Prognosis Forecast of Re-irradiation for Recurrent Nasopharyngeal Carcinoma based on Deep Learning Multi-modal Information Fusion

Shanfu Lu, Xiang Xiao, Ziye Yan, Tingting Chen, Xufang Tan, Rongchang Zhao, Haijun Wu, Liangfang Shen, Zijian Zhang*

***Abstract*— Radiation therapy is the primary treatment for recurrent nasopharyngeal carcinoma. However, it may induce necrosis of the nasopharynx, leading to severe complications such as bleeding and headache. Therefore, forecasting necrosis of the nasopharynx and initiating timely clinical intervention has important implications for reducing complications caused by re-irradiation. This research informs clinical decision-making by making predictions on re-irradiation of recurrent nasopharyngeal carcinoma using deep learning multi-modal information fusion between multi-sequence nuclear magnetic resonance imaging and plan dose. Specifically, we assume that the hidden variables of model data can be divided into two categories: task-consistency and task-inconsistency. The task-consistency variables are characteristic variables contributing to target tasks, while the task-inconsistency variables are not apparently helpful. These modal characteristics are adaptively fused when the relevant tasks are expressed through the construction of supervised classification loss and self-supervised reconstruction loss. The cooperation of supervised classification loss and self-supervised reconstruction loss simultaneously reserves the information of characteristic space and controls potential interference simultaneously. Finally, multi-modal fusion effectively fuses information through an adaptive linking module. We evaluated this method on a multi-center dataset. and found the prediction based on multi-modal features fusion outperformed predictions based on single-modal, partial modal fusion or traditional machine learning methods.**

***Index Terms*— Recurrent Nasopharyngeal Carcinoma, Prognosis forecast, Task-Consistency feature extract, Multi-modal fusion, Deep learning.**

This work is supported by Natural Science Foundation of Hunan Province, No.2021JJ70151, and Clinical medical technology innovation guidance project of Hunan Province, No. 2020SK53706.

S. Lu and Z. Yan are with Perception Vision Medical Technologies Co., Ltd., Guangzhou, 510530, China. (email: lushanfu@pvmedtech.com)

Z. Zhang, H. Wu and F. Shen are with National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China, and Department of Oncology, Xiangya Hospital, Central South University, Changsha, China. (e-mail: wanzzj@csu.edu.cn)

T. Chen is with Department of General practice, Xiangya Hospital, Central South University, Changsha 410008, P.R. China, and National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Changsha 410008, P.R. China.

X. Tan is with Department of Nursing, Xiangya Hospital, Central South University, Changsha 410008, P.R. China.

R. Zhao is with School of Computer Science and Engineering, Central South University, Changsha 410083, P.R. China.

X. Xiao is with Hunan Cancer Hospital/the Affiliated Cancer Hospital of Xiangya School of Medicine, Central South UniversityChangsha 410083, P.R. China.

*Corresponding author

I. Introduction

NASOPHARYNGEAL carcinoma(NPC) is a head and neck cancer common in southeast Asia [1], [2]. The standard treatment for early disease is radiotherapy, and for advanced disease, radiotherapy and chemotherapy may be combined. Radiotherapy and combined modality treatments have significantly improved local control of advanced NPC [3], [4], but local recurrence remains a major cause of treatment failure [5], [6]. In many cases of recurrent NPC, infiltration has already occurred at the time of diagnosis, making re-irradiation therapy the only option for salvage. After salvage therapy, a large portion of patients can survive for years; therefore, aggressive treatment with curative intent is generally advisable [7], [8]. Re-irradiation using intensity-modulated radiation therapy (IMRT) is the primary method of recurrent nasopharyngeal carcinoma treatment. However, radiation-induced nasopharyngeal necrosis, as the leading adverse effect during re-irradiation [9], may cause potentially life-threatening bleeding to patients if it cannot be controlled [10]. Therefore, it is of great significance to improve the prognosis of patients to forecast the risk of nasopharyngeal necrosis before re-irradiation.

The application of radiomics in oncology has garnered significant attention in recent years. The overarching theme of the field is to extracts quantitative characteristics such as form, texture, and wavelet in order to conduct deep representation of disease phenotype and build statistical models for the diagnosis, classification, and prognosis of diseases [11], [12]. Previous successful experiences using deep learning on medical image analysis tasks demonstrated radiomics is able to operate end-to-end modeling over the course of medical interactions, conducting multi-task learning on multiple clinic tasks with a sound predictive effect [13]. Thus, radiomics based on deep learning has potential value for forecasting prognosis and therapeutic reaction of nasopharyngeal carcinoma [14], [15], [16].

With the rapid development of diagnostic imaging, technologies such as magnetic resonance imaging(MRI), computed tomography(CT), and positron emission tomography (PET)

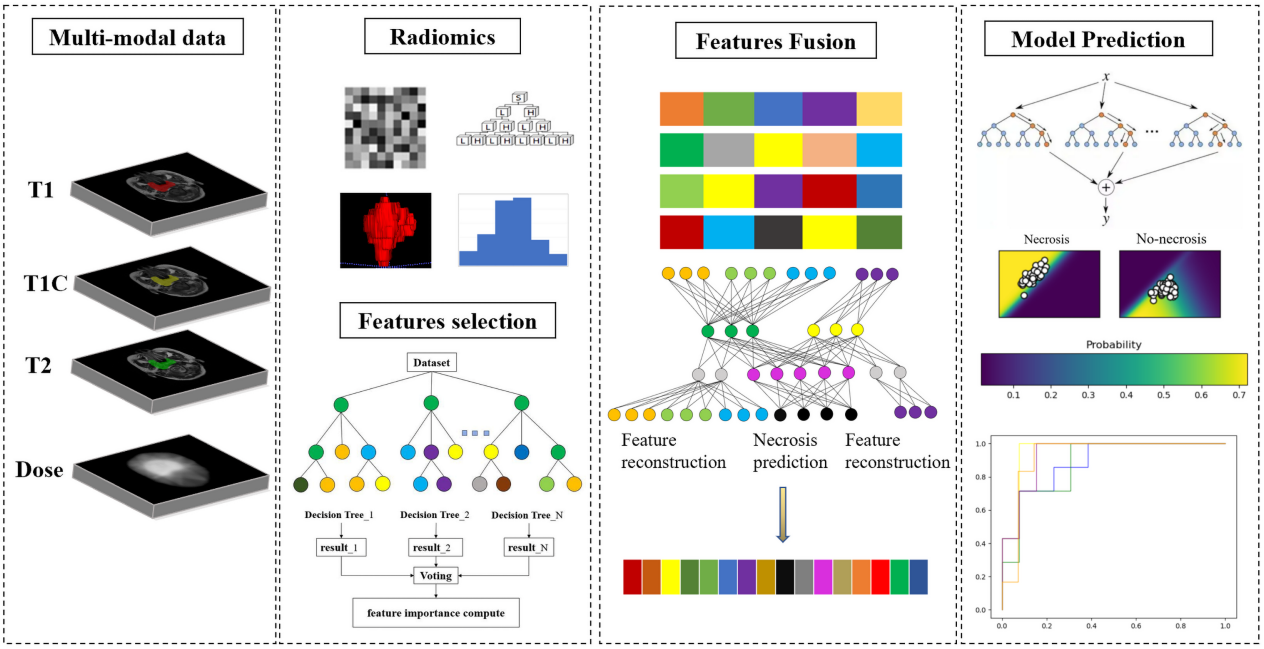

Fig. 1: Flow diagram of re-radiotherapy for recurrent nasopharyngeal carcinoma necrosis forecast. Multi-modal data: obtained the multi-modal data: T1, T1C, T2, and dose. Radiomics and Features selection: to extract the radiomics features of the multi-modal data and reduce the original features dimension. Features Fusion: constructed a multi-modal feature fusion network and train the proposed model. Model Prediction: The test was carried out based on the trained model.

are widely used in the diagnosis and prognosis analysis of many diseases. Liu et al. [17] reviewed the development situation of COVID-19 diagnosis and forecast based on deep learning and medical image analysis. The authors highlight that the application of deep learning in medical images play an important guiding role in COVID-19's diagnosis and forecast. Khvostikov et al. [18] proposed the application of the 3DCNN classification method to research Alzheimer's using sMRI and MD-DTI images. The method fuses sMRI and MD-DTI images on 3DCNN to establish a combined analysis method of diseases. Guo et al. proposed an NPC MR image T-staging method based on weak supervised deep learning [19]. This method used image slices of patients and TNM staging labels to construct weakly supervised label data and built a classification network using ResNet for T staging. Zhong et al. [20] proposed a method based on multiple parameters MR image fusion for prognosis forecast of NPC patients. This method extracts features on MR images with different modalities based on deep neural network and makes feature fusion to build a survival prediction model of patients. Liu et al. [21] discussed the value of prognosis of pathological microscopic features and the effect on treatment decisions using deep learning methods reliable instruments to forecast the survival risks of NPC patients and may have the capability to guide the treatment decision. Yue et al. [22] proposed a multi-loss disentangled representation learning method. Data at different stages underwent multi-feature fusion to enhance the forecast ability of pCR (pathologic complete response) after neoadjuvant chemoradiotherapy (nCRT). Using the regressive analysis method, Yu et al. [23] built a mathematical model based on clinical features. This model is applied for forecasting the locoregional necrosis of recurrent NPC patients were receiving IMRT. However, this model can only provide a risk rating scale. Accurate prediction for individuals cannot be carried out under this model.

Although deep learning and radiomics have achieved great progress on medical image analysis [24], [25], [26], no research has been conducted focusing on nasopharyngeal necrosis before re-irradiation of recurrent NPC. In the current study, we build an automatic forecast model of re-irradiation on recurrent NPC patients and discuss the relationship between multi-modal data and NPC nasopharyngeal necrosis forecast.

The main research questions in the current study include the following points: 1. How radiation dose data can constitute effective features of NPC necrosis forecast; 2. How multi-modal features can be fused effectively; 3. The significance of multi-modal features fusion for nasopharyngeal necrosis forecasting.

Radiation dose data is usually calculated within the gross tumor volume (GTV) at the three-dimensional voxel level. Many types of research indicate that dose volume histograms (DVHs) generated by radiotherapy planning are related to disease prognosis. However, while useful for plan assessment, DVHs are a two-dimensional representation of the dose distribution, removing the spatially relevant information of the inherent three-dimensional dose distribution [27], [28]. Moreover, for medical images, different modalities contain different information. The fusion of multi-modal information can effectively improve the perception ability of the task. The same applies to a variety of radiomics-based tasks. Previous multi-modal omics features fusion methods generally directly concatenated or added to train the target task [29], [30].

TABLE I: Cohorts information of all samples in this study.

	Dataset A			Dataset B		
	No necrosis	necrosis		No necrosis	necrosis	
Gender(%)						
Male	63(31.5%)	43(21.5%)		10(33.3%)	6(20%)	
Female	57(28.5%)	37(18.5%)	P-value=0.9769	9(30%)	5(16.7%)	P-value=1.000
Age(mean±std)	47(24-75)	48(24-75)		49(28-63)	48(24-69)	
Overall stage[a], No.(%)						
I	5(2.5%)	1(0.5%)		2(6.67%)	0(0%)	
II	23(11.5%)	14(7%)		3(10%)	2(6.67%)	
III	35(17.5%)	25(12.5%)		6(20%)	4(13.33%)	
IV	57(28.5%)	40(20%)	P-value=0.7401	8(26.67%)	5(16.66%)	P-value=0.8885
chemotherapy(%)						
Yes	37(18.5%)	31(15.5%)		2(6.67%)	3(10%)	
No	83(41.5%)	49(24.5%)	P-value=0.314	17(56.67%)	8(26.66%)	P-value=0.3268

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

However, this method cannot eliminate the redundancy of multi-modal features for effective fusion. Therefore, it is necessary to explore effective multi-modal feature fusion methods. Recently, disentangled representation learning has played an essential role in multi-modal feature fusion. Disentangled representation learning can disassemble information into numerous independent factors where, every factor comprises a valuable aspect [22], [31], [32]. Xu et al. fused vocal features of the different models to diagnose machinery faults by applying a dense multiscale network [33]. Guo et al. [34] used deep disentangled representation learning to forecast the situation of lymphoma prognosis. Hu et al. [35] propose a method of finding general characteristics and specific characteristics among multi-modal MRI through a disentangled multi-modal antagonistic autoencoder. It used a joint ratio loss function under multi-modal MRI to restrain the structure of latent space.

Inspired by the aforementioned literature, we propose a method of predicting recurrent NPC necrosis based on multi-modal information fusion. For radiotherapy dose data, we extract radiomics features to make the quantitative description of a three-dimensional feature of radiotherapy dose; extract task-consistency features and task inconsistency features from different modal data through a two-stage feature fusion method; build a forecast model for re-irradiation necrosis of recurrent nasopharyngeal carcinoma by making full use of multi-modal MRI (T1, T1C, T2) images, and radiomics feature complementary information of radiotherapy dose. The experimental results indicate that our forecast model for re-irradiation necrosis of recurrent NPC is not only better than unimodal data, but also better than the traditional machine learning method. In addition, compared with some existing model fusion methods, our method also has certain advantages.

Our main contribution can be summarised as follows:

1. Build a method of multi-modal feature fusion to forecast re-irradiation necrosis of recurrent nasopharyngeal carcinoma.

2. Apply radiomics on radiotherapy dose data of NPC tentatively to extract abundant three-dimensional feature information.

3. Fuse omics features of different models using multiscale dense connection method to actualize effective fusion of multi-modal MRI.

4. Fuse MR features and radiotherapy dose features by disentangled representation method, extracting task-consistency features and task-inconsistency features.

5. Design multi-loss function self-supervised reconstruction loss and supervised classification loss, which makes feature fusion of each stage have task-consistency constrain.

II. Materials and Methods

A. Materials

1) Patients: Two datasets (Dataset A and Dataset B) accumulated over different periods were used in this study. The demographic information of all the samples is summarized in Table 1. Dataset A is a retrospective analysis of 200 patients with locally recurrent NPC who received re-irradiation +/- chemotherapy in our hospital between Jun 1, 2013, and Dec 31, 2020. Dataset B is a retrospective analysis of 30 patients with locally recurrent NPC who received re-irradiation +/- chemotherapy in our hospital between Jan 1, 2021, and Feb 31, 2023. All patients were re-staged according to the eighth edition of the International Union Against Cancer/American Joint Committee on Cancer (UICC/AJCC) staging system. Pre-irradiation routine and enhanced MRI scans were required for patients to be eligible for study inclusion. Dataset A was used both to train the model and to evaluate model performance, while dataset B was only used to evaluate model performance. Most patients disease was pathologically confirmed. Other patients with recurrence in inaccessible sites, such as skull base or cavernous sinus, were mainly diagnosed based on imaging representation and clinical symptoms.

2) Magnetic resonance imaging (MRI): All MRI examinations were conducted in the radiology department of our hospital, with a 3.0-T MRI system. High-quality MRI images were obtained using the following protocols: axial T1WI: layer thickness 4 mm, layer spacing: 1 mm, matrix: 320×256, and field of view (FOV) : 24 cm × 24 cm; axial T2WI: layer thickness: 5 mm, layer spacing: 1 mm, matrix: 288×192, and FOV: 24 cm × 24 cm; and axial contrast-enhanced T1WI (T1CE): layer thickness: 3 mm, layer spacing 1 mm, matrix 256 × 256, and FOV 28 cm × 28 cm.

All MRI images were retrieved from the picture archiving and communication system for image feature extraction.

3) Chemotherapy: As shown in Table 1, all patients who received chemotherapy,underwent Cisplatin-based chemotherapy (cisplatin alone or cisplatin plus other one or two anti-tumor drugs).

4) Radiation therapy: All patients were treated with IMRT. The IMRT plan was designed according to the treatment protocol for recurrent NPC at our study hospital. Tumor volumes were delineated in accordance with the International Commission on Radiation Units and Measurements(ICRU) Report No. 62 and ICRU No. 50. The delineation of recurrent gross tumor volumes (GTVnx and GTVnd) was determined from the MR images. The clinical tumor volume (CTV) included GTV plus a 2 to 3 mm margin. Critical normal structures, including the brainstem, spinal cord, parotid glands, optic nerves and chiasm, lens, eyeballs, temporal lobes, temporomandibular joints, mandible, and hypophysis were contoured and set as organs at risk (OARs) during optimization. The prescribed dose was 60–68 Gy to the GTV and 50–54 Gy to the CTV in 30 to 34 fractions. All patients received full-course IMRT with 6-MV x-rays generated by a Trilogy linear accelerator (Varian Medical Systems, Palo Alto, CA, USA). Dose verification was carried out before re-irradiation. The dose error between the measurement and the plan should was less than 2%.

The radiotherapy dose data were retrieved from the treatment planning system for further dose feature extraction.

5) Diagnosis of nasopharyngeal necrosis: Diagnosis of nasopharyngeal necrosis was based on clinical characteristics, including foul nasal smell, refractory headache, filemot necrotic tissue and skull base osteoradionecrosis in nasopharyngeal cavity under endoscopy, discontinuous nasopharyngeal mucosa line and/or tissue defects on MRI, and a heap of red-stained substance without cellular structure in hematoxylin-eosin staining under pathologic examination [36], [37], [38], [39], [40], [41], [42], [43]. Patients who died from intractable epistaxis diagnosed with nasopharyngeal necrosis were recorded as having lethal nasopharyngeal necrosis (LNN).

B. Method overview

The framework of re-irradiation for recurrent NPC necrosis forecast based on multi-modal feature fusion is shown in Fig. 1. The whole process is constituted of 4 sections: 1) At the stage of multi-modal data acquisition, T1, T1C, T2, dose modal data of patients are respectively acquired for lesion area delineation by physicians with extensive clinical experience. 2) Extract radiomics features of multi-modal data by delineation and analyzing the significance of each feature. 3) Conduct multi-modal omics feature fusion by the model mentioned before to achieve information sharing and complement. 4) Forecast the probability of recurrent NPC necrosis based on the multi-modal information fusion.

According to the successful experience of multi-modal radiomics in medical image analysis, we attempt to mine quantified dose features from three-dimensional dose distribution of nasopharyngeal carcinoma. The following progress is executed for the standardization of radiomics feature extraction progress:

First, the multi-modal MRI data should be field-biased corrected to reduce the influence caused by magnetic deviation. Then reserve the region of interest (ROI) based on the carcinoma area delineated by a physician and eliminate background interference. Register the radiotherapy dose data to the corresponding MRI carcinoma ROI. At last, use the Python programming language based software package pyradiomics 3.0.1 to extract radiomics features from T1, T1C, T2 and dose files. The types of extracted features include form feature, first-order feature, grey-level co-occurrence matrix (GLCM), grey-level run-length matrix (GLRLM), grey-level dependence matrix (GLDM), wavelet feature, etc. 1562 features are extracted respectively from every modal data through radiomics feature extraction. To relieve the numerical instability of the model during training, the extracted features are z-score normalized to eliminate dimensional inconsistencies.

The radiomics features extracted from four different sources of modal data have a high information dimension, which may cause redundancy. We used a feature dimension reduction technique to compress the feature dimension. specifically, we computed pair-correlation, and the features with high correlation were eliminated. Furthermore, the random forest method was used for feature dimension reduction. Leveraging the random search, the number of final features of each modal dataset was reduced to 49.

To effectively perform multi-modal feature fusion, we proposed a two-stage feature deep fusion model, namely, multi-modal MR feature fusion and multi-modal MR and radiation dose feature fusion respectively. In the beginning, a multi-scale feature fusion network was used to extract the same embedding space feature representation of different sequences of MRI. Then multi-modal MR fusion and dose input features were coded by latent spacing embedding separately. The combination was performed adaptively; that is, the task consistency feature was constructed for the NPC necrosis prediction task. Next, we described the proposed model architecture, multi-MR information fusion, latent spacing embedding of MR, and dose feature vectors, and adaptive task consistency feature fusion in detail.

The proposed model framework is shown in Fig. 2. The input of networks consists of four modal features, which are T1, T1C, and T2. Different image display methods of the same tumor region, so the features have certain independence and complementarity. The Multi-modal MRI Feature Fusion Block (MMFFB) module was used for information fusion to extract MRI information fusion features with complementary advantages. Then, the MRI omics features and dose omics features

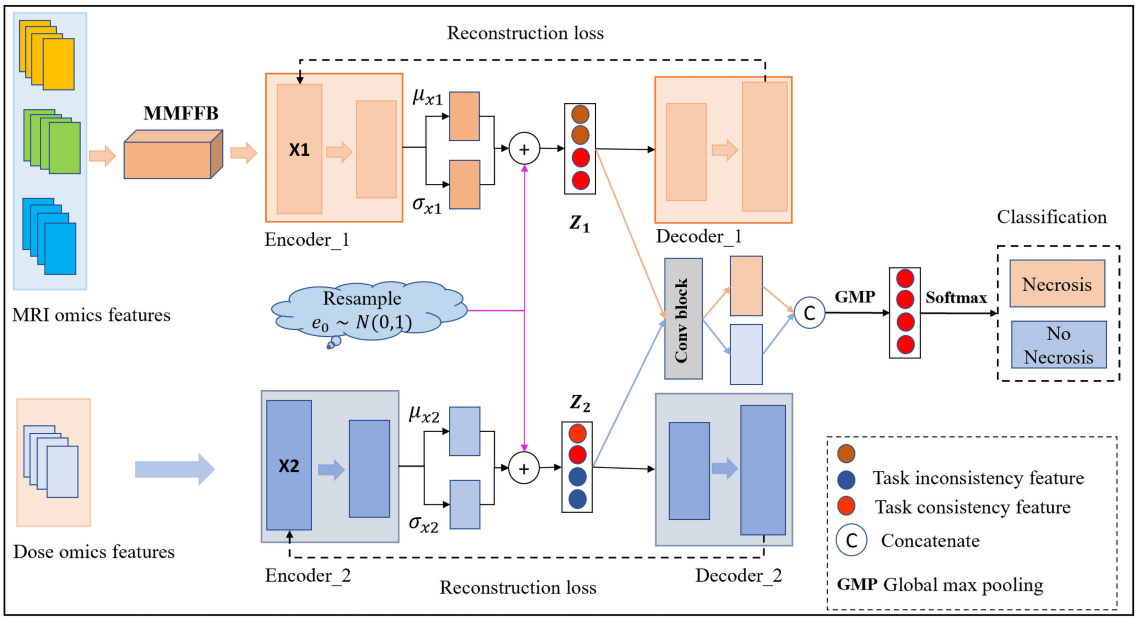

Fig. 2: The architecture of proposed model. The model had four inputs, of which the MMFFB module fused T1, T2, and T2C, and the fused features were output $Z_{mri}\rightarrow$ X1. The omics features were X2. On the one hand, the variational auto-encoder learns X1 and X2 feature distribution. The original features X1 and X2 are reconstructed by resampling feature embedding features Z1 and Z2. On the other hand, Z_1 and Z_2 can adaptively learn task consistency features for necrosis prediction. The two VAEs have the same structure. Z_1 and Z_2 will be globally pooled after a 1x1x10 convolution operation before splicing and finally sent to softmax for classification probability calculation.

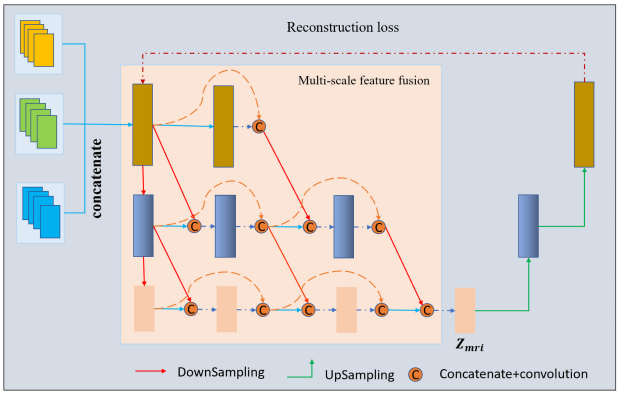

Fig. 3: The architecture of multi-modal MR feature fusion network. It is noteworthy that the input of this module is the MR features of three modes, and deep feature fusion is achieved by constructing a multi-scale dense connection network. The fused features are used to reconstruct the original input feature space and for subsequent necrosis prediction tasks.

are processed through the Variational Autoencoder (VAE) [44] module to obtain the latent space embedding. Latent space embedding may be regarded as low-dimensional manifolds of high-dimensional feature spaces. On the one hand, the decoder reconstructs two sets of latent space embedding to the input feature dimensions. On the other hand, a classifier was constructed by task consistency feature extraction and fusion, which was used as the main module of the model to predict the necrosis of recurrent NPC.

C. Multi-modal MRI Feature Fusion block(MMFFB)

For multi-modal MRI radiomics features, there are common and unique parts among different modes, and extracting both enables the fusion of the representative MRI modal features. Regarding disentanglement representation, different modal features are usually divided into common feature vectors and specific feature vectors in a fixed proportion [22]. By cross-fusing the commonality and distinct parts of different modes and reconstructing them to the corresponding input feature dimensions, objective optimization can be carried out, and the corresponding commonality and distinct parts can be learned. However, several things could be improved in this method. On the one hand, limiting the dimension of the common and distinct feature vectors may lead to incomplete expression of relevant information. On the other hand, it is challenging to construct a cross-fusion objective-constrained optimization method when handling a large number of nodes. We posit that there are feature representations in the three MR modal datasets that are conducive to the final classification goal and inconsistent feature representations that could be more conducive to the task. Therefore, Learning a potential feature representation can represent the three modes themselves well and shows good consistency for the classification task. As shown in Fig. 3, to obtain deep fusion of MR features, MMFFB block was used for latent space embedding features extraction. Specifically, An extended Dense U-shape module was defined to adaptively obtain multi-modal MR features and fusion. First, multi-modal MR features were concatenated and

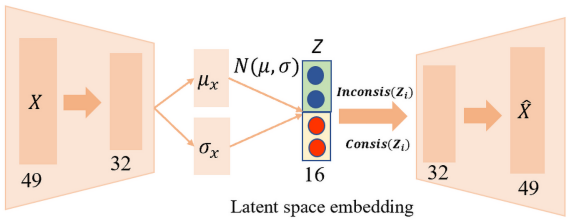

Fig. 4: The VAE latent space embedding module. X is the input feature vector, which generates hidden space embedding Z through the encoder, and then $\hat{X}$ is generated through the decoder.

passed by multi-scale dense blocks to generate the final fusion MRI feature, whose dimension is the same as each single modal data. Then, to maximize the MR feature information of the three modes, the fused features are reconstructed to the original input feature dimension by a decoder, and the self-supervised reconstruction loss constraint is utilized. At the same time, the final fused features are transmitted to the subsequent network for deep fusion with radiation dose features to participate in supervised classification optimization.

Specifically, we take as input MRI features from three different modalities, each with 49 dimensions. First, the three modal features are concatenated. The feature space is then mapped into a low-dimensional representation by twice down-sampling using a 1D convolution operation with a step size of two and a kernel of three. Use zero padding to keep the dimension of the feature space Z at 49. Referring to [28], better feature representation can be obtained by fusing features from the previous layer and the previous scale. Our specific operation is to perform a convolution operation after feature addition, with kernel size k=3, stride=1, and padding=1. In the decoder module, we recover the dimension of the feature space using transposed convolution. As shown in Fig. 3, we perform two downsampling and up-sampling operations. The low-dimensional embedding $\mathcal{Z}_{mri}$ is the fused multi-MRI feature used for both feature reconstruction and prediction of radiation necrosis in recurrent NPC re-treatment.

By constructing the self-supervised reconstruction loss-constrained feature fusion network, the special features and common features of each MR modal can be effectively extracted and deeply fused. The optimization objectives are as follows:

$$Loss_{mri\_recon} = \sum ||X_i - D_i||^2 \quad (1)$$

The minimum mean square error (MSE) is used to optimize the reconstruction objective of multi-modal MR features, where X_i is the feature after the concatenation of three modal MR omics features, and D_i is the reconstructed result of X_i based on the MMFFB module.

D. Latent Spacing Feature Embedding and Task Consistency Feature Representation

To fully use the complementary information of the features of each mode, we carry out task consistency abstraction in the following ways. Specifically, we project the feature vectors of each mode into the low-dimensional latent space to embed the encoding Z. Then, the latent variable Z is decomposed into Consis(Z_i) and Inconsis (Z_i). Consis(Z_i) represents the task consistency feature, and Inconsis (Z_i) represents the task inconsistency features intuitively, this classification process can be understood as the decomposing feature vectors of different modes into vector representations that are beneficial to the target task and vector representations that are detrimental to the target task. In addition, fixing the feature dimensions of Consis(Z_i) and Inconsis(Z_i) may lead to insufficient representation of the feature space of the target task. Therefore, we abstractly decompose the latent space embedding code Z into Consis(Z_i) and Inconsis(Z_i). We can implicitly learn the feature embedding representation of the latent space by optimizing the objective constraints. Therefore, for latent space feature embedding, the following requirements should be met:1) Consis(Z_1) and Consis(Z_2) complement each other as much as possible; 2) Adaptive decoupled latent space embedding Z_i can reconstruct input X well; 3) Consis(Z_1) and Consis(Z_2) can be fused well and increase the predictive power for the probability of necrosis.

We used two VAE to learn latent space feature embeddings. The encoder maps the input features into a Gaussian probability distribution and generates the latent space embedding Z by resampling, which the decoder uses to reconstruct the input features. The advantage of using VAE is that noise is added to the hidden space feature, so the feature reconstruction has an anti-noise ability. The disentanglement decomposition representation based on VAE is widely used [22], [35]. In our study, for radiomics features with different modalities, we first used a multi-layer perceptron neural network as an encoder to generate latent variables. Then, a multi-layer perceptron neural network is used as a decoder to reconstruct the input features. Finally, different from [22], the hidden space is directly divided into task consistency and task inconsistency according to the proportion of the number of features. We implicitly computed task consistency and inconsistency features by an adaptive combination of the latent space embedding Z_i of the two modalities. In our study, the number of neurons in each layer is based on the parameter setting, and the number of neurons in the other two layers of the encoder is set to 32 and 16, respectively. However, the number of task consistency features and task inconsistency features were not specified and obtained by adaptive learning. The module architecture is shown in Fig. 4. The module consists of the encoder, decoder, and latent space embedding. The encoder has two layers, and the number of neurons is 49 and 32, respectively. The decoder has two layers with 32 and 49 neurons, respectively, and the middle layer is a hidden space embedding layer with 16 neurons. The encoder learns the features of X through the network, namely the mean and standard deviation. The hidden space embedding is performed by randomly sampling under a standard normal distribution and reconstructing X through the decoder according to the original feature distribution using a re-parameterization technique. To formally represent the decomposability of the latent space embedding, we assume that the latent space can be intuitively decomposed into task consistency and inconsistency features, that is, the red

component and blue component of Z in Fig. 4. We use VAE reconstruction loss to constrain feature embedding with the optimization objectives:

$$Loss_{vae\_recon} = \sum_{i=1}^{2} ||X_i - D_i(Consis(E_i(X_i)), Inconsis(E_i(X_i)))||^2 + KL(q_\theta(Z_i|X_i)||p(Z_i)) \quad (2)$$

$Loss_{vae\_recon}$ represents different modal characteristics of the re-construction loss,E_i and D_i represent the ith encoder and decoder, respectively. $p(Z_i)$ is the prior imposed on the latent space embedding, $q_\theta(Z_i|X_i)$ is the encoder data distribution, KL is Kullback-Leible divergence.

The latent space features generated by VAE are not explicitly decomposed into two feature vectors to represent task consistency features and task inconsistency features. Therefore, we adopt the method of adaptive feature fusion to select the adaptive combination of latent space features of different modes to achieve the purpose of task consistency feature extraction and fusion. This stage aims to extract discriminative features from the latent space embedding of different modes. These characteristics may reinforce or contradict one another. Therefore, instead of the conventional method, where features are concatenated and directly sent to the classification layer, we adopt a multi-feature fusion layer processing step after feature concatenation. Specifically, we use a 1D convolution with kernel size 1x1x12 for the latent space embedding Z_1, Z_2. The features after the convolution operation are then concatenated. Finally, the concatenated feature vectors were pooled by global Max pooling(GMP) to obtain the final fusion feature, which is then sent to the classifier for classification.

To ensure that the adaptive fusion of latent space features promotes task consistency, we designed a multi-layer perceptron classifier, which takes the adaptively fused features as the input of supervised classification and leverages the supervision of supervised classification to determine the optimization direction of task consistency features. The supervised classification loss is defined as follows:

$$Loss_{cls} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M=2} Y_n^m \log \hat{Y}_n^m \quad (3)$$

Where M represent the number of categories, N represent the number of samples, Y indicates the sample label, and $\hat{Y}$ is the prediction result of the model.

Based on the two-stage multi-modal information fusion, an end-to-end prediction network is constructed. The overall learning objectives of the model are as follows:

$$Loss_{total} = \lambda_1(Loss_{cls}) + \lambda_2(Loss_{mri\_recon}) + \lambda_3(Loss_{vae\_recon}) \quad (4)$$

Where λ_1 is set to 1, $\lambda_2 = \lambda_3 = 0.5$.

III. EXPERIMENTS

A. Dataset and Experimental setting

1) Data preprocessing: Before model training, the input data underwent the following preprocessing: First, we performed N4 bias field correction on MR images to alleviate the interference caused by magnetic field deviation. Second, the omics features were extracted from different modes' MR and radiation dose data. Last, leveraging z-score normalization to eliminate dimensional inconsistency problems.

2) Experimental setting: Our study used dataset A to train and test the proposed method. Dataset B was used to compare the performance of the proposed model to the existing multi-modal fusion method. For model training, a total of 200 cases of data were used. The training set and test set were divided according to 5-fold cross-validation. That is, a total of five model training sessions were performed. Each training takes four-fifths of the total dataset as the training set, with the remainder being used as the test set. Following [45], the random forest method with a max_depth of 100 was used to coarsely remove the high-dimensional omics features on the training dataset before the training. The final feature number was 49 according to the feature weight coefficients ranking after random forest screening. As can be seen in dataset A in Table 1, the number of patients with necrosis was 80, and without necrosis was 120. Sample category imbalance was a problem needing to be addressed. Smote oversampling technique was used to augment data in order to alleviate the problem of category imbalance on the training set. Specifically, patients with necrosis were resampled to match the number of patients without necrosis. In the proposed network architecture, in the multi-modal MR image fusion stage, the feature variables that can simultaneously represent the three modes are extracted, and the number of features after the fusion of the three MR images is set to 49, which is used for subsequent and dose feature fusion. In the subsequent fusion stage of MR features and dose features, the number of feature-embedded variables of different modes was set to 16. The two kinds of features are embedded for adaptive fusion to form the final discriminative features. Our network model is based on PyTorch. In the training stage, a stochastic gradient descent algorithm is used to optimize the model, and the learning rate is 0.0001. Training epochs were set to 1000. All experiments were run on a XEON E5-2698 V4, GPU NVIDIA GeForce P6000 24GB machine. To demonstrate the effectiveness of the proposed method, we verify it from the following aspects: 1) To explore the effect of different combinations of multi-modal data on the prediction of radiotherapy necrosis in recurrent NPC.

2) To explore the performance improvement of multi-modal fusion over single-modal data, we use a variety of machine learning techniques to conduct extensive tests on four types of single-modal data while verifying the performance of the proposed method's single-modal verification form on the four types of modal data.

3) To verify the performance of the multi-modal fusion method proposed in this paper, we compare the proposed method with the following feature fusion methods: 1) EFM: Zhu et al. [46]; 2) DAAE: Hu et al. [47]; 3) HyperDense-Net (HDNET): Dolzde et al. [48]; 4) MLDRL: Yue et al. [22].

TABLE II: The performance of four modal data combination strategies on dataset A

modal combination	AUC	ACC	SEN	SPE
Dose+T1	0.890±0.045	0.8±0.176	0.692±0.742	0.615±0.593
Dose+T2	0.904±0.051	0.75±0.53	0.714±0.69	0.653±0.461
Dose+T1C	0.894±0.043	0.77±0.425	0.832±1.06	0.654±0.73
Dose+T1+T2+T1C	**0.936±0.028**	**0.85±0.145**	**0.857±0.58**	**0.692±0.73**

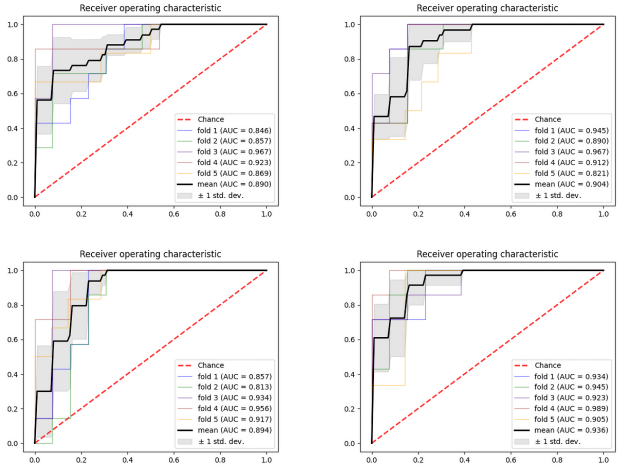

Fig. 5: The Receiver operating characteristic of the proposed method on multi-modal data(dataset A) fusion under different combination strategies is shown from left to right and from top to bottom: Dose+T1, Dose+T2, Dose+T1C, Dose+T1+T1C+T2. The figure shows the AUC indicators of each fold of the 5-fold cross-validation and the mean performance.

B. Evaluation Criterion

To better illustrate the predictive ability of the proposed method to predict necrosis for re-radiotherapy recurrent NPC, all experimental results were determined by 5-fold cross-validation, and we calculated the average AUC, classification accuracy (ACC), sensitivity (SEN), and specificity (SPE) to measure the performance of the proposed framework and other methods. The definitions of ACC, SEN, and SPE are as follows:

$$\begin{aligned} ACC &= \frac{TP+TN}{TP+FP+FN+TN} \\ SEN &= \frac{TP}{TP+FN} \\ SPE &= \frac{TN}{TN+FP} \end{aligned} \tag{5}$$

Where TP, FP, TN, FN are true positive, false positive, true negative, false negative, respectively.

C. Quantitative Results

Advantages of multi-modal information fusion: In the proposed network framework, four different data modes were fused in the following ways: Dose+T1, Dose+T1C, Dose+T2, and Dose+T1+T1C+T2. Table 2 shows the performance of the proposed method on four combinations of multi-modal data. Fusing all modal data achieves the best performance compared with the other three fusion strategies. ROC curves for each fold and the average ROC curve of the 5-fold set of the four combinations are shown in Fig. 5. The ROC analysis shows that the fusion of all modes has a better predictive ability. Specifically, the Dose+T1+T1C+T2 fusion method achieved the best AUC, ACC, SEN, and SPE, which were 0.936, 0.85, 0.857, and 0.692, respectively. Meanwhile, the other three combination strategies showed good performance. Therefore, the strategy based on the fusion of all modes demonstrated appreciable performance gains compared with the partial fusion method.

We conducted extensive experiments based on single-modal data to prove the effectiveness of the multi-modal information fusion proposed in this study. Since the proposed feature fusion method is based on a multi-modal scenario, the proposed model cannot be directly used for single-modal data. Therefore, we removed the multi-feature fusion part of the

TABLE III: The performance on single-modal data of several traditional methods(Dataset A).

method	metric	T1	T1C	T2	Dose
LR	AUC	0.533±0.039	0.538±0.038	0.614±0.065	0.756±0.065
	ACC	0.58±1.29	0.6±1.358	0.61±2.64	0.69±4.53
	SEN	0.442±1.93	0.476±2.50	0.467±1.71	0.587±1.98
	SPE	0.649±6.39	0.664±5.472	0.682±1.76	0.742±2.01
RF	AUC	0.695±0.047	0.698±0.0496	0.794±0.1106	0.750±0.09
	ACC	0.67±1.35	0.64±2.13	0.74±0.851	0.73±1.13
	SEN	0.404±1.77	0.38±3.23	0.576±1.59	0.51±2.07
	SPE	0.803±3.43	0.77±4.14	0.817±1.85	0.83±0.97
SVM	AUC	0.537±0.079	0.547±0.057	0.672±0.1	0.780±0.107
	ACC	0.58±1.27	0.57±1.78	0.63±1.829	0.78±1.593
	SEN	0.3±1.52	0.271±2.34	0.25±4.75	0.609±0.813
	SPE	0.727±0.79	0.728±3.41	0.817±0.83	0.862±1.77
Adaboost	AUC	0.710±0.065	0.705±0.059	0.746±0.0838	0.733±0.093
	ACC	0.69±2.13	0.63±0.84	0.69±0.39	0.73±1.21
	SEN	0.357±0.09	0.323±0.13	0.466±0.413	0.433±0.329
	SPE	0.863±1.251	0.787±0.491	0.802±0.463	0.876±1.14
Ours	AUC	0.551±0.03	0.566±0.043	0.734±0.119	0.779±0.04
	ACC	0.55±0.251	0.59±0.412	0.66±0.736	0.76±0.415
	SEN	0.495±0.23	0.495±0.18	0.638±1.09	0.667±0.085
	SPE	0.573±1.26	0.636±1.13	0.685±0.921	0.801±0.61

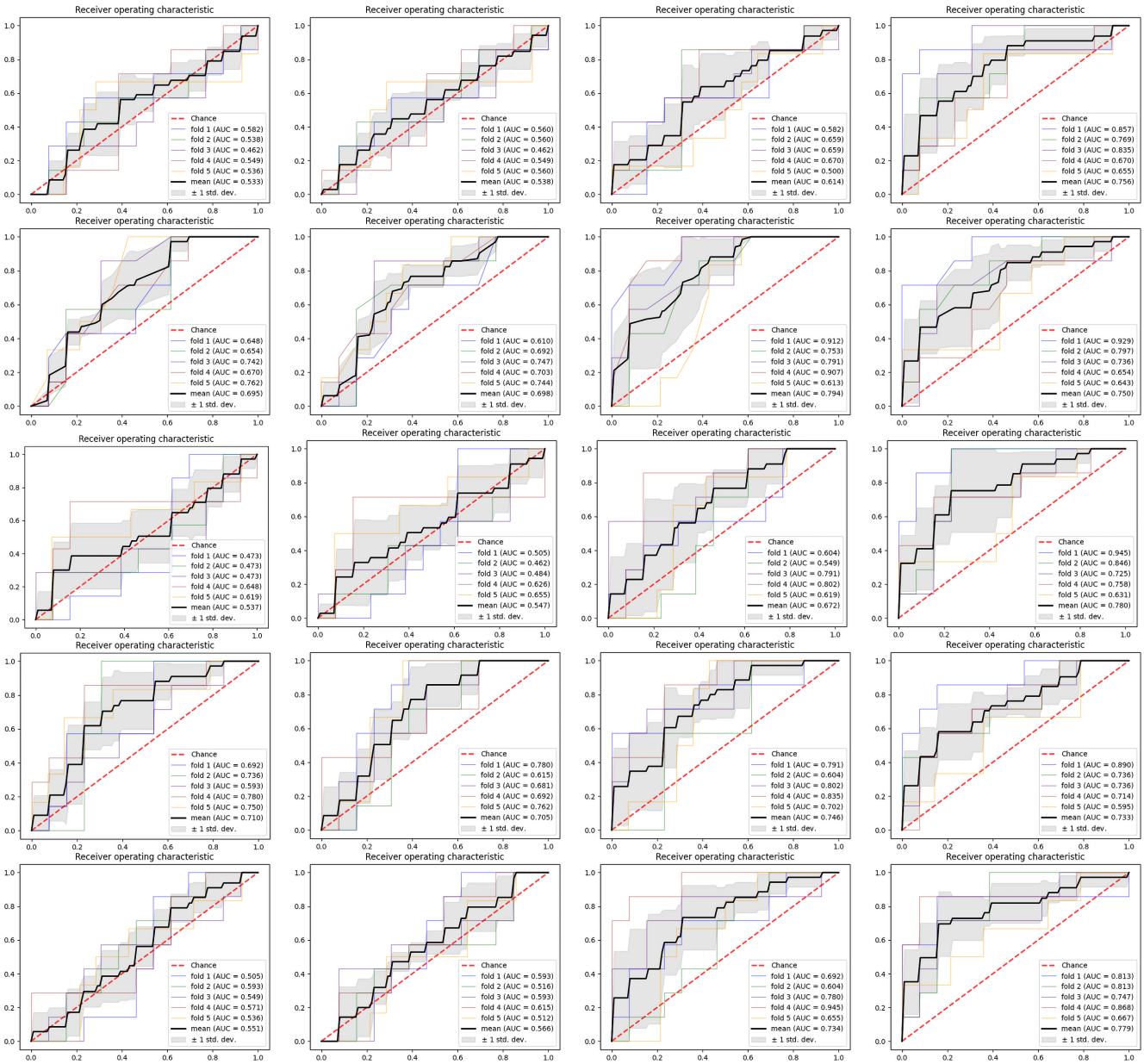

Fig. 6: The test results of the single modal data(dataset A) on different methods, each row represents a classification method, in order: LR, RF, SVM, Adaboost and Ours. Each column represents a data mode: T1, T1C, T2, Dose, respectively.

TABLE IV: The performance comparison of different multi-modal feature fusion models(dataset A).

method	AUC	ACC	SEN	SPE
EFM [46]	0.88±0.69	0.71±0.764	0.73±2.524	0.5±3.46
DAAE [47]	0.864±0.417	0.77±0.551	0.771±1.58	0.492±1.43
HDNet [48]	0.877±0.124	0.78±0.235	0.862±0.415	0.429±1.526
MLDRL [22]	0.925±0.073	0.82±0.174	**0.865±0.32**	0.593±0.941
Ours	**0.936±0.028**	**0.85±0.145**	0.857±0.58	**0.692±0.73**

model and directly trained the supervised classification model on the single-modal features. To eliminate the deviation by model differences and highlight the importance of multi-modal fusion. We used multiple traditional machine learning methods to experiment with single-modal data. The classifiers used include logistic regression (LR), Random Forest (RF), Support Vector Machine (SVM), and Ensemble Model (Adaboost). As listed in Table 3, SVM showed the best performance in dose data, with AUC=0.78 and ACC=0.78. T1 and T1C modal data showed low predictive performance with each classifier, with the lowest AUC of 0.533. Compared with T1 and T1C, T2 and dose modal data have a higher predictive potential. As shown in Fig. 6, dose and T2 achieved good results in numerous prediction methods for single modal data, while T1 and T1C performed poorly. In addition, the proposed method achieved similar performance compared to other traditional single-modal data validation methods. However, after multi-modal fusion, the AUC of the proposed method was 0.936,

TABLE V: Model performance comparison on the supplementary dataset (dataset B).

method	AUC	ACC	SEN	SPE
EFM [46]	0.75±0.562	0.65±0.743	0.64±1.582	0.52±2.42
DAAE [47]	0.768±0.631	0.643±0.945	0.721±1.471	0.54±1.983
HDNet [48]	0.77±0.751	0.67±0.812	0.754±1.782	0.429±2.41
MLDRL [22]	0.872±0.43	0.78±0.652	**0.823±0.842**	0.593±1.74
Ours	**0.893±0.24**	**0.794±0.67**	0.78±0.96	**0.645±1.32**

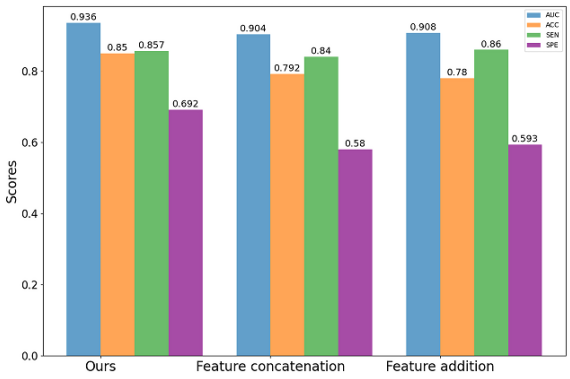

Fig. 7: Performance comparison of different fusion methods on dataset A.

which was higher than the AUC of the single-modal data, which was 0.78. These comparisons imply that multi-modal information fusion played an essential role in the study of re-radiotherapy induced necrosis in recurrent NPC.

Comparison with other fusion methods: This section introduces the horizontal comparison between the proposed method and other fusion methods. The data in this paper have four modalities, and to highlight the effectiveness of dose features, it is necessary to treat multi-sequence MRI image data as a single self-contained modality. Therefore, we add the multi-modal MRI data into a single self-contained modality. We feed the added MRI data and the dose data into the network to compare the existing methods. As can be seen from Table 4, EFM [46], D-AAE [47], HDNET [48], and MLDRL [22] all show competitive performance for multi-modal data fusion, with AUC of 0.88, 0.864, 0.877 and 0.925 respectively. However, our method achieved better performance. Specifically, the AUC (0.936) and ACC (0.85) are better than the other fusion methods. Good performance is also achieved for SPE and SEN, with SEN slightly lower than MLDRL [22]. To further illustrate the performance of the proposed method, we also performed supplementary experiments on the dataset B. As shown in Table 5, our method achieves the best results on the leading metrics, such as AUC and ACC, and other metrics also achieve competitive results. These results indicate that our developed method has increased performance over other commonly used multi-modal feature fusion methods in predicting necrosis in recurrent NPC.

D. Ablation Study

1) Different Components Analysis: In this section, we explore the importance of multi-modal MRI reconstruction loss, VAE reconstruction loss, and supervised classification loss on the prediction of NPC necrosis. These experiments were performed on each component using 5-fold cross-validation on dataset A. As we can see from Table 6, adding different reconstruction losses can improve the model performance compared with only using supervised classification loss. When only using supervised classification loss ($Loss_{cls}$), AUC and ACC are 0.878 and 0.75, respectively; Using $Loss_{cls}$ and $Loss_{mri\_recon}$, AUC and ACC improve by a point to 0.887 and 0.76, respectively; Use $Loss_{cls}$ and $Loss_{vae\_recon}$ achieves similar performance as the second combination method, with an AUC of 0.881, ACC 0.8; The three loss optimization methods together achieved the best performance AUC (0.936) and ACC(0.85). The above results show that these loss functions are practical for target classification tasks. In addition, combining three losses gives the best performance compared to two or single losses. Based on the above analysis, the three loss combinations designed in this study are effective.

2) Significance for Adaptive to learn Task consistency features: In this section, we experiment with several different fusion modalities to show the effect of adaptive task-consistent feature fusion. Three ways of feature fusion are used in the experiments on dataset A through 5-fold cross-validation. 1. The features Z_1 and Z_2 are directly concatenated and connected to a fully connected layer, and the probability is predicted by softmax output. 2. The features Z_1 and Z_2 are directly added into a fully connected layer, and the probability is predicted by softmax activate function output. 3. The adaptive task-consistent feature fusion method proposed in this paper is used. We compare four performance metrics, AUC, ACC, SEN, and SPE, as shown in Fig. 7. As in the figure, similar performance is achieved using direct addition or concatenation of the features. The best performance was achieved using the proposed adaptive task-consistent feature fusion method. Therefore, the proposed task consistency feature is more advantageous than the direct feature fusion methods.

IV. DISCUSSION

This work discusses multi-modal data's predictive effect on recurrent NPC's radiation necrosis after retreatment. In the clinic, radiotherapy is a critical factor of radiotherapy necrosis, so the fusion of radiotherapy dose modality data and MRI multi-modal image data have practical clinical significance. On the one hand, the multi-modal MRI image data reflects the tumor's shape, size, texture, and other related attributes. On the other hand, the dose data reflect the irradiation dose of different tumor regions. Therefore, potential correlations between these two parameters can be found if they are implicitly modeled jointly by fusing multi-modal MRI data and dose data. In addition, the proposed model is mainly used to explore the correlation between multi-modal data fusion and radiation necrosis in recurrent NPC retreatment. Implicitly, the two were combined to model, proving that the multi-modal data

TABLE VI: Ablation study for different components on dataset A.

$Loss_{mri\_recon}$	$Loss_{vae\_recon}$	$Loss_{cls}$	AUC	ACC	SEN	SPE
✗	✗	✓	0.878±0.42	0.75±0.474	0.571±1.283	0.462±1.514
✓	✗	✓	0.878±0.36	0.76±0.357	0.635±0.87	0.650±0.76
✗	✓	✓	0.818±0.132	0.8±0.272	0.686±0.63	0.571±0.97
✓	✓	✓	**0.936±0.028**	**0.85±0.145**	**0.857±0.58**	**0.692±0.73**

had positive significance for predicting radiotherapy induced necrosis in recurrent NPC.

To forecast recurrent necrosis in recurrent NPC for early clinical intervention, we proposed an effective multi-modal information fusion model for predicting necrosis in recurrent NPC after re-irradiation. This model is the first to fuse multi-modal MR image and three-dimensional radiation dose data to predict necrosis in recurrent NPC. To verify the superiority of the proposed method, we discuss the performance results of the model on different modal data. In addition, we evaluated different machine learning methods and analyzed the performance of some existing feature fusion methods for necrosis prediction. Finally, ablation experiments were performed to evaluate the effectiveness of each component.

We defined a two-stage strategy for effective feature fusion of multiple modal data, including a Dense U-shape and VAE module. The main idea is to decompose the features of different modal through self-supervised reconstruction loss and supervised classification loss, that is, to fuse the features of different modal of task consistency to classification prediction. The self-supervised reconstruction loss is mainly used to explore the shared and complementary information in features, and the supervised classification loss is mainly used for task consistency feature fusion.

Experiment show that our multi-modal feature fusion network can effectively utilize the complementary information of different modal features, and the self-supervised reconstruction loss can be used as a supplement to promote the learning of classification models.

In our study, an adaptive fusion strategy was adopted to automatically learn feature combination weights to make the final imaging and radiation dose features more task consistent. Specifically, a 1D convolutional neural network is used to learn the combined parameters, promote the effective feature fusion of different modal, and implicitly carry out consistent feature extraction. Compared with the existing machine learning and multi-modal fusion methods, the characteristic of our method is that it uses different fusion strategies to deal with the data of different modal rather than directly concatenating features. In addition, we applied radiomics methods to radiation dose data; At the same time, the network model is optimized with multiple constrained objectives.

Although our proposed method has achieved good performance and ultimately enriches the research content in the necrosis prediction of re-radiotherapy for recurrent NPC, there are some shortfalls. Firstly, Showing that the model has good generalization is not easy due to the need for more sample data. So additional data are needed to verify the power of the developed model. Secondly, this method manually extracts omics features as the input of deep learning, and the feature selection method is based on traditional machine learning, which fails to build a complete end-to-end deep learning network, and there may be bias during feature selection. Finally, our study was only used to predict the probability of necrosis and did not provide other prognostic indicators. In the future, we aim to collect richer multicenter data, expand the experimental scope, conduct more predictive analysis through this study, and explore the interrelationship between multi-tasks.

V. CONCLUSION

In this study, we propose a multi-modal feature fusion network model for patients of recurrent NPC to accurately forecast the probability of re-irradiation induced necrosis. Specifically, we define self-supervised reconstruction loss to extract the general and specific characteristics of different MR modal information. We used self-supervised construction and classification loss to conduct latent space feature embedding of radiomics and dose omics to extract task consistency characteristics. At last, the model learns the combination mode of different modal through the self-adaption feature fusion method. The experimental results indicate that the proposed method has advantages in necrosis forecast compared to existing forecast methods and multi-modal data fusion methods. Our method can accurately forecast the necrosis probability and be a potential index for personal radiotherapy of recurrent nasopharyngeal carcinoma.

ETHICS STATEMENT

This study was approved by the Institutional Review Board (IRB) of Xiangya Hospital with Approval No. 202207166. According to the relevant guidelines and regulations of the retrospective study, the requirement for informed consent was waived.

REFERENCES

[1] L. Zhong et al., "A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0," *Radiother Oncol*. 2020, vol. 151, pp. 1–9.

[2] L. Tang et al., "Global trends in incidence and mortality of nasopharyngeal carcinoma," *Cancer Letters*., vol. 374, no. 1, pp. 22–30, 2016.

[3] Al-Sarraf, M et al., "Chemoradiotherapy versus radiotherapy in patients with advanced nasopharyngeal cancer: phase III randomized Intergroup study 0099," *J Clin Oncol*., vol. 16, no. 4, pp. 1310–1317, 1998.

[4] J. Wee et al.,"Randomized Trial of Radiotherapy Versus Concurrent Chemoradiotherapy Followed by Adjuvant Chemotherapy in Patients With American Joint Committee on Cancer/International Union Against Cancer Stage III and IV Nasopharyngeal Cancer of the Endemic Variety," *J Clin Oncol*., vol. 23, no. 27, pp. 6370–6378, 2005.

[5] M. K., Kam et al., "Treatment of nasopharyngeal carcinoma with intensity-modulated radiotherapy: the Hong Kong experience," *Int J Radiat Oncol Biol Phys* ., pp. 1440–1450, 2004.

[6] A. W., Lee et al., "Treatment results for nasopharyngeal carcinoma in the modern era: the Hong Kong experience," *Int J Radiat Oncol Biol Phys*., vol. 61, no. 4, pp. 1107–1116, 2004.

[7] A. W., Lee et al., "Retrospective analysis of patients with nasopharyngeal carcinoma treated during 1976-1985: survival after local recurrence," *Int J Radiat Oncol Biol Phys*., vol. 26, no. 5, pp. 773–782, 1993.

[8] K. H. Yu et al., "Survival outcome of patients with nasopharyngeal carcinoma with first local failure: a study by the Hong Kong Nasopharyngeal Carcinoma Study Group," *Head Neck*., vol. 27, no. 5, pp. 397–405, 2005.

[9] H. Chen et al., "Effectiveness and toxicities of intensity-modulated radiotherapy for patients with locally recurrent nasopharyngeal carcinoma," *PloS one*.,8(9), e73918, 2013.

[10] A. S. Ho et al., "argeted endoscopic salvage nasopharyngectomy for recurrent nasopharyngeal carcinoma," *Int Forum Allergy Rhinol*., vol. 2, no. 2, pp. 166–173, 2012.

[11] D. Dong et al., "Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer," *Ann Oncol*., vol. 30, no. 3, pp. 431–438, 2019.

[12] W. L. Bi et al., "Artificial intelligence in cancer imaging: Clinical challenges and applications," *CA Cancer J Clin*., vol. 69, no. 2, pp. 127–157, 2019.

[13] L. Zhong et al., "A deep learning-based radiomic nomogram for prognosis and treatment decision in advanced nasopharyngeal carcinoma: A multicentre study," *EBioMedicine*.,70:103522, 2021.

[14] H. Peng et al., "Prognostic Value of Deep Learning PET/CT-Based Radiomics: Potential Role for Future Individual Induction Chemotherapy in Advanced Nasopharyngeal Carcinoma," *Clin Cancer Res*., vol. 25, no. 14, pp. 4271–4279, 2019.

[15] L. Z. Zhong et al., "A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0," *Radiother Oncol*., 151:1-9, 2020.

[16] D. Dong et al., "Development and validation of a novel MR imaging predictor of response to induction chemotherapy in locoregionally advanced nasopharyngeal cancer: a randomized controlled trial substudy (NCT01245959)," *BMC Med*., 17(1):190, 2019.

[17] T. Liu et al., "Deep Learning and Medical Image Analysis for COVID-19 Diagnosis and Prediction," *Annu Rev Biomed Eng*., vol. 24, no. 1, pp. 179–201, 2022.

[18] A. Khvostikov et al., "3D CNN-based classification using sMRI and MD-DTI images for Alzheimer disease studies," Available: arXiv. https://doi.org/10.48550/arXiv.1801.05968

[19] Q. Yang et al., "Automatic T Staging Using Weakly Supervised Deep Learning for Nasopharyngeal Carcinoma on MR Images," *J Magn Reson Imaging*., vol. 52, no. 4, pp. 1074–1082, 2020.

[20] B. Jing et al., "Deep learning for risk prediction in patients with nasopharyngeal carcinoma using multi-parametric MRIs," *Comput Methods Programs Biomed*., 197:105684, 2020.

[21] K. Liu et al., "Deep learning pathological microscopic features in endemic nasopharyngeal cancer: Prognostic value and protentional role for individual induction chemotherapy," *Cancer Med*., vol. 9, no. 4, pp. 1298–1306, 2020.

[22] H. Yue et al., "MLDRL: Multi-loss disentangled representation learning for predicting esophageal cancer response to neoadjuvant chemoradiotherapy using longitudinal CT images," *Med Image Anal*., 79:102423, 2022.

[23] Y. H. Yu 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*.,35(1):59, 2016.

[24] Lao, J., Chen, Y., Li, ZC. et al., "A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme," Scientific reports 7.1 (2017): 10353.

[25] Zheng, X., Yao, Z., Huang, Y. et al., "Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer," Nat Commun 11, 1236 (2020).

[26] Hu, Zongsheng et al. "A radiomics-boosted deep-learning model for COVID-19 and non-COVID-19 pneumonia classification using chest x-ray images," Medical physics vol. 49,5: 3213-3222, 2022.

[27] A. Wu et al., "Dosiomics improves prediction of locoregional recurrence for intensity modulated radiotherapy treated head and neck cancer cases," *Oral Oncol*., 104:104625, 2020.

[28] Y. Xu et al., "Hierarchical Multiscale Dense Networks for Intelligent Fault Diagnosis of Electromechanical Systems," in *IEEE Transactions on Instrumentation and Measurement*., vol. 71, pp. 1-12, 2022.

[29] Bo, L.; Zhang, Z.; Jiang, Z.; Yang, C.; Huang, P.; Chen, T.; Wang, Y.; Yu, G.; Tan, X.; Cheng, Q.; et al. Differentiation of Brain Abscess from Cystic Glioma Using Conventional MRI Based on Deep Transfer Learning Features and Hand-Crafted Radiomics Features. Front. Med. 2021, 8, 748144.

[30] Liu Z, Jiang Z, Meng L, Yang J, Liu Y, Zhang Y, et al. Handcrafted and deep learning-based radiomic models can distinguish gbm from brain metastasis. Journal of Oncology 2021 (2021).

[31] Y. Zhang, Y. Zhang, W. Guo, X. Cai and X. Yuan, "Learning Disentangled Representation for Multimodal Cross-Domain Sentiment Analysis," in IEEE Transactions on Neural Networks and Learning Systems, doi: 10.1109/TNNLS.2022.3147546.

[32] J. Ouyang, E. Adeli, K. M. Pohl, Q. Zhao, and G. Zaharchuk, "Representation disentanglement for multi-modal brain MRI analysis," in Proc. IPMI, pp. 321–333, 2021.

[33] A. Wu et al., "Robustness comparative study of dose–volume–histogram prediction models for knowledge-based radiotherapy treatment planning," *Journal of Radiation Reasearch and Applied Sciences*., pp. 390-397, 2019.

[34] Y. Guo et al., "Deep Disentangled Representation Learning of Pet Images for Lymphoma Outcome Prediction," in *Proc. IEEE Conf. International Symposium on Biomedical Imaging (ISBI)*., 2020.

[35] T. Q. Chen et al., "Isolating Sources of Disentanglement in Variational Autoencoders," in *Proc. Advances in Neural Information Processing Systems*., 2018.

[36] Y. Hua et al., "Long-term treatment outcome of recurrent nasopharyngeal carcinoma treated with salvage intensity modulated radiotherapy," European journal of cancer 48.18 (2012): 3422-3428.

[37] M. Chen et al., "Clinical findings and imaging features of 67 nasopharyngeal carcinoma patients with postradiation nasopharyngeal necrosis," *Chin J Cancer*., vol. 32, no. 10, pp. 533–538, 2013.

[38] X. Huang et al., "Diagnosis and management of skull base osteoradionecrosis after radiotherapy for nasopharyngeal carcinoma," *Laryngoscope*., vol. 116, no. 9, pp. 1626–1631, 2006.

[39] Y. Hua et al., "Postradiation nasopharyngeal necrosis in the patients with nasopharyngeal carcinoma," *Head Neck*., vol. 31, no. 6, pp. 807–812, 2009.

[40] S. Chin et al., "Necrotic nasopharyngeal mucosa: an ominous MR sign of a carotid artery pseudoaneurysm," *AJNR Am J Neuroradiol*., vol. 26, no. 2, pp. 414–416, 2005.

[41] R. E. Marx et al., "A new concept in the treatment of osteoradionecrosis," *J Oral Maxillofac Surg*., vol. 41, no. 6, pp. 351–357, 1983.

[42] A. W. Lee et al., "Retrospective analysis of nasopharyngeal carcinoma treated during 1976-1985: late complications following megavoltage irradiation," *Br J Radiol*., vol. 65, no. 778, pp. 918–928, 1992.

[43] J. M. Bedwinek et al., "Osteonecrosis in patients treated with definitive radiotherapy for squamous cell carcinomas of the oral cavity and naso- and oropharynx," *Radiology*., vol. 119, no. 3, pp. 665–667, 1976.

[44] C. Doersch, "Tutorial on variational autoencoders," arXiv preprint arXiv:1606.05908, 2016.

[45] Kawakubo, H., Yoshida, H., "Rapid feature selection based on random forests for high-dimensional data," Expert Syst. Appl 40, 6241–6252,2012.

[46] X. Zhu et al., "Canonical feature selection for joint regression and multi-class identification in Alzheimer's disease diagnosis," *Brain Imaging Behav*., vol. 10, no. 3, pp. 818–828, 2016.

[47] D. Hu et al., "Disentangled-Multimodal Adversarial Autoencoder: Application to Infant Age Prediction With Incomplete Multimodal Neuroimages," *IEEE Trans Med Imaging*., vol. 39, no. 12, pp. 4137–4149, 2020.

[48] J. Dolz et al., "HyperDense-Net: A Hyper-Densely Connected CNN for Multi-Modal Image Segmentation," *IEEE Trans Med Imaging*., vol. 38, no. 5, pp. 1116–1126, 2019.