

# RadKAM: Attention-Driven Kolmogorov-Arnold Model for Automatic Radiation-Induced Lymphopenia Prediction by Multimodal Learning

Rongchang Zhao<sup>1</sup>, Zhangyue Wu<sup>1</sup>, Jian Zhang<sup>1</sup>, Zijian Zhang<sup>2(✉)</sup>, and Shuo Li<sup>3</sup>

<sup>1</sup> School of Computer Science and Engineering, Central South University, Changsha, China

<sup>2</sup> Department of Oncology, National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, China  
wanzzj@csu.edu.cn

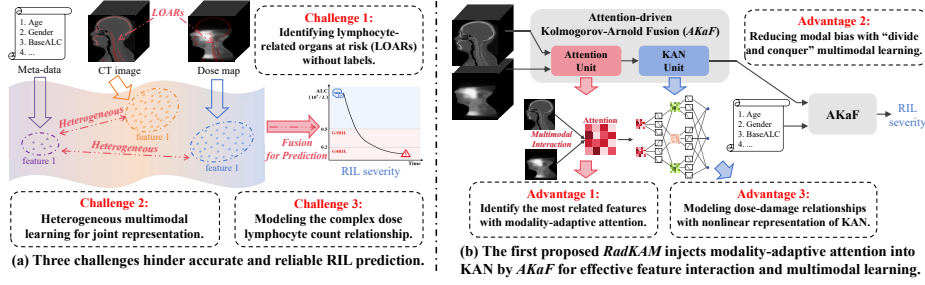
<sup>3</sup> Department of Computer and Data Science and Department of Biomedical Engineering, Case Western Reserve University, Cleveland, USA

**Abstract.** Accurate prediction of radiation-induced lymphopenia (RIL), a common complication of radiation therapy (RT), is clinically crucial for ensuring the safety of cancer treatment. However, accurately predicting RIL before RT is highly challenging due to the complexity of immune damage and various input data. In this study, we propose a novel multimodal learning framework named **RadKAM** to predict RIL severity using heterogeneous data, including CT images, dose maps, and meta-data. The proposed RadKAM leverages a “divide and conquer” strategy to learn the multimodal representation and model the dose-damage relationship for RIL prediction in an end-to-end framework. For the first time, an Attention-driven Kolmogorov-Arnold Fusion (**AKaF**) scheme is designed by injecting modality-adaptive attention into KAN for intra- and inter-modality interactions. Specifically, RadKAM is constructed with Multimodal Interactive AKaF (MI-AKaF) and Cross-modality Guided AKaF (CG-AKaF) to capture features related to lymphocyte-related organs, and model the dose-damage relationships by multimodal feature interactions. By leveraging the advantages of nonlinear representation, RadKAM effectively models the complex interactions of heterogeneous multimodal data, resulting in a comprehensive representation for RIL prediction. Extensive experiments validate the effectiveness of the proposed RadKAM framework, demonstrating its ability to accurately predict RIL severity through multimodal learning.

**Keywords:** Radiation-induced Lymphopenia · Multimodal Learning · Kolmogorov-Arnold Network.

## 1 Introduction

Accurate prediction of radiation-induced lymphopenia (RIL) before the radiation therapy (RT) is clinically crucial for the safe treatment of nasopharyngeal



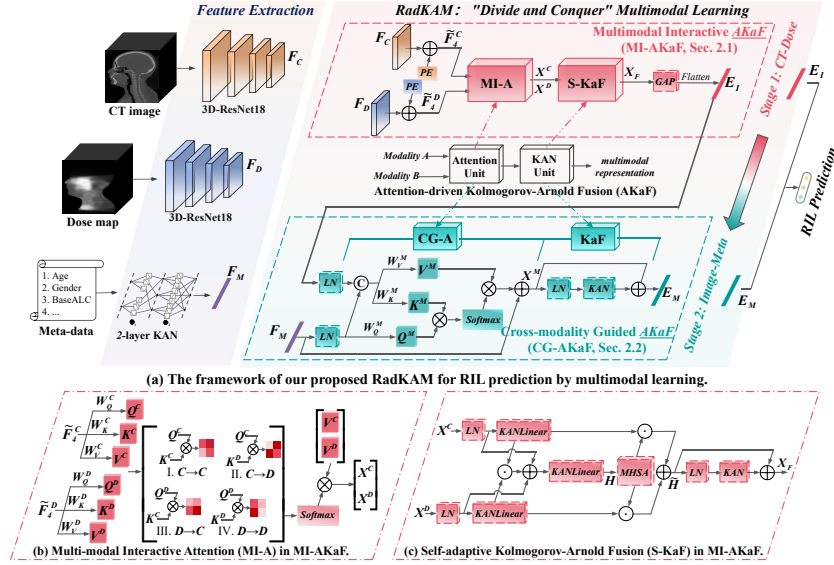
**Fig. 1.** (a) The challenges in automatic RIL prediction. (b) For the first time, the proposed RadKAM effectively injects modality-adaptive attention into KAN by AKaF scheme to learn the multimodal feature interactions for complex dose-damage relationship modeling.

carcinoma (NPC) [4, 5, 17]. RIL, marked by a significant decline in the absolute lymphocyte count (ALC) after radiation exposure, is a common and typically unavoidable complication of RT [2]. Research has consistently shown that patients with head and neck cancers, including NPC, are susceptible to RIL [4, 13]. Importantly, severe RIL has been linked to poorer overall survival, progression-free survival, and distant metastasis-free survival in patients [4, 10, 15, 17].

Predicting RIL for NPC treatment is challenging due to three issues (Fig. 1a): **1)** Identifying lymphocyte-related organs at risk (LOARs) [5, 21] without labels. Accurately identifying LOARs that are particularly sensitive to RIL severity is critical, however, the ground truth is no consensus in clinical settings. **2)** Heterogeneous multimodal learning for joint representation. Accurate RIL prediction often involves multimodal data (CT image, dose maps, meta-data), which exhibit significant heterogeneity and pose a greater hurdle for the unified representation of multimodal data. **3)** Modeling the complex dose-damage relationship related to ALC decrease. Several factors such as radiation dose maps, lymphocyte dynamics, heterogeneity of tissue response and individual differences, become critical impediments to ensuring accurate and reliable RIL prediction [2, 17, 15, 21].

Although notable progress has been made to predict RT complications [11, 18, 22] and RIL severity [5, 19], there are some drawbacks: **1)** Priors bias. Existing approaches [5, 19] are subject to manual designs involving complex steps and biased priors. **2)** Inefficient multimodal learning. Existing methods are trapped in inefficient learning of heterogeneous multimodal data, resulting in biased representation towards specific modalities. **3)** Linear dose-damage relationship. Existing methods often model the dose-damage relationship with a linear radiobiological model or exponential function, struggling with modeling complex, long-term dependencies of the dose-lymphocyte count relationship.

The Kolmogorov-Arnold Networks (KANs) [9] emerge as a potential alternative for modeling complex relationships with nonlinear fitting capabilities. It stems from the Kolmogorov-Arnold representation theorem, which posits that any multivariate function can be expressed as a finite composition of univariate



**Fig. 2.** The RadKAM is a novel “divide and conquer” multimodal learning framework to capture the comprehensive representation of multimodal data (CT image, dose map, meta-data) for RIL prediction. RadKAM incorporates modality-adaptive attention into KAN to form two novel components: Multimodal Interactive AKaF (MI-AKaF) for lymphocyte-related organs-aware features and Cross-modality Guided AKaF (CG-AKaF) for comprehensive representation in dose-damage relationships.

functions and addition. Although KAN shows promising performance in medical image representation [3, 8, 16], incorporating KAN into multimodal learning for RIL prediction presents three challenges: **1)** KAN cannot handle the “where” issue to find which organs or regions are closely related to RIL. **2)** Combining KAN with Transformer suffers from inefficient multimodal learning due to computation catastrophe and limited modeling of multimodal interactions. **3)** KAN leads to modality-biased learning stemming from serious intrinsic modality differences (such as CT images and meta-data).

To address the challenges, we proposed RadKAM (Fig. 1b), a novel Attention-driven Kolmogorov-Arnold Model, to achieve accurate RIL prediction by multimodal learning. RadKAM innovatively injects modality-adaptive attention into KAN by Attention-driven Kolmogorov-Arnold Fusion (AKaF) scheme to learn the comprehensive representation of CT images, dose maps, and meta-data. This approach follows the “divide and conquer” multimodal learning strategy to effectively identify the lymphocyte-related organs-aware features, and then integrate those features into a comprehensive representation for dose-damage relationship modeling. Leveraging long-term feature interactions and nonlinear representation capabilities, RadKAM models the complex relationship between multimodal features and ALC decline, enabling accurate and reliable RIL prediction.

Our main contributions are threefold: 1) For the first time, an effective multimodal learning framework **RadKAM** is proposed to achieve accurate RIL prediction before RT. 2) A novel attention-driven KAN scheme, **AKaF**, is proposed by injecting modality-adaptive attention into KAN to learn the inter- and intra-modality feature interactions in multimodal learning. 3) Two modality-specific blocks (MI-AKaF and CG-AKaF) are well-designed with the “divide and conquer” strategy to develop heterogeneous multimodal data for RIL prediction.

## 2 RadKAM

RadKAM (Fig. 2) is proposed to learn the comprehensive multimodal representation for RIL prediction by injecting modality-adaptive attention into KAN. Conceptually, the core design is the **Attention-driven Kolmogorov-Arnold Fusion (AKaF)** scheme, which leverages flexible attention computing to identify the most related features and KA theorem to model the complex dose-damage relationships for RIL prediction. RadKAM fuses CT images, dose maps, and meta-data into a comprehensive multimodal representation for the RIL prediction with the “divide and conquer” multimodal learning strategy, relying on two specialized AKaF variants: **1) Multimodal Interactive AKaF (MI-AKaF)** to learn the lymphocyte-related organs-aware features for modeling the dose-LOARs correspondence with image modality data (CT images and dose maps); **2) Cross-modality Guided AKaF (CG-AKaF)** to capture the joint representation incorporating individual data (meta-data) for RIL prediction. Finally, the representation is passed into a classification layer to predict RIL severity.

### 2.1 Multimodal Interactive AKaF (MI-AKaF) for lymphocyte-related organs-aware features

MI-AKaF is designed to capture lymphocyte-related organs-aware features by modeling the long-term interaction between CT images and dose maps. MI-AKaF is composed of two modules, **Multimodal Interactive Attention (MI-A)** and **Self-adaptive Kolmogorov-Arnold Fusion (S-KaF)**. Given the input CT image feature  $\mathbf{F}_C \in \mathbb{R}^{c \times h \times w \times d}$  and dose map feature  $\mathbf{F}_D \in \mathbb{R}^{c \times h \times w \times d}$ , position embeddings (PE) [14] is added to get  $\tilde{\mathbf{F}}_C, \tilde{\mathbf{F}}_D \in \mathbb{R}^{d_s \times c}$ , where  $d_s = hwd$ . PE allows MI-AKaF to contain richer spatial position information.  $\tilde{\mathbf{F}}_C$  and  $\tilde{\mathbf{F}}_D$  are inputted into MI-A to identify the most lymphocyte-relevant features through multimodal attention interaction, and are then passed into S-KaF to fuse these features into a joint representation. Finally, MI-AKaF performs the global average pooling (GAP) to fuse the feature channel as  $\mathbf{E}_I \in \mathbb{R}^c$ .

**Multimodal Interactive Attention (MI-A, Fig. 2b).** The MI-A is a modality-specific attention computing unit that models both intra- and inter-modal interactions of CT image features and dose map features, thereby enhancing the comprehensive features of lymphocyte-related organs associated with dose distribution. Specifically, given the query  $\mathbf{Q}_C$ , key  $\mathbf{K}_C$ , and value  $\mathbf{V}_C$  from  $\tilde{\mathbf{F}}_C$

and the query  $\mathbf{Q}_D$ , key  $\mathbf{K}_D$ , and value  $\mathbf{V}_D$  from  $\tilde{\mathbf{F}}_D$ . We can model the long-term dependencies within and across modalities through attention mechanisms, thereby deriving enhanced multimodal features  $\mathbf{X}_C \in \mathbb{R}^{d_s \times c}$  and  $\mathbf{X}_D \in \mathbb{R}^{d_s \times c}$ :

$$\begin{bmatrix} \mathbf{X}_C \\ \mathbf{X}_D \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{C \rightarrow C} & \mathbf{A}_{C \rightarrow D} \\ \mathbf{A}_{D \rightarrow C} & \mathbf{A}_{D \rightarrow D} \end{bmatrix} \begin{bmatrix} \mathbf{V}_C \\ \mathbf{V}_D \end{bmatrix} = \begin{bmatrix} \underbrace{\sigma\left(\frac{\mathbf{Q}_C \mathbf{K}_C^\top}{\sqrt{d_s}}\right)}_{C \rightarrow C} & \underbrace{\sigma\left(\frac{\mathbf{Q}_C \mathbf{K}_D^\top}{\sqrt{d_s}}\right)}_{C \rightarrow D} \\ \underbrace{\sigma\left(\frac{\mathbf{Q}_D \mathbf{K}_C^\top}{\sqrt{d_s}}\right)}_{D \rightarrow C} & \underbrace{\sigma\left(\frac{\mathbf{Q}_D \mathbf{K}_D^\top}{\sqrt{d_s}}\right)}_{D \rightarrow D} \end{bmatrix} \begin{bmatrix} \mathbf{V}_C \\ \mathbf{V}_D \end{bmatrix} \quad (1)$$

where  $\sigma$  is the row-wise softmax activation function. In Eq. 1, MI-A can be decomposed into four terms for intra- and inter-modal interactions: 1) CT-to-CT intra-modal interactive attention  $\mathbf{A}_{C \rightarrow C} \in \mathbb{R}^{d_s \times d_s}$  models the global context association of CT image features to localize key areas; 2) CT-to-Dose inter-modal interactive attention  $\mathbf{A}_{C \rightarrow D} \in \mathbb{R}^{d_s \times d_s}$  models the response mechanism of radiation dose distribution guided by CT image features; 3) Dose-to-CT inter-modal interactive attention  $\mathbf{A}_{D \rightarrow C} \in \mathbb{R}^{d_s \times d_s}$  models the response mechanism of key CT image regions guided by dose map features; 4) Dose-to-Dose intra-modal interactive attention  $\mathbf{A}_{D \rightarrow D} \in \mathbb{R}^{d_s \times d_s}$  models the cumulative effect and spatial continuity of dose distribution.

**Self-adaptive Kolmogorov-Arnold Fusion (S-KaF, Fig. 2c).** The S-KaF is designed to learn a self-adaptive matrix for fusing enhanced CT and dose map features into a joint representation, by leveraging the nonlinear fitting capabilities of KAN. Specifically, S-KaF first mixes the enhanced features  $\mathbf{X}_C$  and  $\mathbf{X}_D$  outputted by MI-A to generate the hybrid enhanced feature  $\mathbf{H} \in \mathbb{R}^{d_s \times c}$ :

$$\mathbf{H} = \text{KANLinear}(\text{KANLinear}(\mathbf{X}_C) \odot \text{KANLinear}(\mathbf{X}_D) + \mathbf{X}_C + \mathbf{X}_D) \quad (2)$$

where  $\odot$  and  $\text{KANLinear}$  represent the element-wise multiplication and 1-layer KAN, respectively. The hybrid enhanced feature is then fed into the MHSA [14] layer to capture the spatial long-term dependency, resulting in a self-adaptive matrix for multimodal fusion:  $\tilde{\mathbf{H}} = \text{MHSA}(\mathbf{H}) \odot \mathbf{X}_C + \text{MHSA}(\mathbf{H}) \odot \mathbf{X}_D$ . Therefore, lymphocyte-related organs-aware features  $\mathbf{X}_F$  can be expressed as  $\mathbf{X}_F = \text{KAN}\left(\text{LN}\left(\tilde{\mathbf{H}}\right)\right) + \tilde{\mathbf{H}}$ .

**Summary of Advantages:** MI-AKaF is designed by incorporating multimodal interaction attention with KAN to capture lymphocyte-related organs-aware features with CT images and dose maps.

## 2.2 Cross-modality Guided AKaF (CG-AKaF) for comprehensive representation in dose-damage relationships

CG-AKaF is specifically designed to bridge the modality gap between image modalities (CT image and dose map) and meta-data for comprehensive representation in the modeling of dose-damage relationships. CG-AKaF enables

**Table 1.** Modality effectiveness study demonstrates that our RadKAM achieves the best performance for RIL prediction with three heterogeneous multimodal data (%).

CT Dose Meta	AUC(↑)	ACC(↑)	PRE(↑)	SEN(↑)	SPE(↑)	F1(↑)
✓	72.73±5.67	72.50±3.29	67.32±3.23	61.91±4.56	84.23±3.07	62.24±4.12
✓	68.93±4.25	70.15±3.20	66.98±14.76	57.24±4.93	81.67±2.14	56.26±6.95
✓	70.63±1.92	59.72±2.88	56.08±4.86	51.46±2.92	75.02±1.33	51.65±3.91
✓	77.18±6.78	77.24±3.67	78.14±5.71	69.26±6.66	86.39±3.78	70.73±5.62
✓	81.53±5.72	77.73±2.05	77.22±5.08	68.99±3.03	86.44±2.28	70.94±2.41
✓	78.93±1.94	74.42±1.76	80.48±5.11	63.66±3.01	83.29±1.98	66.18±2.51
✓	<b>87.32±3.80</b>	<b>83.40±2.47</b>	<b>83.00±3.29</b>	<b>79.11±4.55</b>	<b>90.34±2.56</b>	<b>80.12±2.73</b>

the model to selectively transfer information from image modalities to meta-data modality through a unidirectional guidance strategy. CG-AKaF consists of **Cross-modality Guided Attention (CGA)** and **Kolmogorov-Arnold Fusion (KaF)**. In CGA, meta-data feature  $\mathbf{F}_M \in \mathbb{R}^{d_m}$  is used as the query  $\mathbf{Q}_M$ , and the concatenation of  $\mathbf{F}_M$  and image modality fusion feature  $\mathbf{E}_I$  are used as key  $\mathbf{K}_M$  and value  $\mathbf{V}_M$ , so that information in two image modalities can be transferred to meta-data modality. Then, leveraging the MHSA mechanism, we can get the output  $\mathbf{X}_M \in \mathbb{R}^m$ :

$$\mathbf{X}_M = \sigma \left( \frac{\mathbf{Q}_M \mathbf{K}_M^\top}{\sqrt{d_m}} \right) \mathbf{V}_M + \mathbf{F}_M \quad (3)$$

Then, the comprehensive representation, which considers lymphocyte-related organs, dose-damage relationships, and individual differences, is captured using KAN as  $\mathbf{E}_M = \text{KAN}(\text{LN}(\mathbf{X}_M)) + \mathbf{X}_M$ .

**Summary of Advantages:** CG-AKaF achieves comprehensive representation of CT image, dose map and meta-data in modeling of dose-damage relationships.

### 3 Experiments and Results Analysis

#### 3.1 Experimental Configuration

**Dataset and Data Preprocessing.** This study includes 211 NPC patients, including 40 individuals with G2RIL, 114 individuals with G3RIL, and 57 individuals with G4RIL. The CT images and dose maps, after alignment, removal of excess parts, resampling, and normalization, have a size of  $128 \times 128 \times 128$ . The meta-data comprises gender, age, tumor stage (T-, N-, M-, and overall-stage), baseline lymphocyte count, and baseline neutrophil count.

**Implementation Setup.** All experiments are based on Python 3.9, PyTorch 1.13.0, CUDA 11.7, and a single NVIDIA GeForce RTX 3090 GPU. The loss function for all experiments is the cross-entropy loss. We use a five-fold cross-validation method to verify the effectiveness of the proposed RadKAM. 80% of the data is used as a training set and 20% as a test set. The initial learning rate is set to  $1e-4$  and the weight decay of  $1e-5$ . The decay strategy of the learning rate is cosine annealing. Training epochs are set to 100. The loss function is minimized with Adam optimizer.

**Table 2.** Ablation studies show that the effectiveness of each components in our RadKAM for RIL prediction (%).

MI-AKaF MI-A S-KaF		CG-AKaF	KAN	AUC(↑)	ACC(↑)	PRE(↑)	SEN(↑)	SPE(↑)	F1(↑)
			✓	75.38±4.02	73.94±3.26	77.49±7.13	61.58±4.72	83.58±3.52	61.74±4.11
✓	✓		✓	84.12±4.49	80.10±4.29	80.81±5.51	73.18±6.05	87.99±3.78	74.81±5.28
		✓	✓	80.30±5.36	77.72±4.02	79.13±4.33	69.60±6.71	86.58±3.70	70.99±5.52
✓		✓	✓	82.75±3.14	79.61±3.66	77.63±4.44	71.39±6.72	87.69±3.43	72.81±6.21
	✓	✓	✓	87.22±4.77	80.56±2.03	79.47±3.15	74.35±2.79	88.46±1.29	75.91±2.76
✓	✓	✓		84.31±5.28	81.03±2.43	<b>83.11±3.47</b>	75.51±4.55	88.38±2.34	77.41±4.08
✓	✓	✓	✓	<b>87.32±3.80</b>	<b>83.40±2.47</b>	83.00±3.29	<b>79.11±4.55</b>	<b>90.34±2.56</b>	<b>80.12±2.73</b>

### 3.2 Performance Evaluation and Analysis

Quantitative results (Tab. 1) demonstrate that the proposed RadKAM achieves remarkable performance in RIL prediction, with an area under the curve (AUC) of  $87.32\pm3.80\%$ , an accuracy (ACC) of  $83.40\pm2.47\%$ , a precision (PRE) of  $83.00\pm3.29\%$ , a sensitivity (SEN) of  $79.11\pm4.55\%$ , a specificity (SPE) of  $90.34\pm2.56\%$ , and a F1-score (F1) of  $80.12\pm2.73\%$ .

**Modality Effectiveness Study.** Experimental results (Tab. 1) show that RadKAM achieves the best performance of RIL prediction by learning from multimodal data together. It also should be noted that RIL severity is most sensitive to modality-aligned feature representation. Benefitting from the proposed multimodal learning framework, RadKAM understands the patient’s condition comprehensively to make an accurate RIL prediction by modeling the complementary mapping relationship between multimodal features and ALC decline.

**Ablation Study.** To evaluate the effectiveness of each components of our RadKAM, extensive ablation studies are conducted (Tab. 2). **1) Effectiveness of MI-AKaF:** Experimental results indicate that MI-AKaF effectively captures the lymphocyte-related organs-aware features to advance the RIL prediction. With the MI-AKaF, a great improvement is achieved with the average AUC, ACC and F1 of 8.94%, 6.16% and 13.07%, respectively. Furthermore, we discuss the effectiveness of MI-A and S-KaF in MI-AKaF. The results show that both MI-A and S-KaF can improve the performance of the network, and the organic combination of MI-A and S-KaF is beneficial to MI-AKaF to integrate CT image features and dose map features. **2) Effectiveness of CG-AKaF:** Experimental results indicate that CG-AKaF effectively captures the meta-data features incorporating radiation damage to advance the RIL prediction. With the CG-AKaF, a great improvement is achieved with the average AUC, ACC and F1 of 4.92%, 3.78% and 9.25%, respectively. **3) Effectiveness of KAN:** To evaluate the effectiveness of KAN’s nonlinear modeling capability in RIL prediction, we further replace all KANs in RadKAM with MLPs with the same number of layers. Experimental results show that compared with MLP, KAN can better model various complex relationships in RIL prediction.

**Comparison Experiments. 1) Outperformance of AKaF in multimodal learning:** To evaluate the outperformance of the AKaF in RIL prediction, extensive experiments are conducted by dividing the multimodal learning into two

**Table 3.** Comparison experiments demonstrate the effectiveness of the proposed AKaF for multimodal learning in RIL prediction (%).

Stage 1	Method Stage 2	AUC(↑)	ACC(↑)	PRE(↑)	SEN(↑)	SPE(↑)	F1(↑)
MMTM [6]	Concat	74.16±2.43	73.92±3.04	72.82±10.05	63.32±3.60	84.63±2.53	63.39±4.40
	MLP	79.65±5.80	77.24±2.79	77.80±7.47	69.12±5.87	86.59±3.06	70.41±4.77
	KAN	80.47±5.08	78.18±3.21	78.58±5.59	70.55±6.76	87.09±3.14	71.95±5.50
HcCNN [1]	Concat	72.84±4.96	73.47±2.50	67.38±4.06	63.44±6.42	84.68±2.58	63.30±5.91
	MLP	79.67±3.00	78.19±2.03	78.84±6.60	70.37±4.34	86.80±3.12	72.35±2.63
	KAN	78.19±4.87	77.72±1.41	76.85±5.89	70.90±2.27	86.65±1.89	72.80±2.47
Yang <i>et al.</i> [20]	Concat	73.32±3.37	72.98±2.78	72.14±5.50	61.85±4.85	83.41±2.54	62.56±5.40
	MLP	76.06±4.73	75.72±3.17	80.26±3.66	63.06±4.09	83.41±2.80	64.68±4.82
	KAN	76.45±2.17	76.29±3.46	74.71±3.19	67.35±3.94	85.62±3.04	68.46±2.86
MDL-Net [12]	Concat	77.56±3.41	75.35±2.77	76.62±5.30	63.35±3.45	84.03±2.15	65.30±4.21
	MLP	80.35±5.04	76.28±5.92	78.82±8.56	64.72±8.45	84.62±4.72	67.16±8.76
	KAN	83.91±5.04	77.22±4.77	79.97±5.74	71.47±7.14	86.69±4.10	71.47±6.45
Ours (MI-AkaF)	Ours (CG-AkaF)	<b>87.32±3.80</b>	<b>83.40±2.47</b>	<b>83.00±3.29</b>	<b>79.11±4.55</b>	<b>90.34±2.56</b>	<b>80.12±2.73</b>

**Table 4.** Comparison results of relevant studies (NSCLC: Non-Small Cell Lung Cancer, ESC: Esophageal Cancer, HCC: Hepatocellular Carcinoma, NPC: Nasopharyngeal Carcinom, DVH: Dose-Volume Histogram, and -: Not exist in original paper).

Method	Type	Data	Subjects	AUC(↑)	ACC(↑)	PRE(↑)	SEN(↑)	SPE(↑)	F1(↑)
Zhu <i>et al.</i> [23]	ESC	DVH;Meta	721	0.831	0.769	0.670	0.610	-	0.631
Kim <i>et al.</i> [7]	NSCLC	DVH;Meta	139(117;22)	0.77	-	-	-	-	-
Xu <i>et al.</i> [19]	NSCLC	DVH;Meta	130(87;43)	0.77	0.76	-	0.76	0.76	-
Huang <i>et al.</i> [5]	NPC	CT;Dose;Meta	125(100;25)	0.93	0.88	0.73	0.8	0.9	0.76
<b>Ours (RadKAM)</b>	NPC	CT;Dose;Meta	211(40;114;57)	0.873	0.834	<b>0.830</b>	0.791	<b>0.903</b>	<b>0.801</b>

stages: one to learn from CT images and dose maps and another to learn the comprehensive representation with meta-data and outputs of the first stage. Four image fusion approaches [6, 1, 20, 12] are adopted for stage 1 and *Concat*, *MLP*, *vanilla KAN* are adopted for stage 2. Results (Tab. 3) demonstrate that the proposed AKaF enables the model to achieve the best performance of multimodal learning for both heterogeneous meta-data-image and homogeneous image-dose data. **2) Outperformance of RadKAM for RIL prediction:** Comparison results (Tab. 4) show that the proposed RadKAM achieves the best performance for RIL prediction on most of the evaluation metrics. It should be noted that the performance of the RadKAM in terms of AUC, ACC, and SEN is slightly inferior to the work [5], because there exists a huge difference between this work and [5] on task definition. Work in [5] only makes a binary classification between G4RIL and G2-3RIL, while RadKAM is applied to predict RIL with fine-grained classification of G2RIL, G3RIL and G4RIL. In the approach described in [5], manual segmentation of regions of interest (ROIs) in CT images is required prior to RIL prediction. In contrast, RadKAM is an end-to-end multimodal learning framework that operates without any such prerequisites.

## 4 Conclusion

In this paper, a newly designed “divide and conquer” multimodal learning framework RadKAM is proposed to achieve accurate radiation-induced lymphope-



nia (RIL) prediction before nasopharyngeal carcinoma (NPC) radiation therapy (RT). To the best of our knowledge, this is the first RIL prediction work that does not require the delineation of the region of interest. RadKAM relies on a key idea of Attention-driven Kolmogorov-Arnold Fusion (AKaF) for multimodal learning. Through a series of rigorous experiments, our proposed method achieves accurate RIL prediction. The results show that our method holds great promise in automatically predicting RIL from radiotherapy treatment plans before treatment and assisting doctors in refining radiotherapy treatment plans.

**Acknowledgments.** This work is supported by the National Natural Science Foundation of China (62372474), and the 111 Project (B18059). We are grateful for resources from the High Performance Computing Center of Central South University.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Bi, L., Feng, D.D., Fulham, M., Kim, J.: Multi-label classification of multi-modality skin lesion via hyper-connected convolutional neural network. *Pattern Recognition* **107**, 107502 (2020)
2. Cella, L., Monti, S., Pacelli, R., Palma, G.: Modeling frameworks for radiation induced lymphopenia: A critical review. *Radiotherapy and Oncology* **190**, 110041 (2024)
3. Chen, Y., Zhu, Z., Zhu, S., Qiu, L., Zou, B., Jia, F., Zhu, Y., Zhang, C., Fang, Z., Qin, F., et al.: Skansformer: Fine-grained classification of bone marrow cells via kansformer backbone and hierarchical attention mechanisms. *IEEE Journal of Biomedical and Health Informatics* (2024)
4. Dai, D., Tian, Q., Shui, Y., Li, J., Wei, Q.: The impact of radiation induced lymphopenia in the prognosis of head and neck cancer: A systematic review and meta-analysis. *Radiotherapy and Oncology* **168**, 28–36 (2022)
5. Huang, Q., Yang, C., Pang, J., Zeng, B., Yang, P., Zhou, R., Wu, H., Shen, L., Zhang, R., Lou, F., et al.: Ct-based dosiomics and radiomics model predicts radiation-induced lymphopenia in nasopharyngeal carcinoma patients. *Frontiers in Oncology* **13**, 1168995 (2023)
6. Joze, H.R.V., Shaban, A., Iuzzolino, M.L., Koishida, K.: Mmtm: Multimodal transfer module for cnn fusion. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 13289–13299 (2020)
7. Kim, Y., Chamseddine, I., Cho, Y., Kim, J.S., Mohan, R., Shusharina, N., Paganetti, H., Lin, S., Yoon, H.I., Cho, S., et al.: Neural network based ensemble model to predict radiation induced lymphopenia after concurrent chemoradiotherapy for non-small cell lung cancer from two institutions. *Neoplasia* **39**, 100889 (2023)
8. Li, C., Liu, X., Li, W., Wang, C., Liu, H., Liu, Y., Chen, Z., Yuan, Y.: U-kan makes strong backbone for medical image segmentation and generation. *arXiv preprint arXiv:2406.02918* (2024)
9. Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T.Y., Tegmark, M.: Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756* (2024)

10. Mohan, R., Liu, A.Y., Brown, P.D., Mahajan, A., Dinh, J., Chung, C., McAvoy, S., McAleer, M.F., Lin, S.H., Li, J., et al.: Proton therapy reduces the likelihood of high-grade radiation-induced lymphopenia in glioblastoma patients: phase ii randomized study of protons vs photons. *Neuro-oncology* **23**(2), 284–294 (2021)
11. OuYang, P.Y., Zhang, B.Y., Guo, J.G., Liu, J.N., Li, J., Peng, Q.H., Yang, S.S., He, Y., Liu, Z.Q., Zhao, Y.N., et al.: Deep learning-based precise prediction and early detection of radiation-induced temporal lobe injury for nasopharyngeal carcinoma. *EClinicalMedicine* **58** (2023)
12. Qiu, Z., Yang, P., Xiao, C., Wang, S., Xiao, X., Qin, J., Liu, C.M., Wang, T., Lei, B.: 3d multimodal fusion network with disease-induced joint learning for early alzheimer's disease diagnosis. *IEEE Transactions on Medical Imaging* (2024)
13. Terrones-Campos, C., Ledergerber, B., Vogelius, I.R., Helleberg, M., Specht, L., Lundgren, J.: Hematological toxicity in patients with solid malignant tumors treated with radiation—temporal analysis, dose response and impact on survival. *Radiotherapy and Oncology* **158**, 175–183 (2021)
14. Vaswani, A.: Attention is all you need. *Advances in Neural Information Processing Systems* (2017)
15. Venkatesulu, B., Giridhar, P., Pujari, L., Chou, B., Lee, J.H., Block, A.M., Upadhyay, R., Welsh, J.S., Harkenrider, M.M., Krishnan, S., et al.: Lymphocyte sparing normal tissue effects in the clinic (lymphotec): A systematic review of dose constraint considerations to mitigate radiation-related lymphopenia in the era of immunotherapy. *Radiotherapy and Oncology* **177**, 81–94 (2022)
16. Wang, G., Zhu, Q., Song, C., Wei, B., Li, S.: Medkaformer: When kolmogorov-arnold theorem meets vision transformer for medical image representation. *IEEE Journal of Biomedical and Health Informatics* (2025)
17. Xie, X., Gong, S., Jin, H., Yang, P., Xu, T., Cai, Y., Guo, C., Zhang, R., Lou, F., Yang, W., et al.: Radiation-induced lymphopenia correlates with survival in nasopharyngeal carcinoma: impact of treatment modality and the baseline lymphocyte count. *Radiation oncology* **15**, 1–10 (2020)
18. Xie, X., Yang, L., Zhao, F., Wang, D., Zhang, H., He, X., Cao, X., Yi, H., He, X., Hou, Y.: A deep learning model combining multimodal radiomics, clinical and imaging features for differentiating ocular adnexal lymphoma from idiopathic orbital inflammation. *European Radiology* **32**(10), 6922–6932 (2022)
19. Xu, Z., Yang, L., Yu, H., Guo, L.: A machine learning model for grade 4 lymphopenia prediction during pelvic radiotherapy in patients with cervical cancer. *Frontiers in Oncology* **12**, 905222 (2022)
20. Yang, P., Zhang, Y., Lei, H., Bian, Y., Yang, Q., Lei, B.: Acute ischemic stroke onset time classification with dynamic convolution and perfusion maps fusion. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 558–568. Springer (2023)
21. Yu, H., Chen, F., Lam, K.O., Yang, L., Wang, Y., Jin, J.Y., Ei Helali, A., Kong, F.M.: Potential determinants for radiation-induced lymphopenia in patients with breast cancer using interpretable machine learning approach. *Frontiers in Immunology* **13**, 768811 (2022)
22. Zhang, X., Zheng, W., Huang, S., Haojiang, L., Zhisheng, B., Xin, Y.: Xerostomia prediction in patients with nasopharyngeal carcinoma during radiotherapy using segmental dose distribution in dosiomics and radiomics models. *Oral Oncology* **158**, 107000 (2024)
23. Zhu, C., Lin, S.H., Jiang, X., Xiang, Y., Belal, Z., Jun, G., Mohan, R.: A novel deep learning model using dosimetric and clinical information for grade 4 radiotherapy-

induced lymphopenia prediction. Physics in Medicine & Biology **65**(3), 035014 (2020)