radiother. Oncol. *171*, 107–113. https://doi.org/10.1016/j.radonc.2022.04.017.

30. Zhang, B., Tian, J., Dong, D., Gu, D., Dong, Y., Zhang, L., Lian, Z., Liu, J., Luo, X., Pei, S., et al. (2017). Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. Clin. Cancer Res. *23*, 4259–4269. https://doi.org/10.1158/1078-0432.Ccr-16-2910.

31. Tang, L.L., Guo, R., Zhang, N., Deng, B., Chen, L., Cheng, Z.B., Huang, J., Hu, W.H., Huang, S.H., Luo, W.J., et al. (2022). Effect of radiotherapy alone vs radiotherapy with concurrent chemoradiotherapy on survival without disease relapse in patients with low-risk nasopharyngeal carcinoma: a randomized clinical trial. JAMA *328*, 728–736. https://doi.org/10.1001/jama.2022.13997.

32. Steyerberg, E. (2009). Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. https://doi.org/10.1007/978-0-387-77244-8.

33. Chen, Y.P., Ismaila, N., Chua, M.L.K., Colevas, A.D., Haddad, R., Huang, S.H., Wee, J.T.S., Whitley, A.C., Yi, J.L., Yom, S.S., et al. (2021). Chemotherapy in combination with radiotherapy for definitive-intent treatment of stage II-iva nasopharyngeal carcinoma: CSCO and ASCO guideline. J. Clin. Oncol. *39*, 840–859. https://doi.org/10.1200/jco.20.03237.

34. Lee, A.W., Tung, S.Y., Ngan, R.K., Chappell, R., Chua, D.T., Lu, T.X., Siu, L., Tan, T., Chan, L.K., Ng, W.T., et al. (2011). Factors contributing to the efficacy of concurrent-adjuvant chemotherapy for locoregionally advanced nasopharyngeal carcinoma: combined analyses of NPC-9901 and NPC-9902 Trials. European J. Can. *47*, 656–666. https://doi.org/10.1016/j.ejca.2010.10.026.

35. Ng, W.T., Tung, S.Y., Lee, V., Ngan, R.K.C., Choi, H.C.W., Chan, L.L.K., Leung, T.W., Siu, L.L., Lu, T.X., Tan, T., et al. (2018). Concurrent-Adjuvant chemoradiation therapy for stage III-IVB nasopharyngeal carcinoma-exploration for achieving optimal 10-year therapeutic ratio. Int. J. Radiat. Oncol. Biol. Phys. *101*, 1078–1086. https://doi.org/10.1016/j.ijrobp.2018.04.069.

36. Peng, H., Chen, L., Zhang, Y., Li, W.F., Mao, Y.P., Zhang, F., Guo, R., Liu, L.Z., Lin, A.H., Sun, Y., and Ma, J. (2016). Prognostic value of the cumulative cisplatin dose during concurrent chemoradiotherapy in locoregionally advanced nasopharyngeal carcinoma: a secondary analysis of a prospective phase III clinical trial. Oncol. *21*, 1369–1376. https://doi.org/10.1634/theoncologist.2016-0105.

37. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., and Gee, J.C. (2010). N4ITK: improved N3 bias correction. IEEE Trans. Med. Imag. *29*, 1310–1320. https://doi.org/10.1109/tmi.2010.2046908.

38. Olaf Ronneberger, P.F., and Brox, T. (2015). U-net: Convolutional Networks for Biomedical Image Segmentation (Computer Science Department and BIOSS Centre for Biological Signalling Studies, University of Freiburg), pp. 234–241.

39. Le, N., Bui, T., Vo-Ho, V.K., Yamazaki, K., and Luu, K. (2021). Narrow band active contour attention model for medical segmentation. Diagnostics *11*, 1393. https://doi.org/10.3390/diagnostics11081393.

40. Zwanenburg, A., Vallières, M., Abdalah, M.A., Aerts, H.J.W.L., Andrearczyk, V., Apte, A., Ashrafinia, S., Bakas, S., Beukinga, R.J., Boellaard, R., et al. (2020). The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology *295*, 328–338. https://doi.org/10.1148/radiol.2020191145.

41. Aerts, H.J.W.L., Velazquez, E.R., Leijenaar, R.T.H., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., et al. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat. Commun. *5*, 4006. https://doi.org/10.1038/ncomms5006.

42. Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. Stat. Med. *16*, 385–395. https://doi.org/10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3.

43. Jakobsen, J.C., Gluud, C., Wetterslev, J., and Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med. Res. Methodol. *17*, 162. https://doi.org/10.1186/s12874-017-0442-1.

44. Schemper, M. (2001). The Application of Firth's Procedure to Cox and Logistic Regression.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
R (version 3·6·1)	R software	https://www.r-project.org
3D-Slicer (version 4.9.0)	3D-Slicer software	http://www.slicer.org
3D-Unet framework	Github	https://github.com/Deeplearninghhh/MRI_ML
Python (version 2.12)	Python software	https://www.python.org/downloads/release/python-378/
pyTorch (version:1.4.0)	pyTorch software	https://pytorch.org/
Other		
Research Data Deposit	Sun Yat-Sen University Cancer Center	https://www.researchdata.org.cn
Source code	Github	https://github.com/Deeplearninghhh/MRI_ML

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Yun-Fei Xia (xiayf@sysucc.org.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- De-identified patient standardized data have been deposited at the Research Data Deposit public platform (No.RDDA2022264880), and DOIs are listed in the key resources table. They are available upon request if access is granted. To request access, contact Sun Yat-Sen University Cancer Center.
- All original code has been deposited at the Github and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Patient cohorts

In this multicenter retrospective study, 1072 patients with stage II NPC treated between April 2008 and July 2017 were recruited from three hospitals in China and followed up until October 1, 2021. Figure 1 shows the inclusion criteria, exclusion criteria, and the selection process of patients from the three centers in detail. For training purposes, data were obtained from 575 patients at Sun Yat-Sen University Cancer Center (Center 1). For validation purposes, we included 285 patients from Fujian Medical University Cancer Hospital (Center 2) and 212 patients from Jiangxi Cancer Hospital (Center 3). The institutional review boards of each center approved this study and waived the requirement for informed consent owing to the retrospective nature of this study.

METHOD DETAILS

Treatment and follow up

The induction and adjuvant chemotherapy regimens were as follows: the TP regimen consisted of docetaxel 75 mg/m^2 (or paclitaxel 135–175 mg/m^2) and cisplatin (or nedaplatin 75 mg/m^2) on day 1. The GP regimen comprised cisplatin (or nedaplatin 80 mg/m^2) on day 1 and gemcitabine 1000 mg/m^2 on days 1 and 8, respectively. The PF regimen comprised cisplatin (or nedaplatin 80–100 mg/m^2) on day 1 and 5-Fu 800–1000 mg/m^2/d, 120 h continuous intravenous (CIV). Finally, the TPF regimen consisted of

docetaxel 60-75 mg/m^2 (or paclitaxel 135−175 mg/m^2) on and cisplatin (or nedaplatin 60−75 mg/m^2) on day 1, and 5-Fu 500−750 mg/m^2/d, 120h CIV. A certain proportion of the patients were treated with oxaliplatin (85−130 mg/m^2), lobaplatin (30 mg/m^2), or carboplatin (AUC=5) instead of cisplatin/nedaplatin. The regimen was repeated every 3 weeks for two to four cycles.

Concurrent chemotherapy consisted of 30−40 mg/m^2 cisplatin/nedaplatin administered every week, 80−100 mg/m^2 cisplatin/nedaplatin administered every 3 weeks, 15−30mg/m^2 docetaxel administered every week, 50 mg/m^2 paclitaxel administered every week, 85−130mg/m^2 oxaliplatin administered every 3 weeks, carboplatin (AUC=5) administered every 3 weeks, or 30 mg/m^2 lobaplatin administered every 3 weeks. In addition, a subset of patients was administered a two-drug regimen of TP and PF (dose as previously described) every 3 weeks. More information about chemotherapy was described in Table S1.

All patients were treated with IMRT, including daily radiation therapy five times a week for 6−7 weeks. A cumulative dose of 66−72 Gy/28−33 fractions to the planning target volume (PTV) of the primary gross tumor volume and 64−70 Gy/28−33 fractions to the PTV of the involved lymph nodes. The prescribed doses were 6066 Gy/28−33 fractions to the PTV of high-risk clinical target volume and 54−56 Gy/28−33 fractions to the PTV of low-risk clinical target volume.

Patients were followed up every 3 months for the first 2 years, every 6 months for the next 3 years, and annually thereafter. DM was diagnosed using various examinations, which included chest computed tomography (CT), abdominal ultrasound/CT/MRI, bone scan or positron emission tomography-CT. For some suspect lesions, a pathological examination was conducted. The primary clinical outcome of this study was DMFS. The second endpoint was OS.

MRI information and preprocessing

Axial T1-weighted (T1W), T2-weighted (T2W), and contrast-enhanced T1-weighted (CET1W) images of pretreatment head and neck MRI were acquired. The MRI scanners and parameters are described in Tables S2 and S3. Preprocessing of the MR images first included normalization to correct the scanner-related variations. Then, a bias field correction was applied using the N4ITK algorithm to correct the potential effects on the image from magnetic field inhomogeneities.[37] Image preprocessing was performed in Python using the open-source Pyradiomics package (version 2.12; https://pyradiomics.readthedocs.io/en/2.1.2/). Codes of image preprocessing are available at GitHub (https://github.com/Deeplearninghhh/MRI_ML).

Regions of interest segmentation

For training the segmentation of the 3D-Unet framework as well as for extracting radiomics features, the primary tumor and metastatic lymph node segmentation were contoured on each slice of the three axial MR sequences by an oncologist (Y.J.H.) using 3D-Slicer software (version 4.9.0; Open source: http://www.slicer.org). The outline results were reviewed by a senior head and neck radiologist with 30 years of subspecialist experience (L.Z.L.).

Construction of the DL model

To increase the diversity of the training samples for the segmentation and to prevent over-fitting, data augmentation methods were employed in the training cohort, including rotation and blurring. After data augmentation, 575 × 32×6 (110,400 in total) MR images were obtained for the segmentation.

U-net can extract global contextual information from an original image while preserving spatially continuous detail in the target image.[38] We adopted a three-dimensional (3D)-Unet as the backbone network because the input is an axial sequence of 3D MRI images consisting of axial slices of each patient.[28] The network architecture is illustrated in Figure S2. The design of the 3D-Unet in this study integrates the functions of automatic tumor and metastatic lymph node segmentation with DM risk prediction. The 3D-Unet network was first trained for automatic segmentation and then trained for the prediction of DM after the segmentation achieved stable performance.

Three axial MRI slices were sampled as 192 × 192 by using linear interpolation as the input to the network to apply this model to the MRI of three centers. The encoder embeds the original high-dimensional data into the low-dimensional space, called " the bottleneck layer". Then, the decoder converts the values of the

bottleneck layer back into the original high-dimensional space to reconstruct the segmented images with detailed information.

The encoder consists of the repeated application of two 3 × 3 × 3 convolutions, each followed by a rectified linear unit (ReLU) and a 2 × 2 × 2 max pooling operation with strides 2 for downsampling.[17] The output of the convolution layer was transferred to the decoder before the pooling operation of the encoder. The number of filters in the convolutional layer doubled after each downsampling. However, the decoder first upsampled the feature map using a 2 × 2 × 2 convolution operation, thereby reducing the feature channels by one-half. This was followed by two 3 × 3 × 3 convolutions, each with ReLU. Finally, the segmentation map was generated by a 1 × 1 × 1 convolution operation. In this process, the loss function is a Dice loss,[39] and it is defined as follows:

$$L_{Dice} = 1 - 2\frac{\sum_i^N T_i P_i}{\sum_i^N T_i + P_i} = 2\frac{T \bigcap P}{T \bigcup P} \quad \text{(Equation 1)}$$

The loss function is a measure of the average divergence between the output of the network (P) and the actual function (T) being approximated over the entire input domain (sized, m × n). The variable *i* denotes the index of each pixel in the image spatial space: N = m × n.

On this basis, we added a distant metastasis prediction model. The DM prediction model takes the results of the global averaging pool as the input, followed by a pool layer, a fully connected layer with ReLU activation, and a classification layer with softmax activation. We used the same network structure and training strategy to train the MR images of three axial MR sequences separately and to output the predicted probability value of DM for each sequence. For the output of the combined sequence, the results of the global average pool of the three MR sequences, three tensors of 6144*1*1, were merged to obtain a combined tensor of the combined sequence for subsequent prediction. The loss function used in this part was a cross-entropy loss:[39]

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}[T_i \ln(P_i) + (1 - T_i)\ln(1 - P_i)] \quad \text{(Equation 2)}$$

The batch size was set at 64, and the learning rate was set as 1E-4. We applied the adam optimizer in the torch library and then set the epoch to 500 for the iterations. Deep learning models were trained using pyTorch (version:1.4.0; Open source: https://pytorch.org/). All experiments were conducted on a computing cluster: two NVIDIA Tesla V100 with a 32 TB frame buffer (NVIDIA Corporation, Santa Clara, CA, USA). The codes for the construction of the DL model were uploaded onto a public platform (available at: https://github.com/Deeplearninghhh/MRI_ML).

Construction of clinical model

Before constructing the clinical model, an rENE assessment was conducted. The criteria for unequivocal rENE were G1 (coalescent nodal mass comprising ≥2 adjacent nodes) and G2 (invasion beyond perinodal fat to frankly infiltrate adjacent structures, e.g., muscles, nerves, and parotid glands). Two radiologists evaluated the MRI scans independently (L.Z.L. and Y.P.X.). Evaluation of rENE relied on contrast-enhanced T1-weighted as well as other MR sequences. Inconsistent assessment results were resolved by consensus. Examples of rENE images have been presented in Figure S3. Clinical variables, including age, sex, T category, N category, LDH, EBV DNA, rENE, and chemotherapy, were included in univariate and multivariable analyses of DM in the training cohort. Then, clinical variables with statistically significant differences in univariate and multivariable analyses were incorporated in Cox proportional hazards regression analysis to construct the clinical model. The output of each model is a risk score for each patient, representing the patient's risk of DM.

Construction of the radiomics model

The radiomics model was constructed according to the reported guidelines.[40] First, images were resampled to a 3 × 3 × 3 mm voxel size to standardize the voxel spacing.[16] The MRI intensity values were also discretized using a fixed 25Hbin width to reduce image noise.[41] Then, wavelet filtering and Gaussian filtering were performed per image to extract radiomic features with different frequency domains. Quantitative radiomic features were extracted from manually segmented regions of interest for each sequence separately using the Pyradiomics package of Python (version 2.12; https://pyradiomics.readthedocs.

io/en/2.1.2/). Codes of image preprocessing and feature extraction are available at GitHub (available at: https://github.com/Deeplearninghhh/MRI_ML).

All the extracted features were standardized using the *Z* score method to ensure comparable ranges for the feature values. Features are consistent with the definitions of features described by the Imaging Biomarker Standardization Initiative.[40] Next, we conducted a radiomics feature selection in the training cohort. First, p-values between each feature and DMFS were calculated by univariate analysis. Pearson correlation coefficients (r) were calculated between each pair of features to analyze linear correlations. Only the most important prognostic features were retained in the features pairs with $r > 0.85$ (lower p-values in the univariate analysis). Second, the least absolute shrinkage and selection operator (LASSO) Cox regression method was used to filter out essential features. LASSO regression performs L1 regularization, where some coefficients can become zero and be eliminated from the model.[42] The feature selection steps were performed separately for the features of the T1W, T2W, and CET1W sequence as well as for all features. Therefore, radiomics model of each sequence and combined sequences was conducted to predict DM based on a linear combination of selected features, weighted by their respective coefficients.

Construction of the combined models

The output score of the DL $_{combined}$ model (DL model based on three MR sequences) and the radiomics $_{combined}$ model (Radiomics model based on three MR sequences) was combined with selected clinical features, respectively, using Cox proportional hazards regression analysis to construct an integrated deep learning and clinical (DC) model and a combined radiomics and clinical (RC) model in the training cohort.

Performance assessment of the models

The performance of the 3D-Unet framework in segmentation was evaluated using the Dice similarity coefficient. The prediction performance of each model for DMFS was first evaluated in the training cohort and then verified in the external validation cohort. We then plotted Kaplan–Meier curves to demonstrate the association between the models and prognosis. In addition, Harrell's concordance index (C-index) and time-dependent receiver operating characteristic (TD-ROC) analysis were conducted to evaluate the predictive ability of the models.

QUANTIFICATION AND STATISTICAL ANALYSIS

EBV DNA and lactate dehydrogenase were missing data at complete random; therefore, they were processed using a complete case analysis.[43] Clinical variables were compared between the training and validation cohorts using Pearson's chi-square or Fisher's exact tests for categorical variables and the Mann–Whitney U test for continuous variables. Survival curves were obtained using the Kaplan–Meier method and were compared using the log-rank test. Univariate and multivariate analyses were conducted using the Cox proportional hazard model. We used the 'compareC' and 'timeROC' packages in R software to calculate the *P*-values that compared the C-index and area under curve (AUC) values between models. To determine the optimal cut-off values for each model, we employed the 'surv_cutpoint' function from the 'survminer' R package. This function calculates the cut-off values that yield the lowest log-rank statistic for DMFS. We used 'QualInt' R package to calculate the *P* value of the interaction test in subgroups analysis. To address the problem that subgroup analyses with a small number of events may have monotonic likelihoods when fitting Cox models, resulting in large confidence interval (CI), we applied Firth's penalized partial likelihood correction to the Cox regression model.[44] Statistical analysis was conducted using R software (version 3·6·1; http://www.R-project.org; R Foundation for Statistical Computing, Vienna, Austria). A two-sided *P* value <0.05 indicated a statistically significant difference.