Multi-Kernel Fusion with Fuzzy Label Relaxation for Predicting Distant Metastasis in Nasopharyngeal Carcinoma

Jiabao SHENG
The Dept. of Health Technology and Informatics
Research Institute for Smart Ageing
The Hong Kong Polytechnic University
Email: jia-bao.sheng@connect.polyu.hk

SaiKit LAM
The Dept. of Biomedical Engineering
Research Institute for Smart Ageing
The Hong Kong Polytechnic University
Email: saikit.lam@polyu.edu.hk

Ta ZHOU
The Dept. of Health Technology and Informatics
The Hong Kong Polytechnic University
Email: ta.zhou@polyu.edu.hk

Jiang ZHANG
The Dept. of Health Technology and Informatics
The Hong Kong Polytechnic University
Email: jiang.zhang@connect.polyu.hk

Yuanpeng ZHANG*
The Dept. of Health Technology and Informatics
The Hong Kong Polytechnic University
Email: y.p.zhang@ieee.org

Jing CAI*
The Dept. of Health Technology and Informatics
Research Institute for Smart Ageing
Shenzhen Research Institute
The Hong Kong Polytechnic University
Email: jing.cai@polyu.edu.hk

***Abstract*—The heterogeneity of omics data poses a challenge for feature fusion in the medical field due to source differences. This study aims to construct a fusion method that can reduce the differences between omics data, enabling them to jointly contribute to specific medical tasks. The multi-kernel late-fusion method is capable of reducing the impact of these differences by mapping the features using the most suitable single-kernel function and then combining them in a high-dimensional space that can effectively represent the data. However, the strict label fitting of complex nasopharyngeal carcinoma (NPC) data samples restricts the performance of general classifiers when using high-dimensional features. To address this issue, this study proposes a multi-kernel model for multi-omics feature fusion in predicting distant metastasis of NPC patients. The proposed model employs a multi-kernel-based Radial basis function (RBF) neural network and introduces a label fuzzy softening method to enlarge the margin between two classes. By mapping the original medical omics data and reducing the differences, the proposed method provides more degrees of freedom for label fitting, improving the classification ability. The proposed model is evaluated on multi-omics datasets, and the results demonstrate its strength and effectiveness in predicting distant metastasis of NPC patients.**

***Keywords*—Multi-omics fusion, Multi-kernel Learning, Multi-modality**

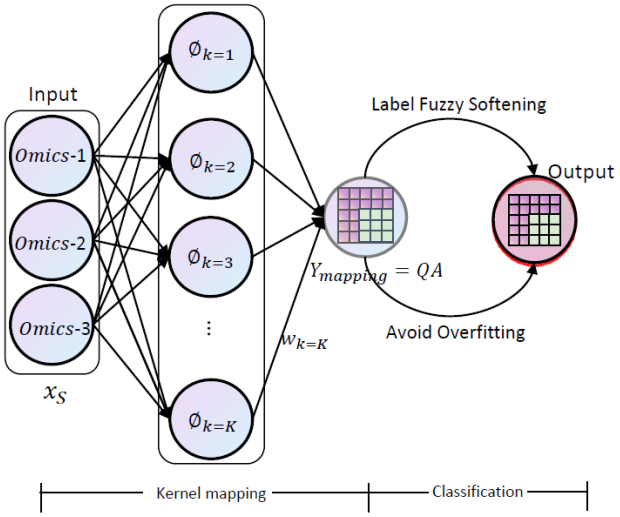


Fig. 1: The illustration of FLS_MK. S is the number of samples. We select the omics images of one patient as x_S, ϕ_k is the mapping in the k-th neuron, w_k is the weight in the k-th neuron, QA is The final mapping result (more details are shown in section III)

I. Introduction

NASOPHARYNGEAL carcinoma (NPC) is prevalent cancer among Asians [1]. Standard treatments include radiotherapy alone for early stages and combined radiotherapy and chemotherapy for advanced lesions [2], [3]. However, more than 30% of patients with advanced NPC fail to respond to treatment, mainly due to distant metastasis rather than local recurrence [4], [5]. While doctors rely on medical imaging to diagnose cancer stage and provide radiation therapy doses, researchers have used single omics approaches such as radiomics studies to predict NPC prognosis, achieving significant success in predicting distant metastasis. However, the use of single omics approaches may not be sufficient, and the integration of multiple omics data may be necessary for better prediction accuracy.

Although signal omics are useful for expressing training features in a model, their limited ability to cover all relevant features can hinder learning and limit predictive performance. In contrast, multi-omics feature fusion can improve sample clustering, provide a better understanding of prognostic and predictive phenotypes, dissect cellular responses to therapy, and assist in translational research through integrative models [6], [7], [8]. Recent evidence has shown the effectiveness of the multi-omics approach in medical prediction tasks [9],

Yuanpeng ZHANG is the co-corresponding author.
Jing CAI is the corresponding author.

979-8-3503-3959-8/23/$31.00

[10]. Specifically, [9] use multilayer perceptron (MLP) to reconstruct each omic with a hierarchical representation and leverage shared self-expression coefficients to embed inter-class and intra-class structural information. Meanwhile, [10] consider the relationship among different omics to learn a latent representation space using all available samples. They use complete multi-modality data to learn a common latent representation and incomplete multi-modality data to learn modality-specific latent representations. In the study of NPC, radiomics and dosiomics are correlated with distant metastasis. However, radiomics data contain many features unrelated to distant metastasis prediction tasks, and temporal and angular differences exist between data from the same patient with different medical images. Existing methods cannot address the feature fusion challenges posed by data noise and out-of-sync data sources.

The study conducted by [11] demonstrates that the late-fusion method utilizing multi-kernel can mitigate the impact of various data differences, resulting in better fusion outcomes. The high-dimensional space created by multiple kernel functions is a combination space that merges multiple feature spaces. This combination space can incorporate the distinct feature mapping abilities of each subspace and merge different heterogeneous data from multiple sources. The most suitable single kernel function individually maps the features, and the data can be more accurately and reasonably expressed in the new combination space. However, for the original complex NPC data samples, the feature mapping process in high-dimensional space lacks classification capability when dealing with downstream tasks. Typically, multi-Kernel learning for classification tasks aims to learn a transformation matrix that can convert the combination of kernels into a binary label matrix. Due to the strict label fitting and limited flexibility, the general classifier is not very effective when employing high-dimensional features.

This study aims to address the problem of preserving the intrinsic geometry structure of transformed samples [12]. To achieve this goal, [13] proposed the margin Fisher analysis method which preserves both the intrinsic geometry structure and the discriminant structure of samples using label information. Some semi-supervised learning methods introduce an adjacency graph to capture the local structure of samples and ensure that similar samples have nearly the same labels [14], [15]. However, to ensure that samples from the same class are kept close together in the transformed space, it is necessary to capture the distribution of samples from the same class by using sample affinity.

This study proposes a novel multi-kernel model for multi-omics feature fusion to predict distant metastasis of NPC patients. This model serves as an automation platform for high-throughput data processing and analysis. In the pre-processing stage, a kernel fusion method with matrix-based mixing weights is utilized, which allows each participating kernel to learn independently. Additionally, to address the problem of overfitting, label fuzzy softening is introduced, which relaxes the strict binary label matrix into a slack variable matrix and constructs a class compactness graph to improve the model's generalization performance.

II. Materials

This section provides a concise overview of the NPC data used in our experiments, covering patient demographics, imaging acquisition protocols, preprocessing steps, and feature extraction techniques.

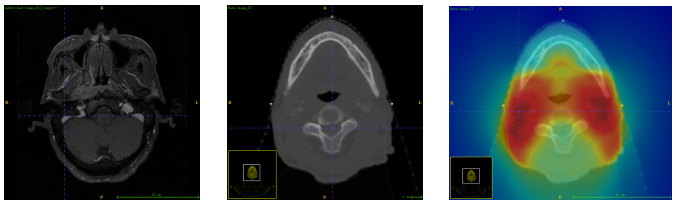
A. Patients and Dataset Statistic

This section provides a brief overview of the NPC dataset used in our experiments. To be included, patients had to meet the following criteria: i. confirmed diagnosis of NPC, ii. locally advanced stage I to IV cancer, and iii. availability of pre-treatment MRI and planning CT data from participating hospitals. The clinical endpoint selected for this study was distant metastasis, determined through fibrotic endoscopy and/or histological/radiological examination. Patients who developed distant metastases after completion of primary therapy were marked as 1, while those who remained disease-free until their last follow-up were marked as 0. The average follow-up period was 36 months.

To construct our NPC-contraparotid dataset, we collected medical data from 102 NPC patients undergoing radiation therapy in Hong Kong. Each patient's dataset included three types of information: MR images, CECT images, and dose data (see Fig.2). According to the clinical state, patients with clinically documented distant metastases were marked as 1, while those without distant metastases were marked as 0. TableI presents NPC-to-parotid statistics.

TABLE I: The statistic of NPC-ContraParotid dataset.

Organ	Datasets	MRI	Dose	CECT	Features	Size
ContraParotid	CTD		✓	✓	7	102
	MRICT	✓		✓	3	102
	MRID	✓	✓		6	102
	MRICTD	✓	✓	✓	14	102



(a) MRI (b) CT (c) Dose

Fig. 2: The sample medical images of NPC-ContraPariod.

B. Data Processing and Feature Extraction

To optimize the consistency, reproducibility, and validity of radiomics studies, it is crucial to preprocess the images prior to radiomics feature extraction, given the variability in image acquisition and reconstruction parameters within and between medical centers. In this study, four key image preprocessing steps were employed, including voxel size resampling, VOI segmentation, image filtering, and gray-level quantization, following the well-established recommendations in the Image

Biomarker Standardization Initiative guideline [16]. An in-house development pipeline tool based on Python v3.7.3 was used to perform these steps. The radiometric features were extracted using the publicly available PyRadiomics v2.2.0 and SimpleITK v1.2.4 packages, which were integrated into an in-house developed Python-based v3.7.3 pipeline. In order to select the features that contribute the most to both Radiomics and dosiomics, a hybrid selection method was utilized.

III. Methodology

A. Multi-Kernel Ridge Regression

Multi-kernel learning is a technique that aims to improve mapping performance by combining different kernel functions or kernel functions with varying parameters. There are various combinations of kernel functions, with the most commonly used being the linear combination. Let us assume that we have L different kernel functions, where ϕ_l represents the l^{th} function with $1 \leq l \leq L$. In this case, a linear combination of these kernel functions can be expressed as follows:

$$\boldsymbol{\phi} = \sum_{l=1}^{L} \boldsymbol{\omega}_l \boldsymbol{\phi}_l, \tag{1}$$

where, $\boldsymbol{\omega}_l$ is the mixing weight for the l^{th} kernel function, and $\boldsymbol{\phi}_l$ is the corresponding kernel function. The goal of multi-kernel learning is to learn the optimal mixing weights for these kernel functions to achieve better performance in mapping the input data to the output.

In this method, we propose a novel architecture for the Radial Basis Function (RBF) neural network [17], which consists of an input layer, a nonlinear hidden layer, and a linear output layer (shown in Fig.1). Let $X \in \mathbb{R}^{a \times S}$ represent an input dataset consisting of S samples, where each sample $\boldsymbol{x} \in \mathbb{R}^{a \times 1}$ is represented by a number of attributes. For all k, let $m_k \subset M \in \mathbb{R}^{a \times K}$, where K is the number of neurons in the hidden layer of the RBF network. Here, $M \in \mathbb{R}^{a \times K}$ comprises K number of $m_k \in \mathbb{R}^{a \times 1}$ vectors, each representing a center point of the kernel of the k^{th} hidden neuron. Then, the output for each neuron can be expressed as

$$\phi_k(\boldsymbol{x}, \boldsymbol{m_k}) = \sum_{l=1}^{L} \omega_{l_k} \phi_{l_k}(\boldsymbol{x}, \boldsymbol{m_k}), \tag{2}$$

Let L be a set of different kernels in the k-th neuron, and let $l \in L$. Then, ϕ_{l_k} is the l-th primary kernel of the k-th neuron, and $\boldsymbol{\omega}_{l_k}$ is its corresponding mixing weight. Two constraints hold on $\boldsymbol{\omega}_{l_k}$: $0 \leq \boldsymbol{\omega}_{l_k} \leq 1$, and $\sum_l^L \boldsymbol{\omega}_{l_k}$=1. The common set of kernel weights for all multi-kernels, combined with these two constraints, ensure that the participating kernels form a convex combination.

According to RBF, for an input sample x, the corresponding output y can be formulated as

$$y = \sum_{k=1}^{K} w_k \sum_{l=1}^{L} \omega_{lk} \phi_{lk}(\boldsymbol{x}, \boldsymbol{m_k}) + b, \tag{3}$$

Eq.3 can be re-formulated as

$$y = \Phi^{\intercal} w \tag{4}$$

Where $w = [b, w_{11}, w_{12}, \ldots, w_{1K}, \ldots, w_{L1}, w_{L2}, \ldots, w_{LK}]^{\intercal}$, $\Phi = [1, \phi_{11}(\boldsymbol{x}, \boldsymbol{m_1}), \phi_{12}(\boldsymbol{x}, \boldsymbol{m_2}), \ldots, \phi_{1K}(\boldsymbol{x}, \boldsymbol{m_K}), \ldots, \phi_{L1}(\boldsymbol{x}, \boldsymbol{m_1}), \phi_{L2}(\boldsymbol{x}, \boldsymbol{m_2}), \ldots, \phi_{LK}(\boldsymbol{x}, \boldsymbol{m_K})]^{\intercal}$. Therefore, for a training set $X = [\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_N}]^{\intercal}$, the corresponding output Y can be formulated as

$$\boldsymbol{Y}_{mapping} = \boldsymbol{QA} \tag{5}$$

$$\boldsymbol{Q} = \begin{bmatrix} 1 & \phi_{l1}(\boldsymbol{x_1}, \boldsymbol{m_1}) & \ldots & \phi_{LK}(\boldsymbol{x_1}, \boldsymbol{m_K}) \\ 1 & \phi_{l1}(\boldsymbol{x_2}, \boldsymbol{m_1}) & \ldots & \phi_{LK}(\boldsymbol{x_2}, \boldsymbol{m_K}) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \phi_{l1}(\boldsymbol{x_N}, \boldsymbol{m_1}) & \ldots & \phi_{LK}(\boldsymbol{x_N}, \boldsymbol{m_K}) \end{bmatrix} \tag{6}$$

$$\boldsymbol{A} = \begin{bmatrix} b & w_{l1}(1) & \ldots & w_{LK}(1) \\ b & w_{l1}(2) & \ldots & w_{LK}(2) \\ \vdots & \vdots & \vdots & \vdots \\ b & w_{l1}(N) & \ldots & w_{LK}(N) \end{bmatrix}^{\intercal} \tag{7}$$

Based on the criterion of empirical and structural risk minimization, we can optimize A by the following objective,

$$\min_{A} \|\boldsymbol{QA} - \boldsymbol{Y}\|_F^2 + \lambda \|\boldsymbol{A}\|_F^2, \tag{8}$$

λ is a positive regularization parameter, and $\lambda \geq 0$.

B. Label Fuzzy Softening

Previous studies have demonstrated that learning discriminative models through overfitting strictly binary label matrices is inadequate. To address this issue, we propose an approach that fuzzy softens the label matrix Y using two matrices, V and Z, similar to the method employed in [18], thereby increasing the margin between classes.

$$\widehat{Y} = Y + V \bigodot Z, \tag{9}$$

where, $\bigodot$ represents the Hadamard operator, and V and Z are defined as follows:

$$V_{ij} = \begin{cases} +1, & \text{if } y_{ij} = 1 \\ -1, & \text{if } y_{ij} = 0 \end{cases} \tag{10}$$

$$Z = \begin{bmatrix} z_{11} & \ldots & z_{1J} \\ \vdots & z_{ij} & \vdots \\ z_{I1} & \ldots & z_{IJ} \end{bmatrix} \tag{11}$$
$$i = 1, 2, ..., I,$$
$$j = 1, 2, ..., J, z_{ij} \geqslant 0, i \neq j.$$

After fuzzy softening the label, the original multi-Kernel regression model in Eq. 8 is updated as:

$$\begin{aligned} &\min_{\boldsymbol{A}, Z} \|\boldsymbol{QA} - (\boldsymbol{Y} + V \odot Z)\|_F^2 + \lambda \|\boldsymbol{A}\|_F^2 \\ &s.t. Z \geq 0 \end{aligned} \tag{12}$$

C. Overfitting Problem

Although label fuzzy softening can expand class margins and improve classification performance, overfitting may occur due to the degree of freedom in fitting. Thus, it is important to suppress overfitting while seeking a more discriminative model. To control data fitting, we perform a regularization term operation. First, we construct an undirected graph to capture the relationships between samples, which is defined as:

$$C_{ij} = \begin{cases} e^{\frac{\|\boldsymbol{x_i}-\boldsymbol{x_j}\|^2}{\sigma}}, \text{if } \boldsymbol{x_i} \text{ and } \boldsymbol{x_j} \text{ have the same label} \\ 0 \end{cases} \tag{13}$$

Where σ is the kernel width. From the above equation, we observe that if two samples belong to the same class, the closer their distance, the larger the weight. Conversely, if they belong to different classes, the weight is 0. Therefore, when the samples are transformed into the label space, we can use the following objective to ensure our assumption:

$$\min_{\boldsymbol{A},Z} \sum_{ij} \|\boldsymbol{g_j} - \boldsymbol{g_j}\|^2 C_{ij} = \min_{\boldsymbol{A}} tr(\boldsymbol{A}^\intercal \boldsymbol{Q}^\intercal L \boldsymbol{Q} \boldsymbol{A}) \tag{14}$$

Where, $\|A\|_F^2 = tr(A^\intercal A) = tr(AA^\intercal)$ represents the squared Frobenius norm of the matrix A, and $tr(.)$ is the trace operator of a matrix. $\boldsymbol{g_i} = \mathbf{x_i}A$ denotes the transformed result of the sample $\mathbf{x_i}$. Minimizing this objective further ensures that $\boldsymbol{g_i}$ and $\boldsymbol{g_j}$ are close in the transformed space. L is the Laplacian matrix that can be computed by $L = G - C$, where G is a diagonal matrix and its diagonal entries are defined as $Gij = \sum_j \mathbf{C}_{ij}$.

By substituting Eq. 14 into Eq. 8, we propose the following objective function for FLS_MK:

$$\begin{aligned} &\min_{\boldsymbol{A},Z} \|\boldsymbol{Q}\boldsymbol{A} - (\boldsymbol{Y} + V \odot Z\|_F^2 + \lambda tr(\boldsymbol{A}^\intercal \boldsymbol{Q}^\intercal L \boldsymbol{Q} \boldsymbol{A}) \\ &s.t. Z \geq 0 \end{aligned} \tag{15}$$

IV. Experiment

We created four multi-omics datasets, each consisting of 102 samples (Table I). Radiomic features were extracted from the planning dose map, pre-treatment contrast-enhanced CT, and T1-weighted contrast-enhanced MR images within the ContrasParid dataset to create a comprehensive feature set. To showcase the proposed classification performance of our model, we selected four models: two improved fuzzy classification models based on SVM (FSVMs [19] and FSVMs_CIP [20]), RBFNN_MK [17] to compare the effect of label softening, and Ridge Regression [21] as the baseline model for regression analysis.

TABLE II: The ACC of NPC-ContraParotid.

Modalities	Rige Regression	RBFNN_MK	FSVMs	FSVMs_CIP	FLS_MK
CTD	91.1765	79.4118	86.1354	82.3529	91.6670
MRICT	94.1176	76.4706	82.3529	88.2358	89.5449
MRID	91.1765	55.8824	88.2353	85.2941	95.8940
MRICTD	82.3529	58.8235	77.6228	79.4118	96.0723

TABLE III: The AUC of NPC-ContraParotid.

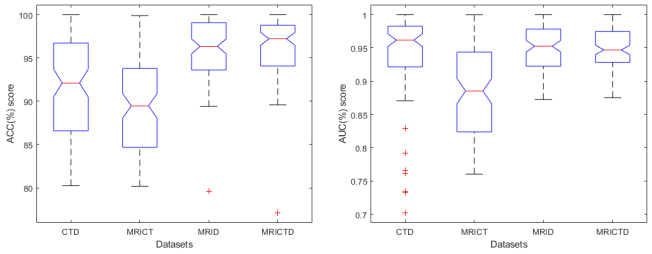
Modalities	Rige Regression	RBFNN_MK	FSVMs	FSVMs_CIP	FLS_MK
CTD	0.8635	0.9086	0.6428	0.6471	0.9404
MRICT	0.8824	0.9123	0.6644	0.7751	0.8812
MRID	0.8304	0.7612	0.7647	0.7059	0.9471
MRICTD	0.6782	0.7474	0.5736	0.5882	0.9497

A. Settings

We employed K-fold cross-validation with a value of $k = 3$ for both training and testing our model. Accuracy and Area Under the Curve (AUC) were utilized as quantitative metrics to evaluate the performance of the model. Specifically, accuracy was calculated as the ratio of correctly classified samples to the total number of samples. On the other hand, AUC is a commonly used evaluation index to measure the effectiveness of binary classification models, indicating the probability that a predicted positive example is ranked higher than a negative example.

B. Expriment results

Table II demonstrates that the FLS_MK model outperforms other models in terms of ACC performance on the CTD, MRID, and MRICTD datasets. Although the ACC score displayed by FLS_MK on the MRICT dataset is not as high as that of ridge regression, it can be observed from the number of features in the dataset that FLS_MK model performs better in fusing data with multiple quantitative features, as shown in Fig 3. Ridge regression, FSVMs, and FSVMs_CIP models can achieve ACC scores of over 90% and 80% on the CTD, MRICT, and MRID datasets containing two modalities. However, their accuracy drops significantly for MRICTD, a dataset with more than two modalities. In contrast, FLS_MK can still maintain a high score, indicating that the fuzzy classification model after high-latitude mapping is more suitable for multimodal datasets. On the other hand, RBFNN_MK has the lowest score compared to other models, indicating that only high-latitude mapping has no noticeable improvement effect on classification tasks.



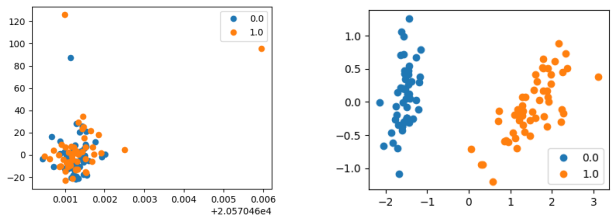
(a) ACC (b) AUC

Fig. 3: The results of FLS_MK on NPC-ContraPariod.

The AUC results in Table III reveal that Ridge Regression, FSVMs, and FSVMs_CIP models have a significant impact on the datasets containing two modalities, but their AUC score drops considerably for datasets with more than two modalities. However, FSL_MK can obtain stable scores on the four divided datasets, with the highest score of 0.9497. By comparing the AUC and ACC scores of RBFNN_MK on the four datasets, it is confirmed that the combination of fuzzy label softening classification and high-latitude mapping has an advantage.

C. Data distribution

Fig. 4 displays the sample distribution of the original NPC patient dataset and the distribution after training. The MRICTD dataset is visualized using the Principal Component Analysis (PCA) dimensionality reduction method [22]. As shown in Fig. 4(a), the original dataset is challenging to separate into two categories. However, after training, the best classification result is achieved, as demonstrated in Fig. 4(b).



(a) Before training (b) After training

Fig. 4: The sample distribution of NPC-ContraPariod MRICTD.

V. Conclusions

We proposed a novel approach called Multi-Kernel model-based Fuzzy Label Softening (FLS_MK), which utilizes kernel mapping to transform data into a high-dimensional space and employs a positive matrix to soften the binary label matrix to enhance label fitting and enlarges margins between classes. Through a series of experiments on multi-omics datasets and comparison with a baseline and other models, we demonstrated the superior performance of our approach on the NPC dataset. Our research provides a promising solution for integrating complex medical data in disease prediction, contributing to the field.

VI. Acknowledgment

We acknowledge the following participants (from Department of Clinical Oncology, Queen Elizabeth Hospital, Hong Kong, Hong Kong SAR, China) who contributed to our work by offering administrative and material support for clinical data and imaging data collection: Francis Kar Ho LEE, Celia Wai Yi YIP, and Kwok Hung AU. This work was supported in part by the Project of RISA (P0043001) of The Hong Kong Polytechnic University, Shenzhen-Hong Kong-Macau S&T Program (Category C) (SGDX20201103095002019), Shenzhen Basic Research Program (JCYJ20210324130209023) of Shenzhen Science and Technology Innovation Committee, NSF of Jiangsu Province (No. BK20201441), Jiangsu Post-doctoral Research Funding Program (No. 2020Z020), and NSFC (Grant No. 82072019).

References

[1] K. C. Wong, E. P. Hui, K.-W. Lo, W. K. J. Lam, D. Johnson, L. Li, Q. Tao, K. C. A. Chan, K.-F. To, A. D. King *et al.*, "Nasopharyngeal carcinoma: an evolving paradigm," *Nature Reviews Clinical Oncology*, vol. 18, no. 11, pp. 679–695, 2021.

[2] X.-S. Sun, X.-Y. Li, Q.-Y. Chen, L.-Q. Tang, and H.-Q. Mai, "Future of radiotherapy in nasopharyngeal carcinoma," *The British journal of radiology*, vol. 92, no. 1102, p. 20190209, 2019.

[3] F. Dionisi, S. Croci, I. Giacomelli, M. Cianchetti, A. Caldara, M. Bertolin, V. Vanoni, R. Pertile, L. Widesott, P. Farace *et al.*, "Clinical results of proton therapy reirradiation for recurrent nasopharyngeal carcinoma," *Acta Oncologica*, vol. 58, no. 9, pp. 1238–1245, 2019.

[4] W. Qu, S. Li, M. Zhang, and Q. Qiao, "Pattern and prognosis of distant metastases in nasopharyngeal carcinoma: A large-population retrospective analysis," *Cancer medicine*, vol. 9, no. 17, pp. 6147–6158, 2020.

[5] S. Chen, D. Yang, X. Liao, Y. Lu, B. Yu, M. Xu, Y. Bin, P. Zhou, Z. Yang, K. Liu *et al.*, "Failure patterns of recurrence and metastasis after intensity-modulated radiotherapy in patients with nasopharyngeal carcinoma: results of a multicentric clinical study," *Frontiers in Oncology*, vol. 11, p. 5730, 2022.

[6] S. Chakraborty, M. I. Hosen, M. Ahmed, and H. U. Shekhar, "Onco-multi-omics approach: a new frontier in cancer research," *BioMed research international*, vol. 2018, 2018.

[7] M. Olivier, R. Asmis, G. A. Hawkins, T. D. Howard, and L. A. Cox, "The need for multi-omics biomarker signatures in precision medicine," *International journal of molecular sciences*, vol. 20, no. 19, p. 4781, 2019.

[8] C. Hu and W. Jia, "Multi-omics profiling: the way toward precision medicine in metabolic diseases," *Journal of Molecular Cell Biology*, vol. 13, no. 8, pp. 576–593, 2021.

[9] Q. Zhu, B. Xu, J. Huang, H. Wang, R. Xu, W. Shao, and D. Zhang, "Deep multi-modal discriminative and interpretability network for alzheimer's disease diagnosis," *IEEE Transactions on Medical Imaging*, 2022.

[10] T. Zhou, M. Liu, K.-H. Thung, and D. Shen, "Latent representation learning for alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2411–2422, 2019.

[11] S.-K. Lam, Y. Zhang, J. Zhang, B. Li, J.-C. Sun, C. Y.-T. Liu, P.-H. Chou, X. Teng, Z.-R. Ma, R.-Y. Ni *et al.*, "Multi-organ omics-based prediction for adaptive radiation therapy eligibility in nasopharyngeal carcinoma patients undergoing concurrent chemoradiotherapy," *Frontiers in oncology*, vol. 11, p. 5406, 2022.

[12] M. Jing, J. Zhao, J. Li, L. Zhu, Y. Yang, and H. T. Shen, "Adaptive component embedding for domain adaptation," *IEEE transactions on cybernetics*, vol. 51, no. 7, pp. 3390–3403, 2020.

[13] Y. Tian and X. Feng, "Large margin graph embedding-based discriminant dimensionality reduction," *Scientific Programming*, vol. 2021, pp. 1–12, 2021.

[14] Z. Kang, C. Peng, Q. Cheng, X. Liu, X. Peng, Z. Xu, and L. Tian, "Structured graph learning for clustering and semi-supervised classification," *Pattern Recognition*, vol. 110, p. 107627, 2021.

[15] Z. Song, X. Yang, Z. Xu, and I. King, "Graph-based semi-supervised learning: A comprehensive review," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[16] A. Zwanenburg, M. Vallières, M. A. Abdalah, H. J. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R. J. Beukinga, R. Boellaard *et al.*, "The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology*, vol. 295, no. 2, pp. 328–338, 2020.

[17] S. M. Atif, S. Khan, I. Naseem, R. Togneri, and M. Bennamoun, "Multi-kernel fusion for rbf neural networks," *Neural Processing Letters*, pp. 1–25, 2022.

[18] X. Fang, Y. Xu, X. Li, Z. Lai, W. K. Wong, and B. Fang, "Regularized label relaxation linear regression," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 4, pp. 1006–1018, 2017.

[19] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines," *IEEE transactions on neural networks*, vol. 13, no. 2, pp. 464–471, 2002.

[20] R. Batuwita and V. Palade, "Fsvm-cil: fuzzy support vector machines for class imbalance learning," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 3, pp. 558–571, 2010.

[21] G. C. McDonald, "Ridge regression," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 1, no. 1, pp. 93–100, 2009.

[22] T. Kurita, "Principal component analysis (pca)," *Computer Vision: A Reference Guide*, pp. 1–4, 2019.