Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Multimodal Deep Learning Fusion Model Based on Radiomics and Dose Features for Predicting Nasopharyngeal Necrosis After Re-irradiation in Recurrent Nasopharyngeal Carcinoma

Meining Chen
Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center

Xuan Guo
Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center

Jiaxin Lin
Manteia Technologies Co, Ltd

Xiao Liu
Manteia Technologies Co, Ltd

Zhenyu Qi
Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center

Qichao Zhou
Manteia Technologies Co, Ltd

Zhirui Huang
Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center

Chong Zhao
Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center

Xiaowu Deng
Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center

Xingwang Gao
Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center

Jingjing Miao
Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center

Yimei Liu
Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center
Jun Zhang
Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center
Yinglin Peng
pengyl@sysucc.org.cn
Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center

---

Research Article

Keywords: deep learning, dose distribution, recurrent nasopharyngeal carcinoma, nasopharyngeal necrosis, radiomics

Posted Date: June 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9711781/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Multimodal Deep Learning Fusion Model Based on Radiomics and Dose Features for Predicting Nasopharyngeal Necrosis After Re-irradiation in Recurrent Nasopharyngeal Carcinoma

Meining Chen[1,†], Xuan Guo[1,†], Jiaxin Lin[2,†], Xiao Liu[2], Zhenyu Qi[1], Qichao Zhou[2], Zhirui Huang[1], Chong Zhao[1], Xiaowu Deng[1], Xingwang Gao[1], Jingjing Miao[1], Yimei Liu[1,*], Jun Zhang[1,*], Yinglin Peng[1,*]

[1] Department of Radiation Oncology, State 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, P. R. China

[2] Manteia Technologies Co, Ltd, Xiamen, 361008, China

† Meining Chen, Xuan Guo and Jiaxin Lin contributed equally to this work.

*** Correspondence**: Yinglin Peng, Department of Radiation Oncology, Sun Yat-sen University Cancer Center, Guangzhou 510060, China. Email: pengyl@sysucc.org.cn; Jun Zhang. Email: zhangjun@sysucc.org.cn; Yimei Liu. Email: liuym1@sysucc.org.cn. Tel/Fax: [+86 02087343036/+86 02087343044].

Abstract

Objective: To develop and evaluate a multimodal deep learning fusion model integrating planning CT, multi-sequence MRI, and three-dimensional dose distribution for predicting nasopharyngeal necrosis after re-irradiation in patients with recurrent nasopharyngeal carcinoma.

Methods: This retrospective study included 113 patients with recurrent nasopharyngeal carcinoma who underwent re-irradiation between February 2008 and April 2017. All patients had complete planning CT, multi-sequence MRI, and three-dimensional dose distribution data. The cohort was randomly divided into a training set (n = 90) and an independent test set (n = 23). Radiomic and dose-related features were extracted and selected for each modality. Deep learning models based on modality-specific encoders and decision-level weighted fusion were constructed to compare the predictive performance of single-modality, dual-modality, and trimodality models. Model performance was assessed using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, calibration curves, and decision curve analysis.

Results: Among the 113 patients, 39 developed nasopharyngeal necrosis, corresponding to an incidence of 34.5%. In the independent test set, the single-modality MRI, Dose, and CT models achieved AUCs of 0.800, 0.800, and 0.775, respectively. Dual-modality fusion models showed numerically higher predictive performance, with the CT+Dose model achieving the most robust performance among dual-modality models (AUC=0.850). The trimodality MRI+CT+Dose model achieved the highest AUC of 0.875. Calibration and decision curve analyses indicated favorable calibration and potential clinical utility of the multimodal fusion models.

Conclusion: The proposed multimodal deep learning fusion model showed favorable performance for predicting nasopharyngeal necrosis after re-irradiation in recurrent nasopharyngeal carcinoma. Multimodal fusion provided better prediction than single-modality modeling, with the trimodality model achieving the highest discriminative ability and the CT+Dose model demonstrating favorable robustness and clinical accessibility.

Keywords: deep learning; dose distribution; recurrent nasopharyngeal carcinoma; nasopharyngeal necrosis; radiomics

1. Introduction

Re-irradiation is an important salvage treatment for locally recurrent nasopharyngeal carcinoma (rNPC) and may provide local control and survival benefits for selected patients [1-5]. However, because the tolerance of the nasopharynx and surrounding skull base tissues is substantially reduced after prior radiotherapy, the therapeutic window of re-irradiation is narrow and the risk of severe late radiation-induced injury is markedly increased [1,3,6]. Nasopharyngeal necrosis (NN) is one of the most serious and potentially fatal complications after re-irradiation for rNPC. It may lead to persistent headache, infection, skull base osteoradionecrosis, and massive nasopharyngeal hemorrhage, thereby severely compromising survival and quality of life [6,7]. Therefore, accurate pretreatment identification of patients at high risk of NN and subsequent optimization of re-irradiation strategies and individualized interventions are clinically important for improving outcomes in rNPC.

Previous studies have attempted to develop risk prediction models for NN or lethal nasopharyngeal necrosis based on clinical factors. Li et al. [8] established a nomogram for predicting post-radiation necrosis in patients with nasopharyngeal carcinoma, and Yu et al. [9] further developed a model to predict lethal nasopharyngeal necrosis after intensity-modulated re-irradiation, suggesting that clinical variables have value for risk stratification. However, the development of NN is influenced not only by clinical factors, but also by the extent of recurrent tumor invasion, structural damage to adjacent tissues, post-radiotherapy microenvironmental alterations, and the spatial distribution of the re-irradiation dose. Conventional clinical models may not adequately characterize the spatial heterogeneity of recurrent tumors and surrounding normal tissues, which limits their predictive performance and individualized assessment capacity.

In recent years, radiomics, dosiomics, and deep learning approaches have provided new opportunities for predicting radiotherapy-related toxicity. MRI offers high soft-tissue contrast and can delineate recurrent tumor invasion, pre-necrotic tissue changes, and local microenvironmental heterogeneity. A multicenter study showed that multiparametric MRI-based radiomic features could predict the risk of NN after re-irradiation for rNPC and outperformed clinical models alone [10]. In addition, dose distribution directly reflects the spatial irradiation pattern of the target volume and adjacent normal tissues and is therefore a key determinant of radiation-induced injury. Previous

studies have also suggested that combining multi-sequence MRI with dose information may improve prediction of NN after re-irradiation for rNPC [11,12].

In addition to MRI and dose information, planning CT also has important value. CT is the fundamental imaging modality for re-irradiation treatment planning, target delineation, and dose calculation, and it has the advantages of easy acquisition, high workflow standardization, and close integration with routine radiotherapy practice. A recent study showed that a multimodal model integrating planning CT, dose, and clinical information could effectively predict NN after re-irradiation for rNPC, suggesting that CT-derived tissue density, anatomical boundaries, and spatial structural information can provide complementary information for risk assessment [13]. Thus, CT, multi-sequence MRI, and three-dimensional dose distribution are complementary in terms of information representation, and their combined modeling may further improve the accuracy and robustness of NN risk prediction.

Multimodal deep learning has become an important research direction in medical artificial intelligence and precision oncology [14-18]. For heterogeneous modalities such as CT, MRI, and dose distribution, which differ in data scale, distribution, and noise structure, the choice of fusion strategy may directly affect model performance. Compared with simple feature concatenation, decision-level fusion with modality-specific encoders can learn modality-specific information separately and integrate it adaptively at the output level, thereby reducing cross-modality interference and improving model stability [14-17]. However, current studies on NN risk prediction after re-irradiation for rNPC have mainly focused on MRI combined with dose information, or CT combined with dose and clinical variables. Studies that uniformly integrate CT, multi-sequence MRI, and three-dimensional dose distribution and systematically compare different modality combinations remain limited [11-13].

Therefore, this study aimed to integrate planning CT, multi-sequence MRI, and three-dimensional dose distribution information and to construct a multimodal deep learning fusion model based on modality-specific encoding and decision-level fusion for predicting NN risk after re-irradiation in patients with rNPC. By systematically comparing the predictive performance of single-modality, dual-modality, and trimodality models, this study evaluated the complementary value of different modalities and provides a potential basis for individualized risk stratification and

treatment decision-making before re-irradiation in rNPC.

2. Materials and Methods

2.1 Patient cohort

This retrospective study included patients with locally recurrent nasopharyngeal carcinoma who underwent re-irradiation at Sun Yat-sen University Cancer Center between February 2008 and April 2017. The inclusion criteria were as follows: (1) no nasopharyngeal necrosis before recurrence; (2) availability of planning CT images for radiotherapy; (3) availability of pre-reirradiation MRI, including T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), and contrast-enhanced T1-weighted imaging (T1WI-C), without artifacts, blurring, slice discontinuity, or disordered slices; (4) receipt of re-irradiation; and (5) no history of other malignancies. The exclusion criteria were as follows: (1) missing clinical data; (2) image artifacts, blurring, slice discontinuity, or disordered slices; and (3) missing imaging or dose modalities. A total of 169 patients with locally recurrent nasopharyngeal carcinoma were screened. After excluding patients with incomplete imaging data (n = 54) and missing definite outcome labels (n = 2), 113 patients with complete MRI, CT, and dose information were included in the final analysis and divided into a training set (n = 90) and an internal test set (n = 23). This retrospective study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of Sun Yat-sen University Cancer Center (No. B2023-103-01). The requirement for written informed consent was waived by the ethics committee. The patient selection process is shown in **Figure 1**.

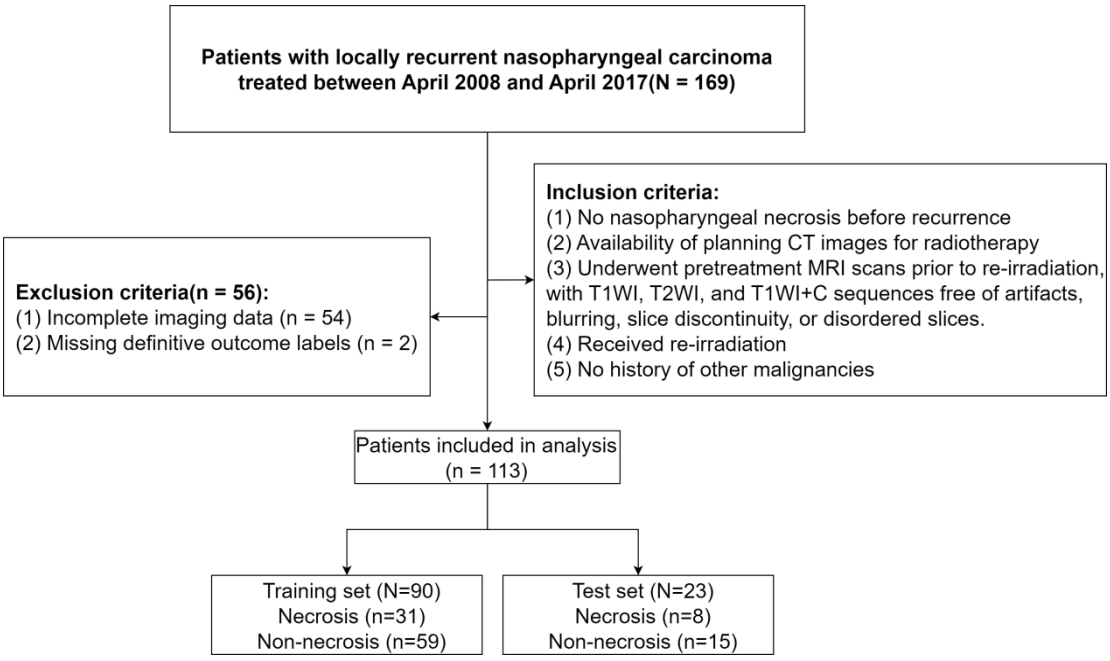

Figure 1. Flowchart of patient selection.

2.2 Treatment and Diagnosis of Nasopharyngeal Necrosis

Curative salvage IMRT was administered to all patients according to established institutional guidelines [19]. The prescribed dose to the GTVnx was 50 – 70 Gy, delivered in 1.80 – 2.50 Gy fractions once daily, 5 days weekly, with a mean EQD2 of 64 Gy [20]. Most patients also received platinum-based chemotherapy as induction, concurrent, or adjuvant treatment. Induction and adjuvant regimens included 2 – 3 cycles of PF (cisplatin with 5-fluorouracil), TPF (adding paclitaxel to PF), or TP (cisplatin with paclitaxel) every 3 weeks, while concurrent cisplatin was given weekly or every 3 weeks during re-irradiation.

NN was identified through a combination of clinical symptoms, imaging, nasopharyngoscopy, and pathological findings. Typical manifestations included fetid nasal odor, persistent headache, necrotic tissue, skull base osteoradionecrosis with mucosal interruption under endoscopy, MRI-detected tissue defects, and abundant acellular eosinophilic material on hematoxylin-eosin staining [8]. Patients who fulfilled the diagnostic criteria for NN and ultimately died of intractable epistaxis were defined as having lethal nasopharyngeal necrosis.

2.3 Multimodal image acquisition and delineation

CT image acquisition and delineation: Patients were positioned supine with the head immobilized using a customized headrest and a thermoplastic mask covering the head and upper neck. CT images were acquired with a slice thickness of 3 mm, covering the region from the superior border of the frontal sinus to 2 cm below the clavicular heads. Detailed acquisition parameters are provided in **Supplementary Table 1**. The CT images were transferred to the Monaco treatment planning system (version 6.00, Elekta, Sweden). According to ICRU Reports 50, 62, and 83 [21-23], the recurrent gross tumor volume (GTV), clinical target volume (CTV), nodal gross tumor volume (GTVnd), and organs at risk (OARs) were delineated slice by slice on CT images. The planning target volume of the recurrent nasopharyngeal lesion (PTVnx) was automatically generated by expanding the GTV by 3 mm and was used as the region of interest (ROI) for subsequent analysis.

MRI acquisition and delineation: All patients underwent multiparametric MRI within 2 weeks before retreatment. A combined head-and-neck coil was used, and the scan range extended from the sellar region to the inferior border of the sternoclavicular joint. The acquired sequences included T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), and contrast-enhanced T1-weighted imaging (T1WI-C). Detailed MRI acquisition parameters are provided in **Supplementary Table 2**. All MRI sequences and dose distributions were rigidly registered to the planning CT coordinate system using MIM software (MIM Vista V7.1.3, MIM Software Inc., USA) before feature extraction. The same PTVnx mask was resampled to each modality to ensure spatial consistency across CT, MRI, and dose maps.

2.4 Image processing and feature extraction

To reduce potential confounding effects, CT, MRI, and dose maps were standardized and resampled before feature extraction. First, all CT images were adjusted to a window width of 300 HU and a window level of 50 HU to enhance tumor-region contrast. N4 bias-field correction was applied to all MRI images [24]. Subsequently, image intensities were scaled to the range of [0, 1] to reduce interscanner and intersequence intensity variations. Finally, all images were resampled to an isotropic resolution of 1 x 1 x 1 mm^3 using B-spline interpolation (sitkBSpline) to ensure spatial

consistency for subsequent feature extraction.

Using PyRadiomics (version 3.1.0) in a Python 3.8.2 environment, a total of 1,502 radiomic features were extracted from each PTVnx. These features included shape-based features, first-order histogram features, texture features based on gray-level co-occurrence, run length, size zone, dependence, and neighboring gray-tone difference matrices, as well as higher-order features generated using Laplacian of Gaussian filtering, wavelet transformation, and gradient filtering. Detailed feature extraction parameters are provided in **Supplementary Table 3**.

For dose-feature extraction, a multi-view deep learning approach was used. First, the three-dimensional dose distribution was sliced along three orthogonal planes: axial, coronal, and sagittal. These slices were used to generate multi-view two-dimensional dose images. Each set of planar images was then fed into a pretrained ResNet-18 convolutional neural network [25] to extract deep feature representations. Global average pooling (GAP) was applied at the network output to reduce dimensionality while preserving the main feature information. Finally, the feature vectors from the three planes were fused to obtain a comprehensive multi-view feature vector with a dimension of 1 x 1536, which was used for subsequent analysis and modeling. Details are provided in **Supplementary A: Dose Feature Extractor**.

2.5 Feature selection

All numerical features were normalized using MinMaxScaler, and missing values were imputed using the mean value. To reduce feature redundancy and retain informative predictors, a two-stage feature selection strategy was adopted. In the initial screening step, variance thresholding and mutual information were evaluated on the training set for each modality. In the refined selection step, XGBoost-based recursive feature elimination (RFE) was applied to iteratively remove features with low importance. Finally, the top 30 features were retained for each modality. All feature selection procedures were performed only in the training set to avoid data leakage. The feature selection parameters for each modality are provided in **Supplementary Table 4.**

2.6 Model construction

The proposed multimodal deep learning fusion model was designed using a hierarchical architecture consisting of modality-specific encoders, residual feature extraction modules, and a decision-level fusion mechanism with learnable global weights. This design enables the integration of heterogeneous information from different modalities, including MRI, CT, and dose distribution, while preserving modality-specific semantic representations and interpretability of the decision process. The model architecture is shown in **Figure 2**.

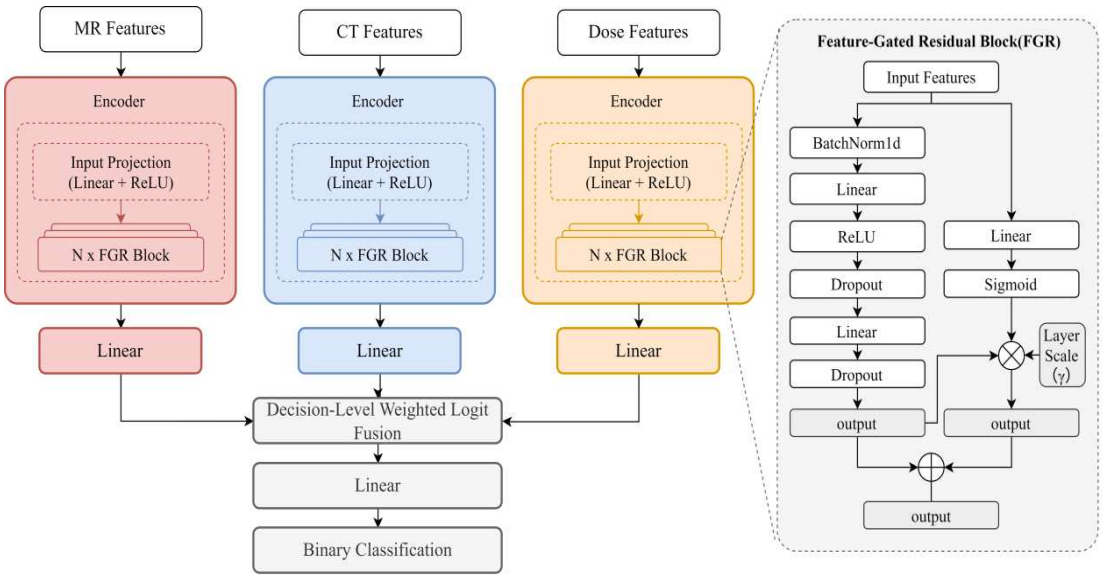

Figure 2. Architecture of the trimodality deep learning fusion network.

The overall framework consisted of three modality-specific branches for MRI, CT, and dose features. In each branch, a linear projection layer with ReLU activation first mapped the original input features into a shared latent representation space. Multiple stacked feature-gated residual blocks were then used for hierarchical feature learning. Each residual block consisted of a feed-forward transformation layer and a gated residual connection. The gating unit used a sigmoid-activated linear layer to adaptively regulate the residual signal, and a learnable LayerScale parameter was introduced to scale the residual amplitude and improve training stability in deeper networks. After modality-specific encoding, each branch generated modality-specific logits through an independent linear prediction head.

(1) Modality-specific encoder

To address distributional shifts across modalities, a separate encoder was constructed for each

modality.

Let m denote the modality index, $m \in MR, CT, Dose$, the input feature vector of each modality $\mathrm{x}^{(\mathrm{m})} \in \mathbb{R}^{\mathrm{d_m}}$, where d_m is the original feature dimension of that modality. First, it is mapped to a unified latent space dimension D through a linear projection layer.:

$$\mathrm{H}_0^{(m)} = ReLU(\mathrm{W}_{proj}^{(m)}\mathrm{x}^{(m)} + \mathrm{b}_{proj}^{(m)}) \quad (1)$$

where $W_{\mathrm{proj}}^{(m)} \in \mathbb{R}^{d_m \times D}$and $b_{\mathrm{proj}}^{(m)} \in \mathbb{R}^{D}$denote the weight matrix and bias vector of the projection layer for the m-th modality, respectively, and $\mathrm{ReLU}(\cdot)$represents the nonlinear activation function. The projected initial feature representation $H_0^{(m)}$is then fed into the encoder E_m, which consists of Lstacked Feature-Gated Residual Blocks, to extract deep feature representations through multilayer nonlinear transformations, formulated as:

$$\mathrm{H}_{\ell}^{(m)} = Feature - Gated\ Residual\ Block(\mathrm{H}_{\ell-1}^{(m)}), \quad \ell = 1,2,\dots,L \quad (2)$$

After Llayers of processing, the deep feature representation for each modality is obtained as $Z^{(m)} = H_L^{(m)} \in \mathbb{R}^{B \times D}$, where Bdenotes the training batch size. An independent classification head is attached to the end of each encoder to map the extracted features into the class prediction space:

$$logit^{(m)} = \mathrm{W}_h^{(m)}\mathrm{Z}^{(m)} + \mathrm{b}_h^{(m)}, \quad logit^{(m)} \in \mathbb{R}^{B \times 2} \quad (3)$$

where $W_h^{(m)} \in \mathbb{R}^{2 \times D}$and $b_h^{(m)} \in \mathbb{R}^{2}$denote the weight matrix and bias vector of the prediction head for the m-th modality, respectively. $\mathrm{logit}^{(m)}$represents the modality-specific logits for the two outcome classes, namely necrosis and non-necrosis.

(2) Feature-Gated Residual Block

To further enhance feature representation and improve training stability, Feature Gating and LayerScale mechanisms were introduced into each residual block. Given an input feature $h \in \mathbb{R}^{B \times D}$, the nonlinear transformation along the residual path is defined as:

$$F(\mathrm{h}) = Dropout(\mathrm{W}_2 \cdot Dropout(ReLU(BN(\mathrm{h})\mathrm{W}_1 + \mathrm{b}_1)) + \mathrm{b}_2) \quad (4)$$

where $\mathrm{BN}(\cdot)$denotes batch normalization; $W_1, W_2 \in \mathbb{R}^{D \times D}$are the weight matrices of the linear layers in the residual path, and $b_1, b_2 \in \mathbb{R}^{D}$are the corresponding bias vectors.

The gating vector $g(h)$is generated through a sigmoid activation function to enable sample-

adaptive feature weighting:

$$g(\mathrm{h}) = \sigma(\mathrm{W}_g \mathrm{h} + \mathrm{b}_g) \tag{5}$$

where $W_g \in \mathbb{R}^{D\times D}$ and $b_g \in \mathbb{R}^D$ are learnable parameters of the gating branch, and $\sigma(\cdot)$ denotes the sigmoid activation function, which maps the gating coefficients to the interval $[0, 1]$, thereby enabling dynamic selection of the contribution of residual features.

The LayerScale mechanism further stabilizes gradient propagation in the deep multimodal network by introducing a learnable scaling parameter $\gamma \in \mathbb{R}^D$, initialized to 10^{-2}. The final output of the residual block is defined as:

$$\mathrm{h}_{out} = \mathrm{h} + \gamma \odot g(\mathrm{h}) \odot F(\mathrm{h}) \tag{6}$$

where $\gamma \in \mathbb{R}^D$ is a learnable scaling parameter initialized to 10^{-2}, and $\odot$ denotes element-wise multiplication. Through the coordinated effect of LayerScale and feature gating, the residual features are adaptively regulated while ensuring stable gradient propagation.

(3) Adaptive Decision-Level Fusion Module

To effectively integrate multimodal information at the decision level, the model introduces a set of globally learnable modality weights, denoted as $\alpha = [\alpha_{\mathrm{MR}}, \alpha_{\mathrm{CT}}, \alpha_{\mathrm{Dose}}]^T$. The modality weights are normalized using the Softmax function to obtain fusion coefficients w_m, satisfying $\sum_m w_m = 1$:

$$w_m = \frac{\exp(\alpha_m)}{\sum_{k\in\{MR,CT,Dose\}} \exp(\alpha_k)}, \quad m \in \{MR, CT, Dose\} \tag{7}$$

The fused logit, $\mathrm{logit}_{\mathrm{fused}}$, is defined as the weighted sum of the modality-specific logits:

$$logit_{fused} = \sum_{m\in\{MR,CT,Dose\}} w_m \cdot logit^{(m)} \tag{8}$$

Finally, the class probabilities p are obtained using the Softmax function:

$$\mathrm{p} = \mathrm{Softmax}(\mathrm{logit}_{\mathrm{fused}}) \tag{9}$$

During decision-level fusion, the learnable global modality weights are normalized by the Softmax function and used to compute a weighted combination of the modality-specific logits [17–18]. The fused representation is subsequently used to generate the final binary prediction of nasopharyngeal necrosis risk.

2.7 Model training

To address class imbalance, a weighted cross-entropy loss function was used to strengthen the learning signal for the minority class (NN). Let the fused output logits of the i-th sample be denoted as $z_i^{(\text{fused})} \in \mathbb{R}^C$, and the corresponding Softmax probability for class $c$be denoted as $p_{i,c}$, where $c = 0$and $c = 1$correspond to the non-necrosis and necrosis classes, respectively. The weighted cross-entropy loss is defined as:

$$p_{i,c} = \frac{\exp(z_{i,c}^{(\text{fused})})}{\sum_{k=0}^{C-1} \exp\,(z_{i,k}^{(\text{fused})})} \tag{10}$$

$$\mathcal{L}\text{oss}_\text{i} = -\text{w}_{\text{y}_\text{i}} \sum_{\text{c}=0}^{1} \text{y}_{\text{i,c}} \log(\,\text{p}_{\text{i,c}}) \tag{11}$$

where y_i denotes the true class label of the i-th sample, and w_{y_i} is the weight coefficient assigned to the corresponding class. By increasing the contribution of minority-class samples to the overall loss, the model was encouraged to learn more discriminative representations for the minority class during parameter optimization.

The model parameters were optimized using the Adam optimizer with an initial learning rate of 0.01 and a weight decay of 1×10^{-3}. During training, the learning rate was adjusted according to the validation loss using the ReduceLROnPlateau strategy; if no improvement was observed for 10 consecutive epochs, the learning rate was reduced by a factor of 0.5. The batch size was set to 32, the maximum number of training epochs was 100, the dropout rate was 0.1, and the number of network layers was 3.

2.8 Model evaluation

All models were independently evaluated in the internal test set. Bootstrap resampling with 1,000 iterations was used to calculate the 95% confidence intervals (CIs) for the area under the receiver operating characteristic curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV). In addition, decision curve analysis (DCA) and calibration curve analysis were performed to evaluate the potential clinical utility and calibration performance of the models.

2.9 Statistical analysis

All statistical analyses were performed using Python 3.8.8. Continuous variables were expressed as mean ± standard deviation (SD) if normally distributed and as median with interquartile range (IQR) otherwise. Categorical variables were summarized as frequencies and percentages. Continuous variables were compared using the Wilcoxon signed-rank test or Kruskal-Wallis test, as appropriate, and categorical variables were compared using the chi-square test. All statistical tests were two-sided, and a P value < 0.05 was considered statistically significant.

3. Results

3.1 Clinical characteristics

A total of 169 patients were screened, and 113 eligible patients were included in the final analysis. The median follow-up duration was 2.5 years, and the final follow-up was completed in September 2024. Patients were randomly assigned to the training set (n = 90) and test set (n = 23) at an approximate ratio of 8:2. Overall, 85 patients were male and 28 were female, and the mean age was 47.5 ± 10.2 years. The median EQD2 was 64 Gy in the training set and 61.1 Gy in the test set. The distributions of sex, age, KPS, tumor stage, treatment strategy, EQD2, and necrosis status were comparable between the training and test cohorts, with no statistically significant differences observed (all P > 0.05). Overall, 39 patients (34.5%) developed NN, whereas 74 patients (65.5%) did not. The incidence of NN was 34.4% (31/90) in the training set and 34.8% (8/23) in the test set, with no significant difference between the two cohorts (P = 1.000). Detailed baseline characteristics are shown in Table 1.

Table 1. Clinical characteristics of the patients

Characteristics	Overall (N=113)	Training set (n=90)	Test set (n=23)	P valuea
Gender, n (%)				
Female	28 (24.8)	24 (26.7)	4 (17.4)	0.516
Male	85 (75.2)	66 (73.3)	19 (82.6)	
Age, mean (SD)	47.5 (10.2)	47.4 (10.6)	47.8 (8.3)	0.871

KPS, median [Q1,Q3]	90.0 [90.0,90.0]	90.0 [90.0,90.0]	90.0 [90.0,90.0]	0.557
Height, mean (SD)	166.1 (7.1)	166.0 (7.3)	166.5 (6.0)	0.743
Weight, mean (SD)	60.8 (11.4)	60.4 (11.9)	62.3 (9.3)	0.395
History of alcohol consumption, n (%)				
No	100 (88.5)	79 (87.8)	21 (91.3)	1.000
Yes	13 (11.5)	11 (12.2)	2 (8.7)	
History of comorbidities, n (%)				
No	88 (77.9)	70 (77.8)	18 (78.3)	1.000
Yes	25 (22.1)	20 (22.2)	5 (21.7)	
Presence of brain metastasis, n (%)				
No	103 (91.2)	82 (91.1)	21 (91.3)	1.000
Yes	10 (8.8)	8 (8.9)	2 (8.7)	
Histological type, n (%)				
Undifferentiated non-keratinizing carcinoma	87 (77.0)	70 (77.8)	17 (73.9)	0.782
Differentiated non-keratinizing carcinoma	8 (7.1)	6 (6.7)	2 (8.7)	
Differentiated keratinizing carcinoma	2 (1.8)	2 (2.2)		
Other	2 (1.8)	2 (2.2)		
Missing	14 (12.4)	10 (11.1)	4 (17.4)	
rTNM8th[b], n (%)				
I	1 (0.9)	1 (1.1)	0	0.938
II	9 (8.0)	8 (8.9)	1 (4.3)	
III	52 (46.0)	41 (45.6)	11 (47.8)	
IV	51 (45.1)	40(44.4)	11 (47.8)	

Targeted therapy, n (%)				
No	80 (70.8)	64 (71.1)	16 (69.6)	1.000
Yes	33 (29.2)	26 (28.9)	7 (30.4)	
Concurrent chemotherapy, n (%)				
No	60 (53.1)	46 (51.1)	14 (60.9)	0.547
Yes	53 (46.9)	44 (48.9)	9 (39.1)	
EQD2, median [Q1,Q3]	64.0 [61.1,64.8]	64.0 [61.1,65.1]	61.1 [60.0,64.4]	0.107
Number of Reirradiation Fractions, median [Q1,Q3]	31.0 [30.0,32.0]	32.0 [30.0,32.0]	30.0 [29.5,32.0]	0.453
Tumor Necrosis, n (%)				
No	74 (65.5)	59 (65.6)	15 (65.2)	1.000
Yes	39 (34.5)	31 (34.4)	8 (34.8)	

Note: rNPC, recurrent nasopharyngeal carcinoma; KPS, Karnofsky Performance Status; UICC/AJCC, Union for International Cancer Control/American Joint Committee on Cancer; EQD2, equivalent dose in 2-Gy fractions; IQR, interquartile range; SD, standard deviation.

[a]*P* values were calculated using the chi-square test for categorical variables and non-parametric test for continuous variables.
[b]According to the 8th edition of the International Union against Cancer/American Joint Committee on Cancer (UICC/AJCC) staging manual.

3.2 Predictive performance of deep learning models constructed from different modalities

Deep learning prediction models based on MRI, CT, dose distribution, and their combinations were constructed and evaluated in the training and independent test sets for predicting NN after re-irradiation in rNPC. The ROC curves, calibration curves, and decision curves are shown in **Figure 3**, and quantitative performance metrics are summarized in **Tables 2 and 3**.

1. Predictive performance of single-modality models

In the training set, the three single-modality models showed different predictive performance. The CT model achieved the highest performance, with an AUC of 0.996 (95% CI, 0.986-1.000), an accuracy of 97.8%, a sensitivity of 100.0%, and a specificity of 96.6%. The dose-distribution model achieved an AUC of 0.902 (95% CI, 0.842-0.963), with an accuracy, sensitivity, and specificity of

84.4%, 80.6%, and 86.4%, respectively. The MRI model achieved an AUC of 0.755 (95% CI, 0.650-0.837), with an accuracy of 73.3%, sensitivity of 29.0%, and specificity of 96.6%.

In the independent test set, all single-modality models demonstrated moderate discriminative ability. The MRI and dose-distribution models both achieved an AUC of 0.800, with 95% CIs of 0.555-0.971 and 0.595-0.962, respectively. Both models had an accuracy of 78.3%; however, the MRI model showed higher specificity (86.7%), whereas the dose-distribution model showed higher sensitivity (75.0%). The CT model achieved an AUC of 0.775 (95% CI, 0.529-0.965), with an accuracy of 82.6%, sensitivity of 75.0%, and specificity of 86.7%.

2. Predictive performance of dual-modality fusion models

Among the dual-modality fusion models, all modality combinations achieved AUCs greater than 0.930 in the training set. The MRI+Dose model achieved the highest training AUC of 0.990 (95% CI, 0.975-0.999), with both sensitivity and NPV reaching 100.0%. The MRI+CT model achieved an AUC of 0.983 (95% CI, 0.958-0.997), also with a sensitivity and NPV of 100.0%. The CT+Dose model achieved an AUC of 0.934 (95% CI, 0.863-0.980), with an accuracy of 86.7%, sensitivity of 93.5%, and specificity of 83.1%.

In the independent test set, the CT+Dose model showed the best performance among the dual-modality models, with an AUC of 0.850 (95% CI, 0.687-1.000), an accuracy of 82.6%, a sensitivity of 75.0%, and a specificity of 86.7%. The MRI+Dose and MRI+CT models both achieved an AUC of 0.825, with a sensitivity of 75.0%. The accuracy and specificity were 78.3% and 80.0% for the MRI+Dose model and 73.9% and 73.3% for the MRI+CT model, respectively.

3. Predictive performance of the trimodality fusion model

The trimodality fusion model integrated MRI, CT, and dose-distribution information. In the training set, this model achieved an AUC of 0.939 (95% CI, 0.876-0.979), with an accuracy of 88.9%, sensitivity of 96.8%, specificity of 84.7%, and NPV of 98.0%.

In the independent test set, the trimodality fusion model achieved the highest AUC among all models, with an AUC of 0.875 (95% CI, 0.696-1.000). At the default threshold of 0.5, the accuracy, sensitivity, specificity, PPV, and NPV were 78.3%, 50.0%, 93.3%, 80.0%, and 77.8%, respectively. These results suggest that the trimodality model had the best discriminative ability, but at the current threshold it tended to favor specificity, with relatively limited sensitivity.

4. Model calibration and decision curve analysis

Calibration curves showed that the predicted probabilities of the models were generally consistent with the observed outcomes, and the fusion models showed relatively stable calibration. The Hosmer-Lemeshow goodness-of-fit test indicated no significant deviation from good calibration. Decision curve analysis showed that the multimodal fusion models provided a higher net benefit than the treat-all or treat-none strategies across a range of threshold probabilities, suggesting potential clinical value for risk stratification. The corresponding results are shown in **Figure 3.**

5. Overall comparison of model performance

Overall, multimodal fusion models achieved higher AUCs than single-modality models in the independent test set. The AUCs of the single-modality models ranged from 0.775 to 0.800, whereas those of the dual-modality fusion models ranged from 0.825 to 0.850, and the AUC of the trimodality model further increased to 0.875. Compared with their performance in the training set, some models, particularly the CT, MRI+Dose, and MRI+CT models, showed a decline in performance in the test set after achieving near-perfect or very high performance in the training set, suggesting possible overfitting. In contrast, the CT+Dose model maintained favorable discriminative ability and relatively balanced sensitivity and specificity in both the training and test sets. Although the trimodality model achieved the highest test-set AUC, its sensitivity was markedly affected by the default threshold. Therefore, future applications should optimize the classification threshold according to specific clinical objectives, such as prioritizing sensitivity for high-risk screening or prioritizing specificity to reduce false-positive predictions.

Table 2. Performance of single-modality, dual-modality, and trimodality deep learning models in the training set

Modality	PPV(95%CI)	ACC(95%CI)	SEN(95%CI)	SPE(95%CI)	NPV(95%CI)	AUC(95%CI)
MR	0.818 (0.555, 1.000)	0.733 (0.633, 0.822)	0.290 (0.120, 0.447)	0.966 (0.914, 1.000)	0.722 (0.618, 0.812)	0.755 (0.650, 0.837)

Dose	0.758 (0.625, 0.889)	0.844 (0.777, 0.911)	0.806 (0.686, 0.926)	0.864 (0.792, 0.941)	0.895 (0.803, 0.962)	0.902 (0.842, 0.963)
CT	0.939 (0.844, 1.000)	0.978 (0.944, 1.000)	1.000 (1.000, 1.000)	0.966 (0.907, 1.000)	1.000 (1.000, 1.000)	0.996 (0.986, 1.000)
MR+Dose	0.795 (0.667, 0.900)	0.911 (0.856, 0.956)	1.000 (1.000, 1.000)	0.864 (0.786, 0.937)	1.000 (1.000, 1.000)	0.990 (0.975, 0.999)
MR+CT	0.689 (0.540, 0.813)	0.844 (0.756, 0.911)	1.000 (1.000, 1.000)	0.763 (0.635, 0.862)	1.000 (1.000, 1.000)	0.983 (0.958, 0.997)
CT+Dose	0.744 (0.600, 0.889)	0.867 (0.789, 0.933)	0.935 (0.833, 1.000)	0.831 (0.730, 0.922)	0.961 (0.900, 1.000)	0.934 (0.863, 0.980)
MR+CT+Dose	0.769 (0.611, 0.897)	0.889 (0.822, 0.944)	0.968 (0.880, 1.000)	0.847 (0.758, 0.926)	0.980 (0.932, 1.000)	0.939 (0.876, 0.979)

Note: AUC, area under the receiver operating characteristic curve; ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval; MR, magnetic resonance (combining T1WI, T2WI and T1WI+C features); CT, computed tomography; Dose, plan dose features.

Table 3. Performance of single-modality, dual-modality, and trimodality deep learning models in the independent test set

Modality	PPV(95%CI)	ACC(95%CI)	SEN(95%CI)	SPE(95%CI)	NPV(95%CI)	AUC(95%CI)
MR	0.714 (0.331, 1.000)	0.783 (0.565, 0.913)	0.625 (0.250, 1.000)	0.867 (0.667, 1.000)	0.812 (0.571, 1.000)	0.800 (0.555, 0.971)
Dose	0.667 (0.333, 1.000)	0.783 (0.609, 0.914)	0.750 (0.333, 1.000)	0.800 (0.571, 1.000)	0.857 (0.688, 1.000)	0.800 (0.595, 0.962)
CT	0.750 (0.398, 1.000)	0.826 (0.696, 0.957)	0.750 (0.500, 1.000)	0.867 (0.666, 1.000)	0.867 (0.700, 1.000)	0.775 (0.529, 0.965)
MR+Dose	0.667 (0.333, 1.000)	0.783 (0.609, 0.957)	0.750 (0.500, 1.000)	0.800 (0.583, 1.000)	0.857 (0.692, 1.000)	0.825 (0.571, 1.000)
MR+CT	0.600 (0.272, 0.875)	0.739 (0.564, 0.913)	0.750 (0.500, 1.000)	0.733 (0.499, 0.929)	0.846 (0.635, 1.000)	0.825 (0.599, 1.000)
CT+Dose	0.750 (0.400, 1.000)	0.826 (0.696, 0.957)	0.750 (0.428, 1.000)	0.867 (0.706, 1.000)	0.867 (0.714, 1.000)	0.850 (0.687, 1.000)
MR+CT+Dose	**0.800 (0.333, 1.000)**	**0.783 (0.652, 0.957)**	**0.500 (0.199, 0.875)**	**0.933 (0.786, 1.000)**	**0.778 (0.600, 0.950)**	**0.875 (0.696, 1.000)**

Note: AUC, area under the receiver operating characteristic curve; ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval; MR, magnetic resonance (combining T1WI, T2WI and T1WI+C features); CT, computed tomography; Dose, plan dose features.

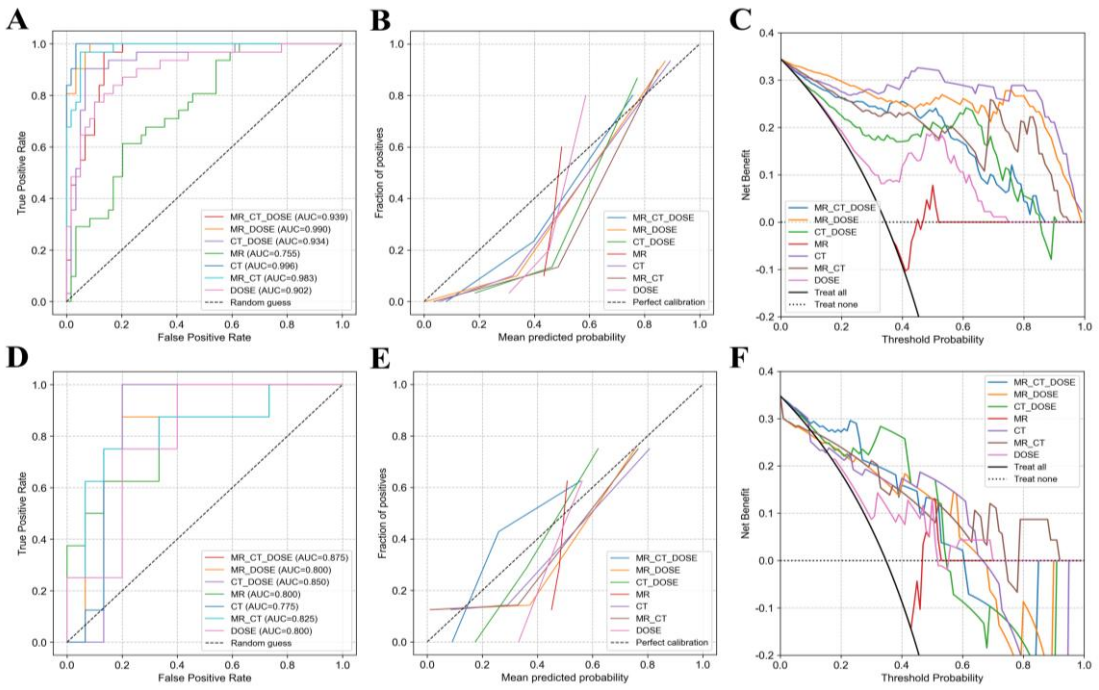

Figure 3. Model performance in the training and test sets. Performance evaluation of single-modality, dual-modality, and trimodality models included receiver operating characteristic curves, calibration curves, and decision curve analysis. (A) ROC curves in the training set. (B) Calibration curves in the training set. (C) Decision curve analysis in the training set. (D) ROC curves in the test set. (E) Calibration curves in the test set. (F) Decision curve analysis in the test set.

4. Discussion

Nasopharyngeal necrosis is one of the most serious late complications after re-irradiation for recurrent nasopharyngeal carcinoma, and accurate pretreatment risk prediction is essential for individualized treatment planning[8,9]. In this study, we developed a multimodal deep learning fusion model integrating planning CT, multisequence MRI, and three-dimensional dose distribution to predict the risk of nasopharyngeal necrosis after re-irradiation. The results showed that multimodal fusion models generally outperformed single-modality models. In the independent test set, the CT+Dose model showed robust performance with an AUC of 0.850, whereas the trimodality MR+CT+Dose model achieved the highest AUC of 0.875. However, the trimodality model showed relatively low sensitivity at the default threshold of 0.5, suggesting that threshold optimization is necessary before clinical application.

Previous studies have developed clinical or radiomic models for predicting radiation-related necrosis in nasopharyngeal carcinoma. Clinical nomograms based on demographic and treatment-related variables have shown value in risk stratification, but they may not fully capture the spatial heterogeneity of recurrent tumors and surrounding irradiated tissues [8,9]. MRI-based radiomic studies further demonstrated that imaging phenotypes could reflect tissue tolerance, tumor heterogeneity, and microenvironmental changes associated with nasopharyngeal necrosis [10]. Consistent with these findings, the MRI model in this study achieved moderate predictive performance, supporting the value of multisequence MRI in pretreatment risk assessment.

Dose distribution is another key determinant of radiation-induced tissue injury. Traditional dose-volume parameters have been associated with radiation necrosis in nasopharyngeal carcinoma and other head and neck settings, while dosiomics provides additional spatial and textural information from three-dimensional dose distributions [26-29]. In this study, the Dose model achieved an AUC of 0.800 in the test set, and the CT+Dose model further improved the AUC to 0.850. These findings suggest that anatomical information from planning CT and spatial dose information may provide complementary value for predicting nasopharyngeal necrosis. Given that both CT and dose data are routinely available during radiotherapy planning, the CT+Dose model may have favorable clinical accessibility and translational potential.

Compared with previous studies that mainly focused on MRI combined with dose information or CT combined with dose and clinical variables [11-13], the present study simultaneously incorporated CT, multisequence MRI, and three-dimensional dose distribution. The trimodality model achieved the highest discriminative performance, indicating that these modalities may provide complementary information for risk prediction. Methodologically, the modality-specific encoding and decision-level weighted fusion strategy allowed each modality to be learned independently before adaptive integration, which may reduce cross-modality interference and improve model stability. This is consistent with recent studies suggesting that fusion strategy is important for multimodal biomedical data modeling [14-18].

The clinical significance of this study lies in its potential to support individualized decision-making before re-irradiation. For patients predicted to be at high risk of nasopharyngeal necrosis, clinicians may consider optimizing dose distribution, adjusting target delineation, intensifying

follow-up, or selecting alternative treatment strategies. Nevertheless, the model should be interpreted cautiously. The trimodality model achieved the highest AUC but had limited sensitivity under the default threshold, indicating that different cut-off values should be explored according to clinical objectives, such as high-sensitivity screening or high-specificity confirmation.

This study has several limitations. First, it was a single-center retrospective study with a relatively small sample size, especially in the independent test set. Second, the number of positive events was limited, which may affect the stability of model performance estimates. Third, external validation was not performed, and the generalizability of the model across institutions, scanners, and treatment protocols remains uncertain. Fourth, the interpretability of the deep learning fusion model remains limited. Future studies should include larger multicenter cohorts, external validation, threshold optimization, and model interpretability analysis to further evaluate clinical applicability.

This study has several limitations. First, it was a single-center retrospective study with a limited sample size and few positive events in the test set, which may affect the stability and generalizability of the results. External validation in larger multicenter cohorts is therefore required, as insufficient validation remains a major barrier to the clinical translation of AI-based toxicity prediction models [30]. Second, the trimodality model showed limited sensitivity at the default threshold of 0.5, indicating that threshold optimization based on clinical objectives is needed. Third, model interpretability remains limited; future studies should incorporate feature-importance analysis, modality-contribution assessment, and visualization methods, and may further integrate clinical variables, EBV DNA, and molecular biomarkers to improve biological interpretability.

In conclusion, the proposed multimodal deep learning fusion model showed promising performance in predicting nasopharyngeal necrosis after re-irradiation in recurrent nasopharyngeal carcinoma. CT, MRI, and dose distribution provided complementary information, with the trimodality model achieving the highest AUC and the CT+Dose model showing favorable robustness and clinical accessibility. Further multicenter validation is warranted before clinical implementation.

Author Contributions

(I) Conception and design: Yinglin Peng, Jun Zhang, Yimei Liu, Meining Chen.

(II) Administrative support: Yinglin Peng, Jun Zhang, Yimei Liu.

(III) Provision of study materials or patients: Chong Zhao, Xiaowu Deng, Xingwang Gao, Jingjing Miao, Zhenyu Qi, Qichao Zhou.

(IV) Collection and assembly of data: Meining Chen, Xingwang Gao, Jingjing Miao, Zhirui Huang.

(V) Data analysis and interpretation: Xuan Guo, Jiaxin Lin, Xiao Liu, Zhenyu Qi, Qichao Zhou.

(VI) Manuscript writing: All authors.

(VII) Final approval of manuscript: All authors.

Acknowledgments

Not applicable

Ethics approval and consent to participate

This retrospective study complied with the Declaration of Helsinki, and it was approved by the Ethics Committee of Sun Yat-sen University Cancer Center (No. B2023-103-01). The requirement for written informed consent was waived by the Ethics Committee of Sun Yat-sen University Cancer Center.

Funding

This work was jointly supported by the National Natural Science Foundation of China (grant numbers:12505426 & 12475360)

Conflicts of Interest:

The authors have no conflicts of interest to declare.

Data Sharing Statement:

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Ng WT, Soong YL, Ahn YC, AlHussain H, Choi HCW, Corry J, et al. International recommendations on reirradiation by intensity modulated radiation therapy for locally recurrent nasopharyngeal carcinoma. Int J Radiat Oncol Biol Phys. 2021;110:682-95.
2. Pandrangi VC, Liao JJ, de Almeida JR, El-Sayed IH, Hanna G, Su SY, Tsang R, Won TB, Witterick I, Choby G, Kuan EC, Geltzeiler M. Current trends in the management of recurrent nasopharyngeal carcinoma. Head Neck. 2025;47:2611-21.
3. Leong YH, Soon YY, Lee KM, Wong LC, Tham IWK, Ho FCH. Long-term outcomes after reirradiation in nasopharyngeal carcinoma with intensity-modulated radiotherapy: a meta-analysis. Head Neck. 2018;40:622-31.
4. 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;36:891-9.
5. Tian YM, Zhao C, Guo Y, Huang Y, Huang SM, Deng XW, Lin CG, Lu TX, Han F. Effect of total dose and fraction size on survival of patients with locally recurrent nasopharyngeal carcinoma treated with intensity-modulated radiotherapy: a phase 2, single-center, randomized controlled trial. Cancer. 2014;120:3502-9.
6. Tian YM, Guan Y, Xiao WW, Zeng L, Liu S, Lu TX, Zhao C, Han F. Long-term survival and late complications in intensity-modulated radiotherapy of locally recurrent T1 to T2 nasopharyngeal carcinoma. Head Neck. 2016;38:225-31.
7. Wu WB, Zhang XB, Liu YP, Zou X, You R, Xie YL, Duan XT, Li HF, Wen K, Peng L, Hua YJ, Huang PY, Sun R, Chen JH, Chen MY. Stent pretreatment for internal carotid artery exposed to necrotic lesions in nasopharyngeal carcinoma. Rhinology. 2023;61:246-54.
8. Li XY, Sun XS, Liu SL, Li HP, Luo MJ, Chen QY, et al. The development of a nomogram to predict post-radiation necrosis in nasopharyngeal carcinoma patients: a large-scale cohort study. Cancer Manag Res. 2019;11:6253-63.
9. Yu YH, Xia WX, Shi JL, Ma WJ, Yao M, Du XJ, et al. A model to predict the risk of lethal nasopharyngeal necrosis after re-irradiation with intensity-modulated radiotherapy in nasopharyngeal carcinoma patients. Chin J Cancer. 2016;35:59.
10. Liu T, Dong D, Zhao X, Wang K, Sun R, Tian J, et al. Radiomic signatures reveal multiscale intratumor heterogeneity associated with tissue tolerance and survival in re-irradiated nasopharyngeal carcinoma: a multicenter study. BMC Med. 2023;21:464.
11. Gao X, Peng Y, Lu S, An Y, Chen M, Zhang J, et al. Deep learning-based multi-omics model to predict nasopharyngeal necrosis of re-irradiation for recurrent nasopharyngeal carcinoma. Front Oncol. 2025;15:1607218.
12. Lu S, Xiao X, Yan Z, Cheng T, Tan X, Zhao R, Wu H, Shen L, Zhang Z. Prognosis forecast of re-irradiation for recurrent nasopharyngeal carcinoma based on deep learning multi-modal information fusion. IEEE J Biomed Health Inform 2023. doi: 10.1109/JBHI.2023.3286656.
13. Gao X, Peng Y, Miao J, Lin J, Fei Z, Liu X, et al. A multimodal fusion model for predicting nasopharyngeal necrosis after re-irradiation in recurrent nasopharyngeal carcinoma. Appl Radiat Isot. 2026;231:112513.
14. Yang H, Yang M, Chen J, Yao G, Zou Q, Jia L. Multimodal deep learning approaches for precision oncology: a comprehensive review. Brief Bioinform. 2025;26:bbae699.

15. Kline A, Wang H, Li Y, Dennis S, Hutch M, Xu Z, et al. Multimodal machine learning in precision health: a scoping review. NPJ Digit Med. 2022;5:171.
16. Li Y, Daho MEH, Conze PH, Fablet R, Heimann T, Nysjö J, et al. A review of deep learning-based information fusion techniques for multimodal medical image classification. Comput Biol Med. 2024;177:108635.
17. Stahlschmidt SR, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. Brief Bioinform. 2022;23:bbab569.
18. Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit Med. 2020;3:136.
19. Huang RD, Sun Z, Wang XH, Chen QY, Tang LQ, Mai HQ, et al. Development of a comorbidity-based nomogram to predict survival after salvage reirradiation of locally recurrent nasopharyngeal carcinoma in the intensity-modulated radiotherapy era. Front Oncol. 2020;10:625184.
20. McMahon SJ. The linear quadratic model: usage, interpretation and challenges. Phys Med Biol. 2018;64:01TR01.
21. International Commission on Radiation Units and Measurements. Prescribing, Recording, and Reporting Photon Beam Therapy. ICRU Report 50. Bethesda, MD: ICRU; 1993.
22. International Commission on Radiation Units and Measurements. Prescribing, Recording, and Reporting Photon Beam Therapy (Supplement to ICRU Report 50). ICRU Report 62. Bethesda, MD: ICRU; 1999.
23. International Commission on Radiation Units and Measurements. Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT). ICRU Report 83. J ICRU 2010;10:1-106.
24. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010;29:1310-20.
25. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016; p. 770-8.
26. Su SF, Huang Y, Xiao WW, Huang SM, Han F, Xie CM, Lu TX. Clinical and dosimetric characteristics of temporal lobe injury following intensity-modulated radiotherapy of nasopharyngeal carcinoma. Radiother Oncol. 2012;104:312-6.
27. Zhou X, Ou X, Xu T, Wang X, Shen C, Ding J, Hu C. Effect of dosimetric factors on occurrence and volume of temporal lobe necrosis following intensity-modulated radiation therapy for nasopharyngeal carcinoma: a case-control study. Int J Radiat Oncol Biol Phys. 2014;90:261-9.
28. Yang SS, OuYang PY, Guo JG, Chen YJ, Deng W, Ye HY, et al. Dosiomics risk model for predicting radiation-induced temporal lobe injury and guiding individual intensity-modulated radiation therapy. Int J Radiat Oncol Biol Phys. 2023;115:1291-300.
29. Liang B, Yan H, Tian Y, Chen X, Yan L, Zhang T, et al. Dosiomics: extracting 3D spatial features from dose distribution to predict incidence of radiation pneumonitis. Front Oncol. 2019;9:269.
30. Ugwu GC, Jalali F, Liu G, Li G, Langendijk JA, Alizadeh BZ. The actual performance of ML/AI models in predicting radiation-induced toxicity in head and neck cancer: a systematic review and meta-analysis. Radiother Oncol. 2026;216:111350.

Supplementary Material

Supplementary Tables 1-4 and Supplementary A are available as separate supplementary files.

Figure Legends

Figure 1. Flowchart of patient selection.

Figure 2. Architecture of the trimodality deep learning fusion network.

Figure 3. Model performance in the training and test sets. Performance evaluation of single-modality, dual-modality, and trimodality models included receiver operating characteristic curves, calibration curves, and decision curve analysis. (A) Receiver operating characteristic curves in the training set. (B) Calibration curves in the training set. (C) Decision curve analysis in the training set. (D) Receiver operating characteristic curves in the test set. (E) Calibration curves in the test set. (F) Decision curve analysis in the test set. ROC, receiver operating characteristic.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.docx