Hindawi
Computational and Mathematical Methods in Medicine
Volume 2020, Article ID 7562140, 15 pages
https://doi.org/10.1155/2020/7562140

Research Article

A Collaborative Dictionary Learning Model for Nasopharyngeal Carcinoma Segmentation on Multimodalities MR Sequences

Haiyan Wang,[1] Guoqiang Han,[1] Haojiang Li,[2] Guihua Tao,[1] Enhong Zhuo,[1] Lizhi Liu,[2] Hongmin Cai,[1] and Yangming Ou[3]

[1]*School of Computer Science and Engineering, South China University of Technology, 510000, China*
[2]*Department of Radiology, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, 510060 Guangdong, China*
[3]*Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA*

Correspondence should be addressed to Hongmin Cai; hmcai@scut.edu.cn

Received 16 June 2020; Revised 6 August 2020; Accepted 12 August 2020; Published 28 August 2020

Academic Editor: Nadia A. Chuzhanova

Nasopharyngeal carcinoma (NPC) is the most common malignant tumor of the nasopharynx. The delicate nature of the nasopharyngeal structures means that noninvasive magnetic resonance imaging (MRI) is the preferred diagnostic technique for NPC. However, NPC is a typically infiltrative tumor, usually with a small volume, and thus, it remains challenging to discriminate it from tightly connected surrounding tissues. To address this issue, this study proposes a voxel-wise discriminate method for locating and segmenting NPC from normal tissues in MRI sequences. The located NPC is refined to obtain its accurate segmentation results by an original multiviewed collaborative dictionary classification (CODL) model. The proposed CODL reconstructs a latent intact space and equips it with discriminative power for the collective multiview analysis task. Experiments on synthetic data demonstrate that CODL is capable of finding a discriminative space for multiview orthogonal data. We then evaluated the method on real NPC. Experimental results show that CODL could accurately discriminate and localize NPCs of different volumes. This method achieved superior performances in segmenting NPC compared with benchmark methods. Robust segmentation results show that CODL can effectively assist clinicians in locating NPC.

1. Introduction

Nasopharyngeal carcinoma (NPC) is an enigmatic malignancy with marked racial and geographical differences, being particularly prevalent in southern China, Southeast Asia, and northern Africa [1, 2]. Although advances in therapeutic techniques have contributed to improve clinical outcomes for patients with NPC, the mortality rate remains high. Early detection and accurate tumor localization of NPC are vital for surgical planning. Magnetic resonance imaging (MRI) is the first choice in primary tumor delineatio and a presurgical tool for localization and evaluation of the tumor entity [3–5]. In practice, the patient is usually scanned by T1-weighted (T1-w) or T2-weighted (T2-w) MR imaging. The T2-weighted (T2-w) imaging provides better fine structural information on soft tissues than by T1-w imaging. A contrast-enhanced T1-weighted (CET1-w) imaging is sometimes operated to provide direct evidence on tumor occurrence. Currently, identification and comprehensive assessment of the carcinoma entity NPC remain a great challenge. The infiltrative and migratory characteristics of NPC make it difficult to be discriminated from surrounding tissues.

To achieve automatic (or semiautomatic) segmentation of the NPC, traditional image processing has been used to fulfill the task. For example, [6] proposed a semiautomatic workflow, including masking, thresholding, and seed growing, to segment NPC from both T2-w and CET1-w from 7 patients to help radiation therapy. [7] proposed an automatic NPC segmentation method based on region growing and clustering and used neural networks to classify suspicious regions. [8] proposed to use a genetic algorithm for selecting the informative features and the support vector machine for classifying NPC. With the great

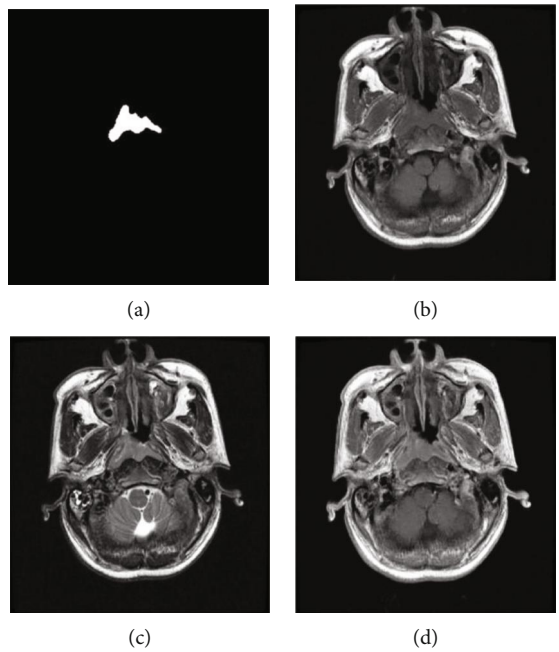

Figure 1: Example MR slices with three sequences. From left to right: (a) ground truth, (b) T1-w, (c) T2-w, and (d) CET1-w.

success of deep learning models in computer vision, [9] proposed to use deep convolutional neural networks and graph cut on T1-w images from 30 NPC patients. [10] tested a deep deconvolutional neural network, composing of an encoder network and a decoder network, on CT images from 230 patients. [11] reported an automatic NPC segmentation method based on the convolutional neural network (CNN) architecture with dynamic contrast-enhanced MRI. [12] used fully convolutional networks with auxiliary paths to achieve automatic segmentation of NPC on PET-CT images. [13] used a modified U-Net model to automatically segment NPC on CT images from 502 patients. [14] proposed an automated method based on CNN for NPC segmentation on dual-sequence MRI (i.e., T1-w and T2-w) from 44 patients. Furthermore, the tumor volume varies greatly and many of them are small. Such sample characteristics raise a large difficulty in constructing representative learning models using deep networks.

Recently, multiview learning models have been developed to analyze images from various imaging modalities or views. Fruitful advances have been made in reconstruction, face recognition, human motion recognition, and other object recognition issues [15–17]. In the current study, each patient underwent MRI by three sequences (i.e., T1-w, T2-w, and CET1-w) to enjoy the merits of different imaging characteristics (see Figure 1). The study is aimed at achieving the identification and segmentation of the NPC with high accuracy. Different views usually provide supplemental information. The problem of NPC segmentation can be formulated as a voxel-wise dictionary learning problem with three different views.

However, existing multiview learning methods cannot be tailored directly to be applied in NPC localization and segmentation. From the methodological aspect, most NPCs only occupy a small area in the entire slice. Such imbalance also results in a high false positive rate in applying learning models directly. To solve this difficulty, we preprocessed the data, that is, using a specially designed deep learning model with a fully convolutional network (FCN) structure to roughly locate the suspicious tumor area. In light of the advantages of multiview subspace learning, we propose to use a multiview learning collaborative dictionary model, which we call CODL, to further refine the detailed structure of NPC. The flowchart of NPC segmentation is illustrated in Figure 2.

The major contributions of our work are as follows:

(1) An original collaborative dictionary model for multiview learning (CODL) is proposed to achieve fine segmentation. The CODL integrates cooperative information from multiple views to find latent intact space for the data and renders the latent space discriminative. The latent space is constructed by collaborative dictionary learning incorporating membership to possess discriminative power. Our approach takes into account the label of the samples to latent intact space. This gives a consistent indicator matrix discriminative capability

(2) The numerical scheme involved in solving the CODL is provided. It treated the proposed unified framework into solvable subproblems, each with an explicit solution and a fast computation

(3) While using all three MR sequences (T1-w, T2-w, and CET1-w) achieved the highest accuracy, we show that, for patients having kidney diseases that prevent the use of contrast agent necessary in CET1-w imaging [18, 19], using T1-w and T2-w alone does not significantly undermine the segmentation accuracy. This

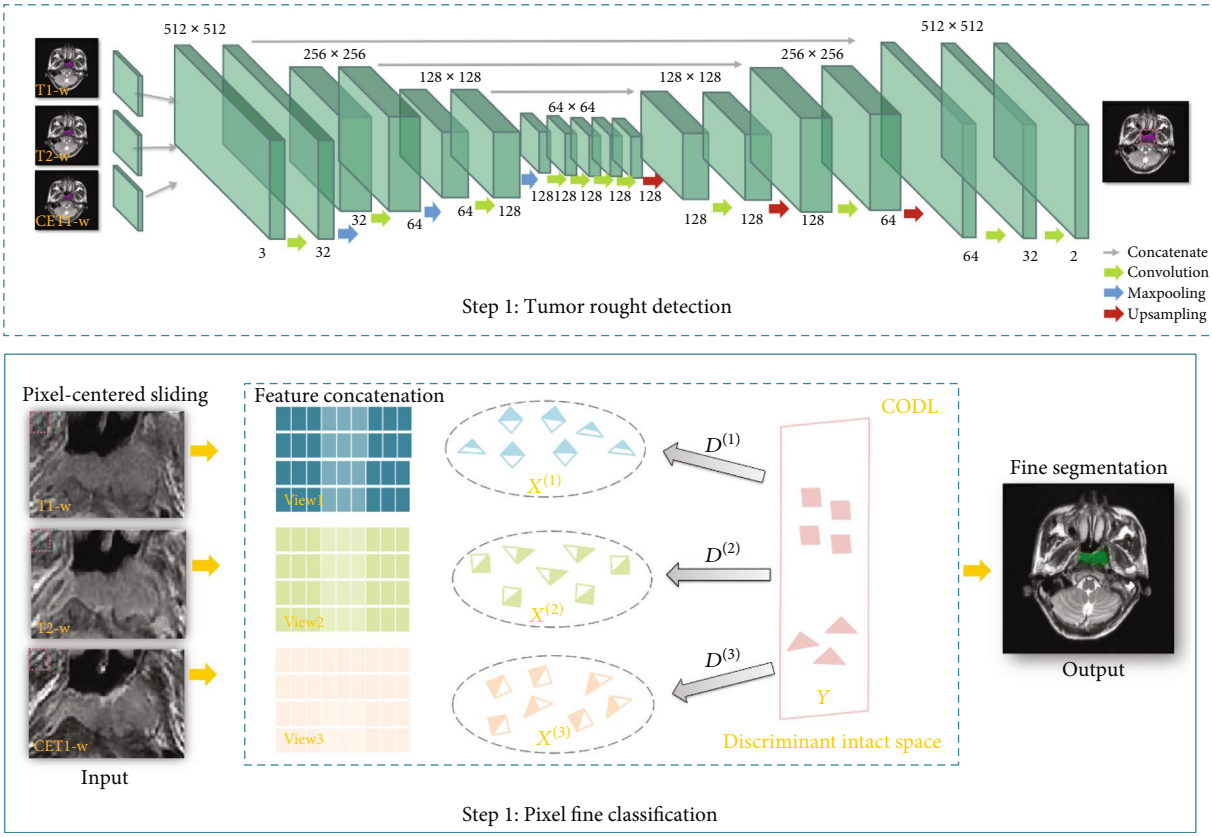

Figure 2: The workflow of location and segmentation of NPC. It consists of two steps, rough location by FCN and pixel-wise fine classification by CODL.

highlights the sensitivity and stability of the proposed CODL algorithm and improves the applicability of the proposed framework

2. Literature Review

Sparse codes generated by dictionaries can be directly used as features to train classifiers for recognition [20]. This two-stage scheme has been extensively used in many existing methods [21–25], such method uses the discriminative nature of sparse representation to perform classification. However, generated sparse codes are often insufficiently discriminative in complex recognition tasks [26]. One alternative is to unify dictionary learning and classifier training in an optimization model [27]. However, most of the supervised dictionary methods only employ single-view information in the learning process, which will result in the data not having the optimal expressibility. Besides, the model will also depend on the peculiarities of training data.

A naive way of multiview learning is feature fusion [28]. However, consolidating each single view may be suboptimal if the different views belong to different sample spaces. To address the drawback, weighted combinations [29] have been proposed. Alternatively, recent advances are aimed at learning the multiview data via finding an intact space, such as the multiview intact space learning (MISL) [15] and the multiview discriminant analysis with view consistency (MvDA-VC) [16]. In such approaches, a latent subspace shared by multiple views is learned by assuming that the input views are generated from this latent subspace. For multiview intact space learning, however, class membership is seldom used to find the optimal latent subspace and has little power available to handle problems in supervised learning.

As the goal of this paper is to develop a multiview dictionary learning method for voxel-wise classification. We first give a brief review of the supervised dictionary and multiview subspace learning methods related to our work. In this paper, we propose a novel collaborative dictionary model for multiview learning, which also takes into account the label of the samples to latent intact space. The construction of the latent space is guided by the supervised dictionary learning within each individual view and equipped to have discriminative power.

3. Materials and Methods

3.1. Dataset. A total of 24 patients with nonmetastatic NPC at the Sun Yat-sen University Cancer Center (SYSUCC) were enrolled in this study. MRI was performed on a 3.0-T scanner (Discovery MR750; GE Healthcare, Milwaukee, WI, USA). The imaging parameters are as follows: axial T1-w imaging (FSE, TR = 540 ms, and TE = 11.8 ms), axial T2-w imaging (FSE, TR = 4000 ms, and TE = 99 ms), and axial CET1-w imaging (FSE, TR = 540 ms, and TE = 11.8 ms). The number of slices per patient was 16, 32, or 36. Not every layer of MR images has lesions. The interval between each layer of images is 5 mm, in which imaging has a high resolution of 0.43 mm

$\times 0.43$ mm. T1-w, T2-w, and CET1-w MR sequences were assessed for each patient. Regions of interest (ROI) were drawn by four experienced radiologists (>3 years of clinical experience) using semiautomatic methods. They were required to draw all discernable tumor regions cautiously along axial directions. Any disagreements were resolved through negotiating until full consent was derived by the four.

The purpose of this study is to develop a multiview dictionary learning method for voxel-wise classification. We first give a rigid quality control on the selection of slices. Following the principle of multiple modalities sequences alignment, in total, 90 slices covering 30 instances of distinct tumor sizes were selected for our experiment. Each instance has three MR sequences (i.e., T1-w, T2-w, and CET1-w) and well-aligned before feeding into models.

3.2. A Collaborative Dictionary Model for Multiview Classification (CODL). In this paper, we proposed a collaborative multiview learning model to fuse multiple image modalities into a consolidated space. By integrating each single modality and exploiting its characteristics comprehensively, the information among different modalities is actively learned and reinterpreted in a latent space. The supervised membership is used to render the latent space being discriminative, and thus, the sample classification is finally conducted within the learned latent space.

3.2.1. Formulation of Multiview Collaborative Classification Model. Mathematically, let $X^{(v)} = [x_1^{(v)}, x_2^{(v)}, \cdots, x_s^{(v)}] \in \mathbb{R}^{n\times s} (v = 1, 2, \cdots, m)$ denote a dataset containing s samples from the vth view, with each sample characterized by a n-dimensional vector. We want to consolidate the multiview data into a latent space, denoted by $Y = [y_1, y_2, \cdots, y_s] \in \mathbb{R}^{d\times s}$, where d is the dimensionality of the latent space.

Let $D^{(v)} \in \mathbb{R}^{n\times d} (v = 1, 2, \cdots, m)$ denote the dictionary learned in the vth view. The label for the training samples is denoted by L. Our aim is to learn an informative latent space from multiple modalities and then achieve accurate classification task within the latent space. To this end, we proposed the following model to achieve latent space learning and classification simultaneously.

$$\underset{Y, D^{(v)}, \beta}{\operatorname{argmin}} \sum_{v=1}^{m} \frac{1}{2} \left\| X^{(v)} - D^{(v)} Y \right\|_2^2 + \frac{1}{2} \lambda_1 \left\| L - Y^T \beta \right\|_2^2 + \lambda_2 \|\beta\|_1. \tag{1}$$

The first term in Equation (1) controls data fidelity by minimizing the reconstruction errors in the latent space Y through the dictionary $D^{(v)}$. The second term renders the latent space with discriminative power. The two terms work collaboratively to yield a sharable latent space for different views. The third term encourages the loading coefficient β to be sparse to achieve economic expression. Besides, it also helps to stabilize the optimization due to large freedom in the objective function. The hyperparameters λ_1 and λ_2 are aimed at penalizing the reconstruction error and sparsity.

Once we obtain the learned dictionaries $D^{(v)} (v = 1, 2, \cdots, m)$ and the latent space Y, we can map a query sample $q_i \in \mathbb{R}^n$ to its representation $\widehat{q} \in \mathbb{R}^d$ in the latent space. The latent representation $\widehat{q}$ is estimated by minimizing the following energy function:

$$\underset{\widehat{q}}{\operatorname{argmin}} \sum_{v=1}^{m} \frac{1}{2} \left\| q_i - D^{(v)} \widehat{q} \right\|_2^2. \tag{2}$$

Finally, we can classify the sample $\widehat{q}$ in the latent space Y using benchmark classification models, e.g., k-nearest neighbor.

The proposed CODL not only integrates complementary information in multiple views to find a latent intact space for the data but also renders the latent space discriminative.

3.2.2. Numerical Scheme for Solving CODL. The objective function Equation (1) is convex with respect to $D^{(v)}$ and Y. Therefore, we used a heuristic alternating direction method to solve it. By minimizing one variable while fixing the others, the alternating direction method iteratively updates each variable until convergence. The alternate minimization method enjoys an excellent characteristic. It can decompose a large complex problem into small-sized subproblems, thus enabling parallel solving to have a quick convergence. In particular to our problem, it decomposes Equation (1) into three subproblems with respect to the three variables $D^{(v)}$, Y, and β.

Step 1 to update $D^{(v)}$: by fixing Y and β and discarding irrelevant terms, the objective function Equation (1) could be simplified as

$$\underset{D^{(v)}}{\operatorname{argmin}} \sum_{v=1}^{m} \frac{1}{2} \left\| X^{(v)} - D^{(v)} Y \right\|_2^2. \tag{3}$$

It is convex and differentiable with respect to the variable $D^{(v)}$. By setting the gradient to zero, one has an explicit solution:

$$D^{(v)} = X^{(v)} Y^T \left(Y Y^T \right)^{-1}. \tag{4}$$

Step 2 to update β: by fixing the variables $D^{(v)}$ and Y, the objective function Equation (1) could be simplified as:

$$\underset{\beta}{\operatorname{argmin}} \frac{1}{2} \lambda_1 \left\| L - Y^T \beta \right\|_2^2 + \lambda_2 \left\| \beta \right\|_1. \tag{5}$$

It resembles the classical least absolute shrinkage and selection operator (LASSO) problem. By using a proximal gradient, its solution could be obtained by the iterative soft-thresholding algorithm (ISTA) [30]:

$$\beta^k = \underset{\beta}{\operatorname{argmin}} \frac{1}{2} \left\| L - Y^T \beta \right\|_2^2 + \|\beta\|_1 = S_{(\lambda_1/\lambda_2)t} \left(\beta^{(k-1)} + tY \left(L - Y^T \beta^{(k-1)} \right) \right), \tag{6}$$

where t is the step size and $S_{\lambda t}(\beta)$ is the soft-thresholding operator. One could further accelerate the ISTA to achieve

Input: $X = \{X^{(v)} \mid 1 \le v \le m\}$, L, λ_1, λ_2.
1: Initialize $D^{(v)}$, Y, and β;
2: **repeat**
3: Update $D^{(v)}$ for $(v = 1, 2, \cdots, m)$ by solving Equation (3);
4: Update β by solving Equation (5);
5: Update Y by solving Equation (8);
6: **until** convergence
7: Get $\widehat{q}$ by solving subproblem Equation (2);
8: Get classification result by applying k-Nearest Neighbor.
Output: The classification result.

Algorithm 1: The algorithm for solving CODL.

fast convergence

$$\beta^{(k)} = S_{(\lambda_1/\lambda_2)t}\left(g + tY\left(L - Y^T g\right)\right), g = \beta^{(k-1)} + \frac{k-2}{k+1}\left(\beta^{(k-1)} - \beta^{(k-2)}\right). \tag{7}$$

Step 3 to update Y: by fixing $D^{(v)}$ and β, the objective function Equation (1) could be simplified as

$$\underset{Y}{\operatorname{argmin}} \sum_{v=1}^{m} \frac{1}{2}\left\|X^{(v)} - D^{(v)}Y\right\|_2^2 + \frac{1}{2}\lambda_1\left\|L - Y^T\beta\right\|_2^2. \tag{8}$$

Setting the gradient with respect to Y to be zero, one has

$$Y = \left(\sum_{v=1}^{m}\left(D^{(v)}\right)^T D^{(v)} + \lambda_1\beta\beta^T\right)^{-1}\left(\sum_{v=1}^{m}\left(D^{(v)}\right)^T X^{(v)} + \lambda_1\beta L^T\right). \tag{9}$$

The above three schemes are iteratively updated until convergence.

In the testing phase, one needs to find the new representation $\widehat{q}$ for query samples q through the dictionary $D^{(v)}$ by solving Equation (2). It is a standard least square minimization problem with an explicit solution

$$\widehat{q} = \left[\sum_{v=1}^{m}\left(D^{(v)}D^{(v)}\right)\right]^{-1}\sum_{v=1}^{m} D^{(v)}q. \tag{10}$$

The pseudocode for solving CODL is provided in Algorithm 1.

3.2.3. Complexity Analysis. The computational time of solving the proposed model is mainly taken by updating the $D^{(v)}$, β, and Y. As mentioned in Section 3.2.1, $D^{(v)} \in \mathbb{R}^{n\times d}$, $\beta \in \mathbb{R}^{d\times 1}$, and $Y \in \mathbb{R}^{d\times s}$, where n is the dimensionality of the vth view, d is the dimensionality of the latent space, and s is the number of multiview objects. According to Algorithm 1, the main computational cost of CODL is incurred in the iterative calculations of $D^{(v)}$, β, and Y. In each inner iteration, the computational cost of solving $D^{(v)}$ by Equation (4) is $O(nsd + d^2s + d^3 + nd^2)$, the computational cost of solving β by Equation (6) is $O(d^3 + d^2s)$, and the computational cost of solving Y via Equation (9) is $O(d^2n + d^2s + dns + d^3)$. Therefore, the total computational complexity is $O(dns + d^2n + d^2s + d^3)$.

4. Experiments and Results

We applied the proposed model on both a synthetic dataset and a real NPC dataset. For a fair comparison, each method was run on the synthetic data 10 times, and the averaged results were recorded. On a real NPC dataset, we tested the performance of each method using 10-fold cross-validation scheme. Classification accuracy was measured in terms of average accuracy across ten trials on different training and testing sets. Moreover, the parameters in each compared method are tuned to meet the best performance in the suggested range. For CODL, we empirically set the parameters, that is, $\lambda_1 = 0.01$, $\lambda_2 = 0.7$ for single view, $\lambda_1 = 1.0$, $\lambda_2 = 0.2$ for two views, and $\lambda_1 = 3.8$, $\lambda_2 = 0.2$ for three views, throughout all experiments. All of our experiments were performed on a desktop computer with a 4.20 GHz Intel(R) Core (TM)i7-7700K CPU, 16.0 GB of RAM, and MATLAB R2017a (×64).

4.1. Evaluation Metrics and Baseline Methods for Performance Comparisons. Six widely used metrics, including the sensitivity (SENS), the dice similarity coefficient (DICE), the area under the receiver operating characteristic curve (AUC), intersection over union (IoU), mean pixel accuracy (MPA), and Hausdorff distance (HD) were employed to measure the performances of each tested method. These qualitative metrics were defined as follows:

$$\begin{aligned} \text{SENS} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{DICE} &= \frac{2\text{TP}}{\text{TP} + \text{FN} + \text{TP} + \text{FP}}, \\ \text{IoU} &= \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}, \\ \text{MPA} &= \frac{\text{TPR} + \text{TNR}}{2}, \end{aligned} \tag{11}$$

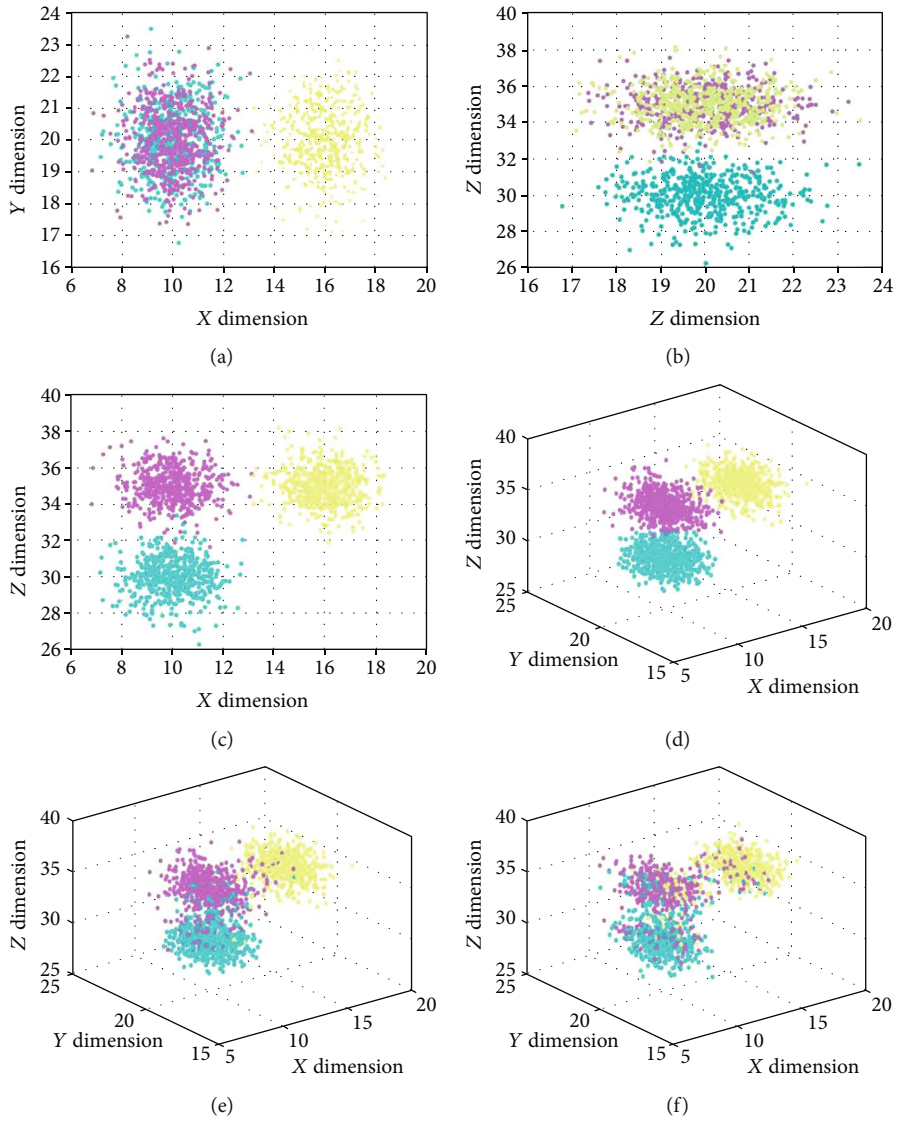

FIGURE 3: A toy example to demonstrate the discrimination power of the CODL. The data is collected from three views on (a) X-Y plane, (b) Y-Z plane, and (c) X-Z plane. The reconstructed results in X-Y-Z space by CODL on (d) the intact noiseless data, (e) the noisy data with std = 0.5, and (f) its std = 1 noisy counterpart are also shown. Different classes are highlighted in different colors.

where TP, FP, TN, FN, TPR, and TNR represented true positive, false positive, true negative, false negative, true positive rate, and true negative rate, respectively. We also plotted the receiver operating characteristic curve (ROC) for each method. The area under the ROC curve (AUC) was then estimated. For two point sets A and B, the Hausdorff distance between these two sets is defined as follows:

$$\mathrm{HD}(A, B) = \max\left(\mathrm{hd}(A, B), \mathrm{hd}(B, A)\right), \tag{12}$$

where $\mathrm{hd}(A, B) = \max_{x \in X} \min_{y \in Y} \|x - y\|_2$, $\mathrm{hd}(B, A) = \max_{y \in Y} \min_{x \in X} \|x - y\|_2$. For this study, we have used the Euclidean norm $\|x - y\|_2$.

Several benchmark methods are borrowed to serve as baseline methods for comparisons. They are widely used multiview methods and most relevant to our method.

(i) Support vector machine (SVM) [31]: we concatenate the features of all views and perform support vector machine classification

(ii) Multiview intact space learning (MISL) [15]: it is aimed at integrating the encoded complementary information from different views into a latent intact space. It shows theoretically that combining multiple views can obtain abundant information for latent intact space learning

(iii) Multiview discriminant analysis with view consistency (MvDA-VC) [16]: it seeks for a single discriminant common space for multiple views in a nonpairwise manner by jointly learning multiple view-specific linear transforms. MvDA-VC method has achieved good performance in addressing the problem of object recognition from multiple views

Table 1: The performance of different methods on noisy synthetic datasets (mean ± standard deviation).

Method	Noise free	std = 0.25	std = 0.5	std = 1
SVM (V1)	0.664 ± 0.007	0.647 ± 0.008	0.614 ± 0.009	0.527 ± 0.013
SVM (V2)	0.606 ± 0.008	0.563 ± 0.009	0.561 ± 0.008	0.462 ± 0.034
SVM (V3)	0.938 ± 0.016	0.932 ± 0.016	0.843 ± 0.018	0.671 ± 0.021
CODL (V1)	0.668 ± 0.011	0.659 ± 0.007	0.621 ± 0.010	0.523 ± 0.014
CODL (V2)	0.664 ± 0.011	0.648 ± 0.013	0.591 ± 0.011	0.496 ± 0.012
CODL (V3)	0.994 ± 0.002	0.969 ± 0.005	0.874 ± 0.008	0.679 ± 0.014
SVM (FeaConcat)	0.946 ± 0.007	0.912 ± 0.007	0.843 ± 0.021	0.735 ± 0.042
MISL (fusion)	0.984 ± 0.004	0.966 ± 0.006	0.908 ± 0.011	0.733 ± 0.013
MvDA-VC (fusion)	0.995 ± 0.002	0.968 ± 0.005	0.876 ± 0.010	0.675 ± 0.011
CODL (fusion)	0.987 ± 0.002	0.971 ± 0.004	0.911 ± 0.006	0.749 ± 0.011

*The V1, V2, and V3 denote X-Y, Y-Z, and X-Z views, respectively. *FeaConcat means that we concatenate features of all views to generate a combined feature. *Fusion means that we construct a multiview latent intact space learning by fusing all individual views. *Classification performance is measured in terms of average accuracy.

Table 2: Architecture of the FCN network for tumor localization.

	Type	Input size	Output size	Filter size	Stride	# filters
Layer 1	Conv.	512 × 512 × 3	512 × 512 × 32	3 × 3	1 × 1	32
Layer 2	Max-pool.	512 × 512 × 32	256 × 256 × 32	2 × 2	2 × 2	—
Layer 3	Conv.	256 × 256 × 32	256 × 256 × 64	5 × 5	1 × 1	64
Layer 4	Max-pool.	256 × 256 × 64	128 × 128 × 64	2 × 2	2 × 2	—
Layer 5	Conv.	128 × 128 × 64	128 × 128 × 128	7 × 7	1 × 1	128
Layer 6	Max-pool.	128 × 128 × 128	64 × 64 × 128	2 × 2	2 × 2	—
Layer 7	Conv.	64 × 64 × 128	64 × 64 × 128	3 × 3	1 × 1	128
Layer 8	Conv.	64 × 64 × 128	64 × 64 × 128	3 × 3	1 × 1	128
Layer 9	Conv.	64 × 64 × 128	64 × 64 × 128	3 × 3	1 × 1	128
Layer 10	Conv.	64 × 64 × 128	64 × 64 × 128	3 × 3	1 × 1	128
Layer 11	Upsampling	64 × 64 × 128	128 × 128 × 128	2 × 2	2 × 2	—
Layer 12	Conv.	128 × 128 × 128	128 × 128 × 128	7 × 7	1 × 1	128
Layer 13	Upsampling	128 × 128 × 128	256 × 256 × 128	2 × 2	2 × 2	—
Layer 14	Conv.	256 × 256 × 128	256 × 256 × 64	5 × 5	1 × 1	64
Layer 15	Upsampling	256 × 256 × 64	512 × 512 × 64	2 × 2	2 × 2	—
Layer 16	Conv.	512 × 512 × 64	512 × 512 × 32	3 × 3	1 × 1	32
Layer 17	Conv.	512 × 512 × 32	512 × 512 × 2	1 × 1	1 × 1	2

*The convolutional layer is denoted by Conv., and the max pooling by max-pool.

(a) Zhao et al. [12]: it uses fully convolutional networks with an auxiliary path to achieve automatic segmentation of NPC on dual-modality PET-CT images. The proposed method improves NPC segmentation by guiding the training of lower layers by auxiliary paths

(b) Li et al. [13]: it proposes a modified version of the U-Net, which performs well on NPC segmentation by modifying the downsampling layers and upsampling layers to have a similar learning ability and predict the same spatial resolution as the source image

4.2. Discriminative Capability Tests of CODL on Synthetic Data. We first constructed a synthetic data to test the discrimination power of the proposed methods. The synthetic data consisted of three classes, and they were separable within a three-dimensional space, but inseparable when projected orthogonally into two-dimensional (2D) plane (i.e., X-Y and Y-Z planes). The projected samples into each 2D plane were

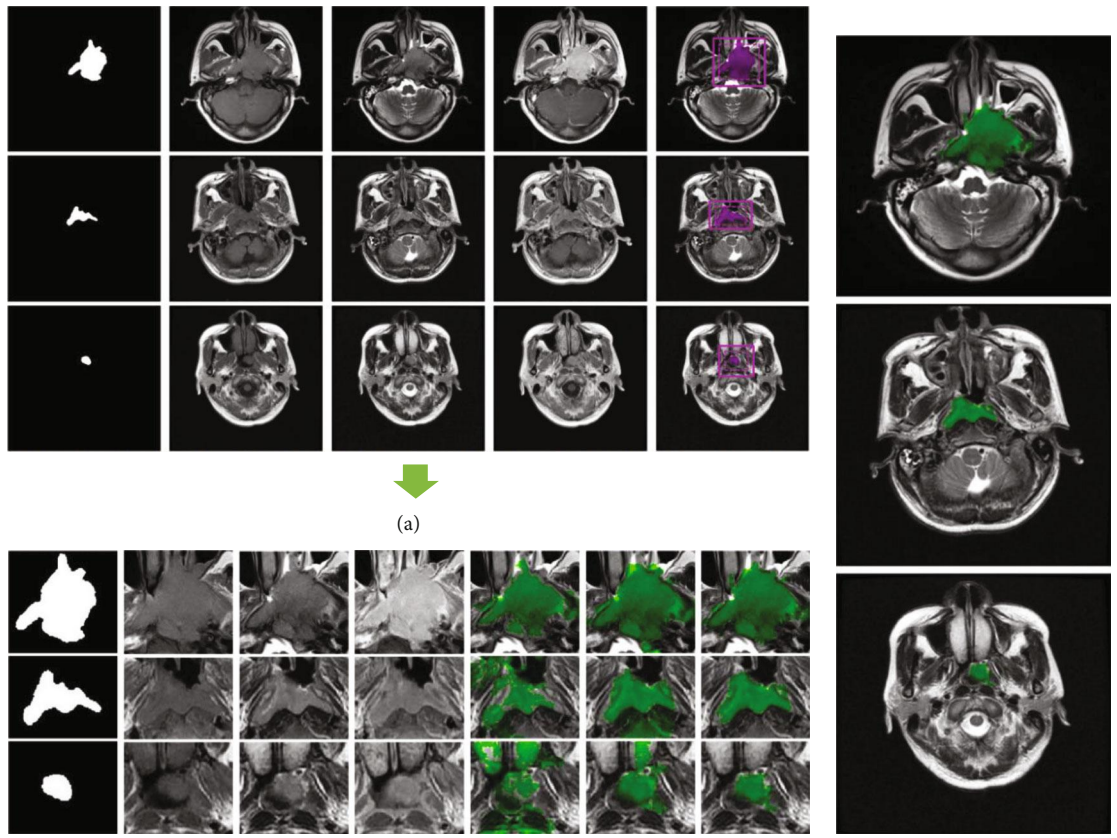

Figure 4: NPC segmentation results on three typical examples. (a) Rough location results with bounding boxes identified by FCN, highlighted in red dots. The extended areas used for fine classification were indicated by solid red lines. (b) Fine segmentation results with fusing T1-w, T2-w, and CET1-w MR sequences. The last three columns are the tumor regions located by MISL, MvDA-VC, and CODL, respectively. (c) Results of our method on the whole slice in case of a combination of three modalities.

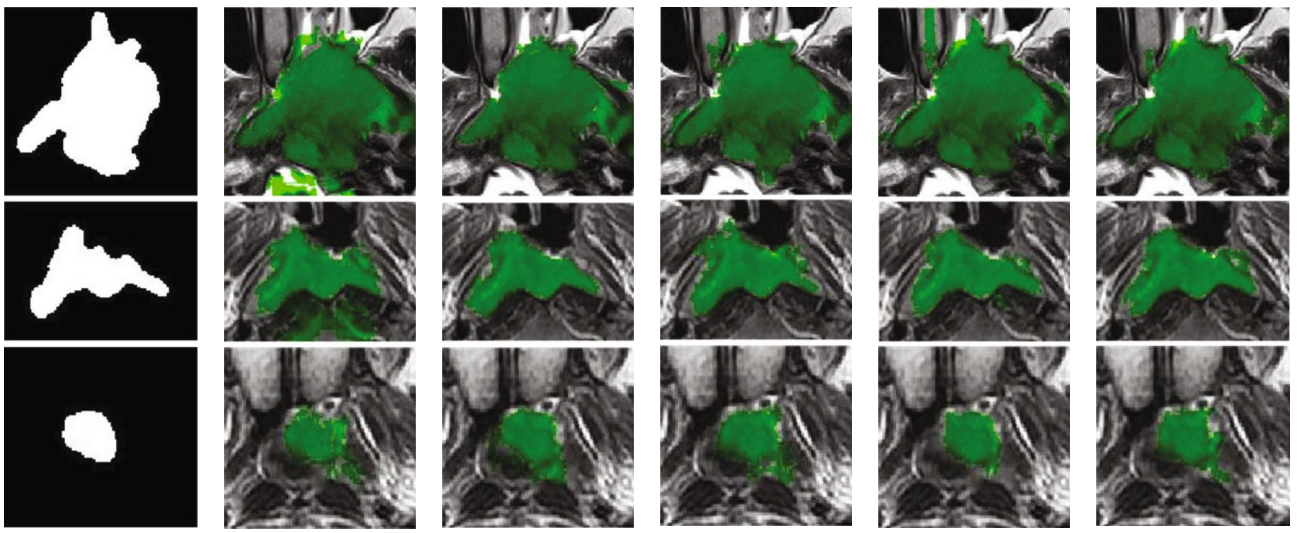

Figure 5: Typical segmentation results of three instances using CODL. (a) Ground truth. From second to last column: the tumor regions identified by CODL on modality T1-w (b), T2-w (c), CET1-w (d), both T1-w and T2-w (e), and T1-w, T2-w, and CET1-w (f), respectively.

Table 3: Metric results (mean ± standard deviation) of different methods on the cropped NPC dataset.

Method	SENS	DICE	AUC	IoU	MPA	HD
SVM (T1-w)	0.570 ± 0.292	0.558 ± 0.237	0.729 ± 0.107	0.419 ± 0.199	0.728 ± 0.113	28.566 ± 13.664
SVM (T2-w)	0.609 ± 0.269	0.652 ± 0.229	0.790 ± 0.144	0.518 ± 0.216	0.780 ± 0.134	23.212 ± 12.273
SVM (CET1-w)	0.733 ± 0.158	0.731 ± 0.093	0.827 ± 0.078	0.584 ± 0.112	0.829 ± 0.075	20.979 ± 8.505
CODL (T1-w)	0.832 ± 0.118	0.733 ± 0.088	0.847 ± 0.055	0.586 ± 0.110	0.847 ± 0.054	23.241 ± 9.168
CODL (T2-w)	0.812 ± 0.138	0.745 ± 0.119	0.860 ± 0.090	0.607 ± 0.145	0.854 ± 0.077	23.057 ± 10.456
CODL (CET1-w)	0.828 ± 0.102	0.767 ± 0.074	0.868 ± 0.050	0.627 ± 0.094	0.864 ± 0.047	22.827 ± 10.103
SVM (FeaConcat2)	0.377 ± 0.182	0.505 ± 0.204	0.692 ± 0.087	0.360 ± 0.168	0.682 ± 0.091	24.470 ± 13.020
MISL (Fusion2)	0.412 ± 0.254	0.310 ± 0.172	0.531 ± 0.136	0.195 ± 0.121	0.530 ± 0.115	38.052 ± 8.738
MvDA-VC (Fusion2)	0.901 ± 0.072	0.718 ± 0.090	0.853 ± 0.046	0.567 ± 0.111	0.858 ± 0.043	24.807 ± 7.010
CODL (Fusion2)	0.827 ± 0.094	0.808 ± 0.075	0.886 ± 0.055	0.683 ± 0.099	0.877 ± 0.052	16.618 ± 9.524
SVM (FeaConcat3)	0.211 ± 0.139	0.327 ± 0.194	0.611 ± 0.086	0.211 ± 0.139	0.606 ± 0.070	31.288 ± 16.200
MISL (Fusion3)	0.530 ± 0.262	0.452 ± 0.219	0.680 ± 0.123	0.317 ± 0.185	0.667 ± 0.125	32.490 ± 9.290
MvDA-VC (Fusion3)	0.893 ± 0.086	0.713 ± 0.094	0.846 ± 0.055	0.562 ± 0.116	0.853 ± 0.050	24.918 ± 7.243
CODL (Fusion3)	0.836 ± 0.111	0.820 ± 0.062	0.889 ± 0.054	0.699 ± 0.087	0.885 ± 0.049	16.683 ± 9.447

*FeaConcat2 and Fusion2 denote concatenating and fusing T1-w and T2-w, respectively. *FeaConcat3 and Fusion3 denote concatenating and fusing T1-w, T2-w, and CET1-w, respectively.

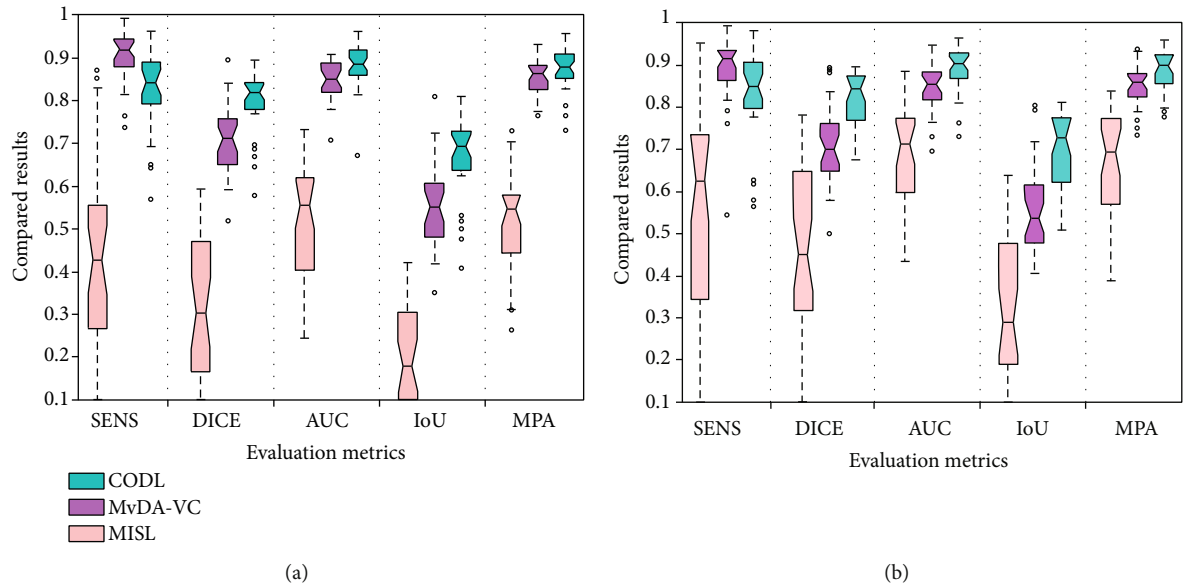

Figure 6: Quantitative results of MISL, MvDA-VC, and CODL on multiple sequences of (a) T1-w and T2-w and (b) T1-w, T2-w, and CET1-w.

considered an observed individual view. The synthetic data contained 3000 samples from three classes, each following a multivariate normal distribution with mean values $\mu_1 = (10\ 20\ 30)$, $\mu_2 = (10\ 20\ 35)$, $\mu_3 = (16\ 20\ 35)$, and covariances

$$\sum = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (13)$$

respectively.

To test the robustness of the model over noise contaminations, the synthetic data were corrupted by Gaussian white noises with a standard deviation of 0.25, 0.5, and 1, respectively. The synthetic data was shown in Figure 3. The first row was the three different views along different planes (i.e., X-Y, Y-Z, and X-Z planes), respectively. The corresponding classified results by the proposed CODL were shown in the second row of Figure 3. Classification performance was measured in terms of average accuracy across ten trials. The percentage of training sets and test sets in each trial is 1 : 1. The averaged results were recorded and summarized in Table 1.

Table 4: Metric results (mean ± standard deviation) of different methods on the whole slices.

Method	SENS	DICE	AUC	IoU	MPA	HD
Zhao et al. [12] (Fusion2)	0.723 ± 0.242	0.662 ± 0.160	0.814 ± 0.121	0.511 ± 0.149	0.858 ± 0.119	31.365 ± 19.268
Li et al. [13] (Fusion2)	0.469 ± 0.338	0.523 ± 0.300	0.723 ± 0.179	0.407 ± 0.274	0.734 ± 0.168	38.769 ± 28.383
Ours (Fusion2)	0.823 ± 0.096	0.804 ± 0.077	0.908 ± 0.048	0.678 ± 0.100	0.910 ± 0.048	16.918 ± 9.553
Zhao et al. [12] (Fusion3)	0.713 ± 0.223	0.664 ± 0.165	0.806 ± 0.125	0.518 ± 0.178	0.854 ± 0.110	30.388 ± 11.953
Li et al. [13] (Fusion3)	0.689 ± 0.237	0.741 ± 0.197	0.826 ± 0.128	0.618 ± 0.195	0.844 ± 0.118	21.928 ± 11.037
Ours (Fusion3)	0.828 ± 0.109	0.813 ± 0.066	0.910 ± 0.060	0.690 ± 0.090	0.913 ± 0.054	16.895 ± 9.624

*Fusion2 denote fusing T1-w and T2-w. *Fusion3 denote fusing T1-w, T2-w, and CET1-w.

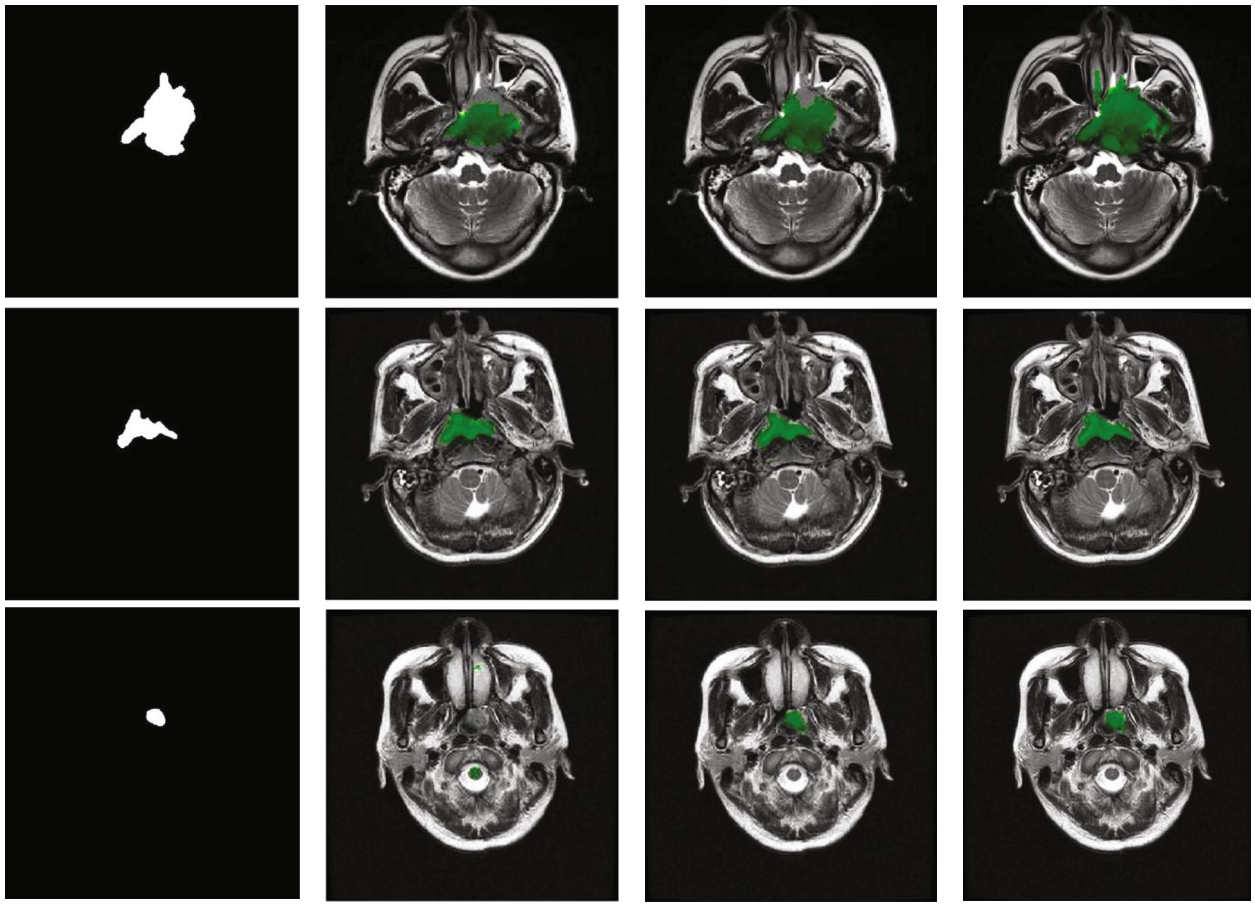

Figure 7: NPC segmentation results by fusing T1-w and T2-w modalities on the whole slices. (a) Ground truth. From second to last column: the tumor regions located by (b) Zhao et al. [12], (c) Li et al. [13], and (d) our approach, respectively.

Since the individual view cannot reveal the intrinsic structure of the data, one may note that the classification on each individual view may not obtain accurate results. When the synesthetic data was noise free, the classification by MvDA-VC obtained the highest accuracy by fusing X-Y, Y-Z, and X-Z views. However, when the noise level increased, its performances were inferior to the MISL and CODL. Throughout the experiments, the proposed CODL achieved the best performance uniformly. With the increasing noise level, the reduction of our method's classification performance was significantly lower than that of other methods. Even when the data was heavily contaminated by the noises (std = 1), the CODL remained superior performance with the highest accuracy of 74.9%.

4.3. Realistic Experiments on Nasopharyngeal Carcinoma Data

4.3.1. Image Preprocessing. Most of the NPCs have a small volume and thus are very difficult to discriminate from its large

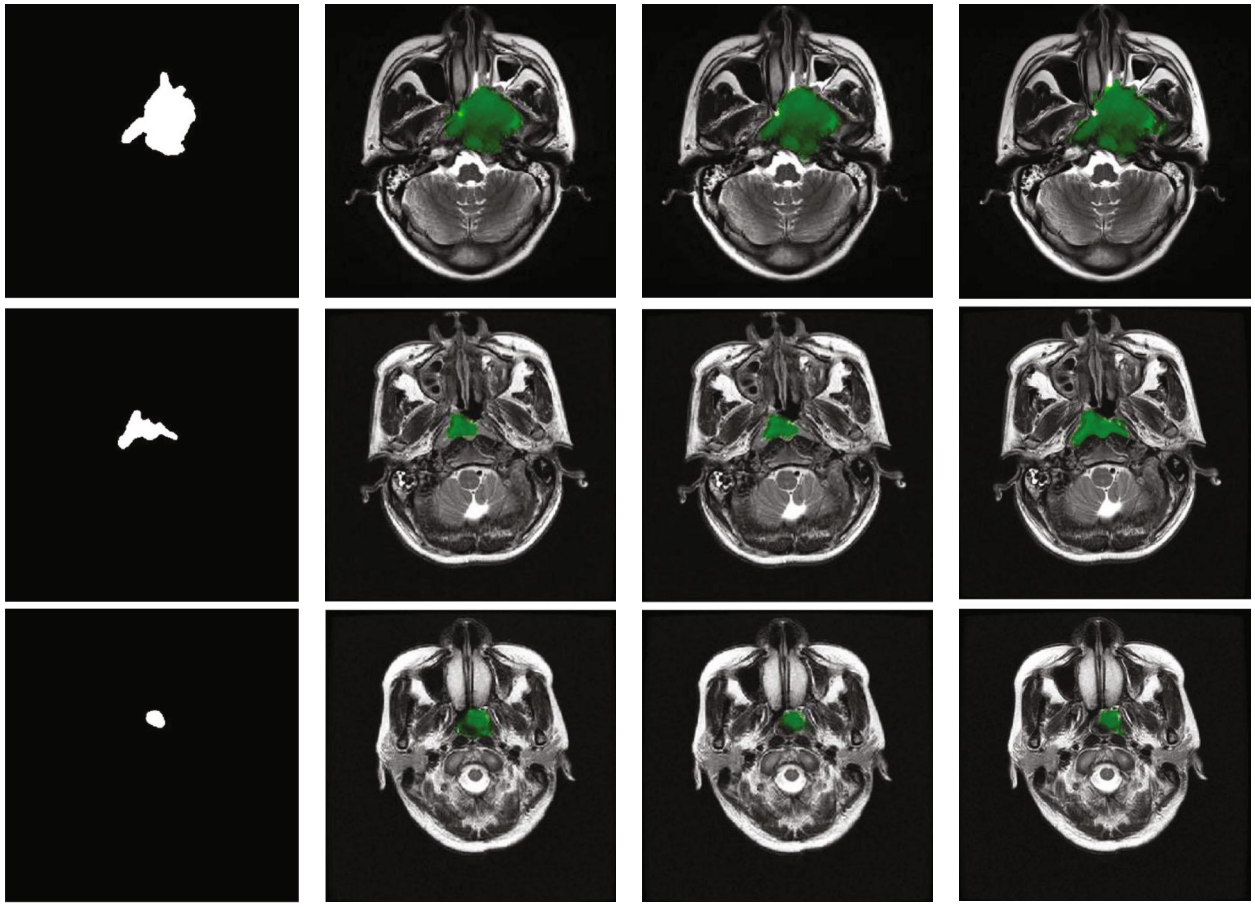

Figure 8: NPC segmentation results by fusing T1-w, T2-w, and CET1-w modalities on the whole slices. (a) Ground truth. From second to last column: the tumor regions located by (b) Zhao et al. [12], (c) Li et al. [13], and (d) our approach, respectively.

surrounding. Such imbalance also results in a large false positive rate in applying learning models directly. To solve these difficulties, we firstly designed a fully convolutional network (FCN) to locate a rectangular box bounding the suspicious tumor. The network contains standard layers, including convolution, maximum pooling, and upsampling [32]. Our network used a jump structure to exploit deep and shallow semantic information. It also used multiscale convolution kernels to obtain a comprehensive global structure. The network was trained to predict a rectangular bounding box for the NPC.

The detailed architecture of the FCN network for NPC location is summarized in Table 2. Figure 4(a) showed the MR slices with bounding boxes identified by FCN, highlighted in red dots. We selected an outer area by extending the located bounding box by fifteen pixels outward to ensure that it sufficiently covers the tumor region.

4.3.2. Radiomics Feature Extraction and Classification. In the bounding box, each voxel is classified into a binary label of tumor vs. normal. The features for each pixel were estimated within a sliding window of 11×11 centered itself. A total of 192 radiomics features (i.e., 32 Gabor, 5 Momentum, 154 GLCM, and 1 Pixel) were extracted for each sliding window. See section S1 in the Supplementary Material for more information on radiomics feature. If the border size is 103×78, it resulted in a sample matrix with 8034 samples and 192 features. The methods for extracting features from T1-w, T2-w, and CET1-w sequences are the same. We use z-score for standardization. Finally, we use an adaptive median filter function to perform a simple postprocessing on the entire slice to retain the largest connected area.

We tested the performance of CODL using a 10-fold cross-validation scheme. The percentage of training sets and test sets per fold cross-validation is 9 : 1. A total of 30 instances (training cohort: 27, testing cohort: 3) were enrolled in the voxel classification analysis. Classification accuracy was measured in terms of average accuracy across ten trials on different training and testing sets.

4.3.3. Experimental Results. Figure 5 visualizes NPC segmentation results on three typical instances, having large, medium, and small size tumors, respectively. Each row stands for segmentation results for one instance of MR sequences. From Figure 5, one would find that the segmentation results of CODL with fusing T1-w, T2-w, and CET1-w MR sequences obtained a highly accurate segmentation.

Figure 4 shows the overall segmentation process. As is illustrated in Figure 4(a), we select the outer area by

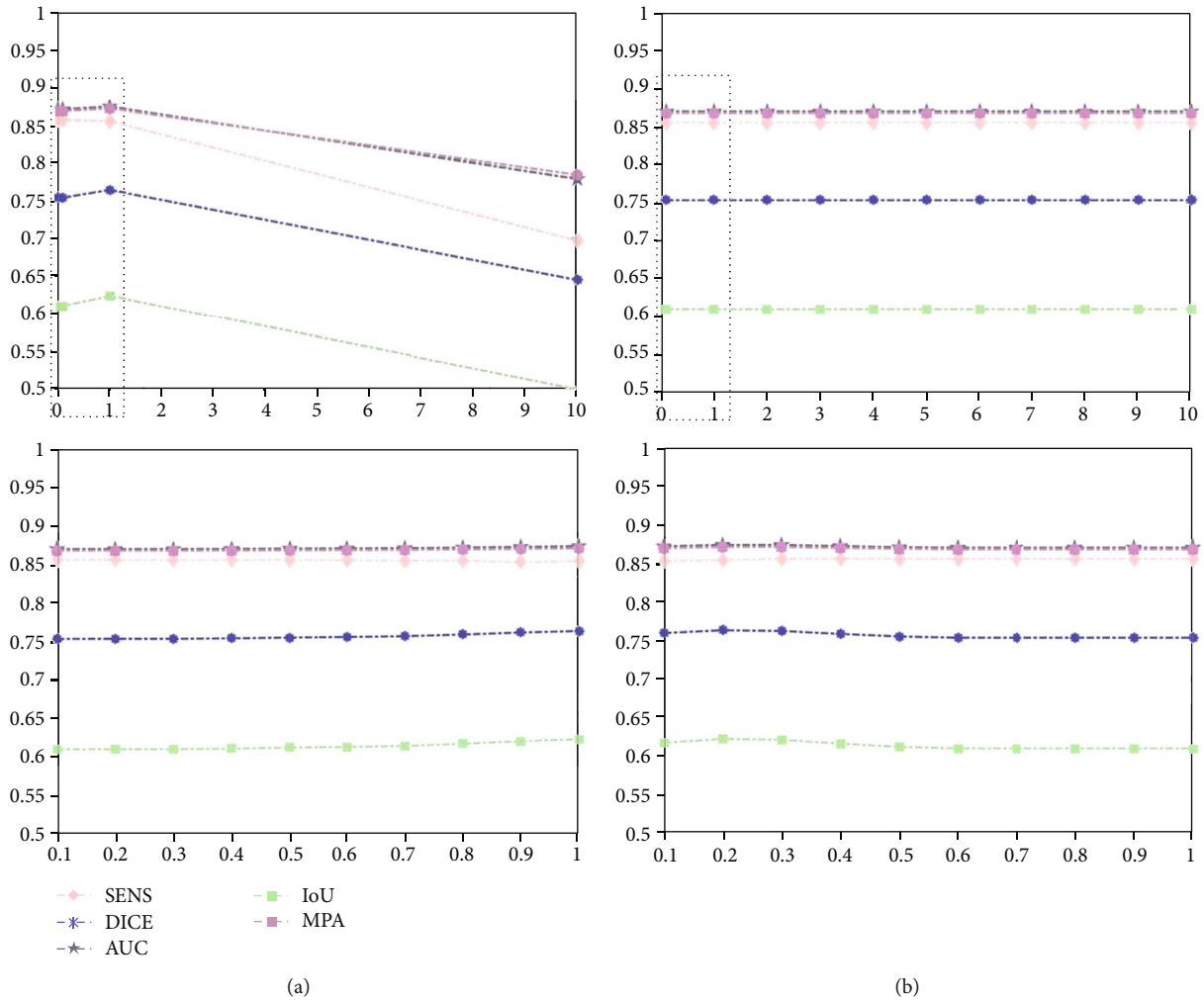

Figure 9: Performance of our model on NPC dataset with different parameter settings by fusing T1-w and T2-w modalities: (a) hyperparameter λ_1 and (b) hyperparameter λ_2.

expanding the positioned bounding box 15 pixels outward. The extended areas used for fine classification were indicated by solid red lines. Figure 4(b) shows pixel-wise fine classification results using MISL, MvDA-VC, and CODL. One may observe that CODL obtained the highest accuracy. Our method performed stably in identifying tumors of different volumes. Specifically, Figure 4(c) showed the identified tumors in the whole slices. One may observe that the proposed method identifies the tumor successfully with its boundary almost perfectly overlapped with the actual one.

We report the detailed numerical results on cropped NPC dataset in Table 3.

In the first section in Table 3, we firstly tested the classification performance on each individual image modalities. CODL performed uniformly better than SVM. The superior performance of CODL is consistent with the synthetic data. Moreover, the CET1-w provides a more accurate classification than T2-w or T1-w. The AUCs by CODL were 0.868 ± 0.050, 0.860 ± 0.090, and 0.847 ± 0.055 on CET1-w, T2-w, and T1-w, respectively.

Considering that some NPC patients do not get CET1-w scans due to kidney diseases, we used two modalities T1-w and T2-w to rerun the experiments. The results were summarized in the second section in Table 3. Overall, the accuracy has increased, which is higher than using any single MR modality. CODL with the fusion of T1-w and T2-w modalities scored the highest accuracy.

Finally, we used three MR modalities. One may observe that CODL achieved superior performances in classifying the nasopharyngeal carcinoma. The DICE, AUC, IoU, and MPA for CODL were uniformly larger than those by the other methods. Incorporating the imaging of CET1-w achieved minor improvement (0.889 ± 0.054) than without it (0.886 ± 0.055) by CODL. It implies that the CODL could exploit fully discriminative information in the modality of T1-w and T2-w, such that the loss of accuracy after dropping CET1-w is only mild.

Quantitative results of each method were shown by box plots in Figure 6. In terms of DICE, AUC, IoU, MPA, and HD, the performance of CODL is superior to the other methods. Another noticeable characteristic of the CODL lies in its robustness. One would find that the variances by the different metrics are dramatically smaller than by other methods. Such high robustness coincides with the experiments on synthetic data.

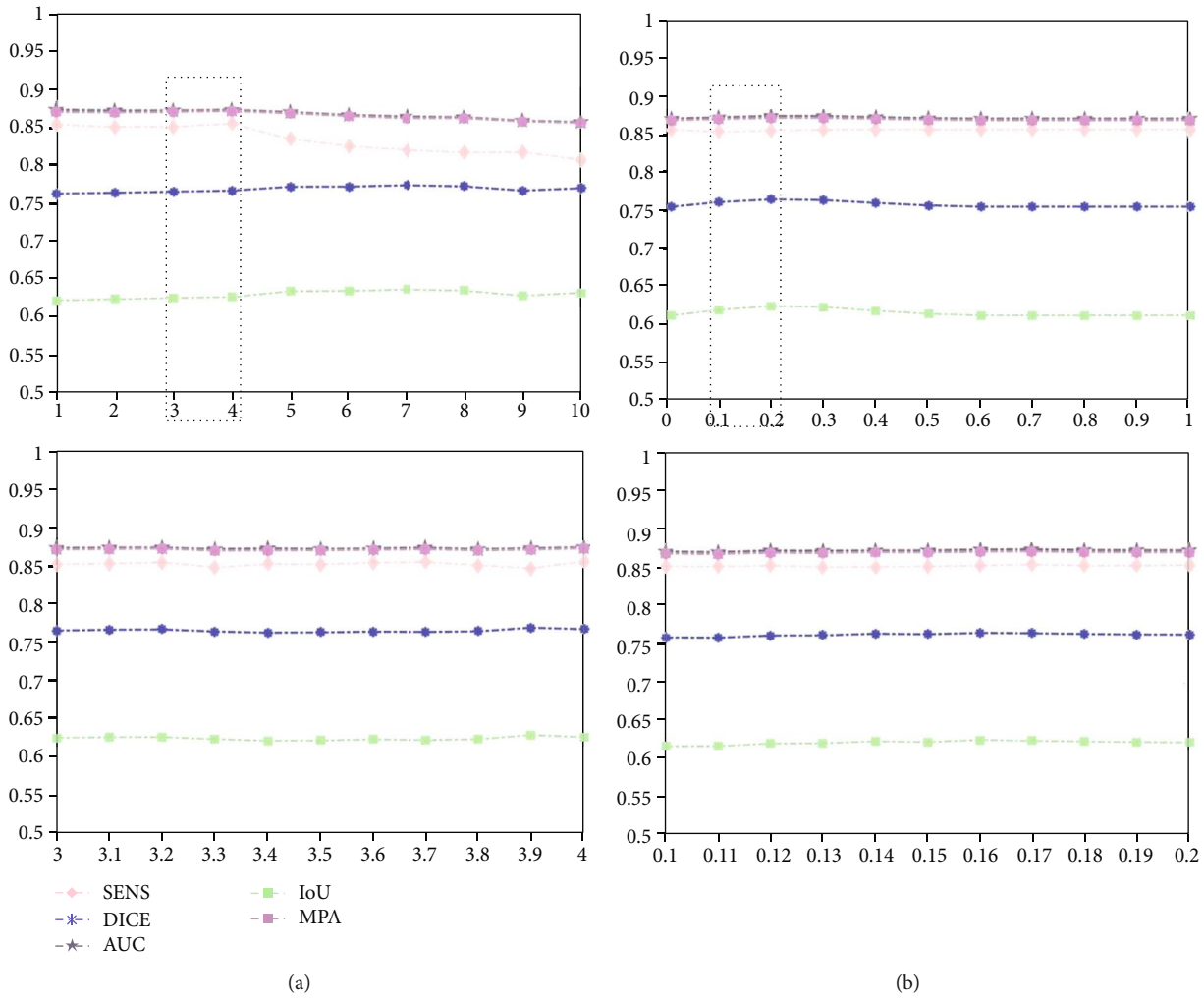

Figure 10: Performance of our model on the NPC dataset with different parameter settings by fusing T1-w, T2-w, and CET1-w modalities: (a) hyperparameter λ_1 and (b) hyperparameter λ_2.

We report the detailed numerical results on whole MR slices in Table 4.

In the first section in Table 4, we firstly tested the segmentation performance on two modalities T1-w and T2-w. One may observe that our approach achieved superior performances in NPC segmentation.

Finally, we used three modalities (i.e., T1-w, T2-w, and CET1-w) to rerun the experiments. The results were summarized in the second section in Table 4. The SENS, DICE, AUC, IoU, and MPA for our approach were uniformly larger than other methods. There were good overlaps in DICE and HD values for our method between segmented contours and ROIs drawn by radiologists. By checking the results, one can find that the variances by the six metrics are dramatically smaller than by other methods.

Figure 7 shows NPC segmentation results in case of fusing two modalities (i.e., T1-w and T2-w). From Figure 7, one may observe that the proposed method identifies the tumor successfully with its boundary almost perfect overlapped with the ground truth drawn by radiologist. Our approach achieved superior performances in segmenting NPC compared with other methods.

Figure 8 visualizes NPC segmentation results in case of fusing three modalities (i.e., T1-w, T2-w, and CET1-w). From Figure 8, one would find that the segmentation results of our approach obtained a highly segmenting performance. It can be seen that our approach could help make their wanting segmentation better.

5. Discussion

In our model, there are two regularization parameters (i.e., λ_1 and λ_2) balancing the effect of approximation error and sparse term. In the following, we study the influence of parameters λ_1, λ_2 on the NPC dataset in terms of SENS, DICE, AUC, IoU, and MPA by setting them to different values, e.g., $[1, 2, \cdots, 10]$. We vary a parameter at a time while keeping others fixed. Due to the limitation of space, we only show the results of a combination of two (i.e., T1-w and T2-w) and three modalities (i.e., T1-w, T2-w, and CET1-w).

From Figure 9, we can see that our method is relatively insensitive to its parameters as long as the parameters are in a suitable range. Moreover, we find that our method

performs well when parameter $\lambda_1 \in (0.1,1.0)$, $\lambda_2 \in (0.1,1.0)$. Thus, we select $\lambda_1 = 1.0$, $\lambda_2 = 0.2$ in our experiment. Similarly, from Figure 10, we find that our method performs well when parameter $\lambda_1 \in (3.0,4.0)$, $\lambda_2 \in (0.1,0.2)$. Consequently, we choose $\lambda_1 = 3.8$, $\lambda_2 = 0.2$ for experiments.

6. Conclusions

In this study, we have proposed a voxel-wise classification method for locating and segmenting NPC from normal tissues. Specifically, each voxel is classified into a binary label of tumor vs. normal. The located NPC is refined to obtain its accurate segmentation by an original multiview collaborative dictionary classification model. The proposed CODL integrates complementary information from multiple views and collaboratively constructs a discriminative latent intact space through rendering with supervised membership. Experimental results show that CODL could accurately discriminate NPCs and effectively assist clinicians in locating NPC.

Data Availability

The NPC image data used to support the findings of this study have not been made available for the private protection of patient information.

Conflicts of Interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Acknowledgments

This work was partially supported in part by the Key-Area Research and Development of Guangdong Province under Grant 2020B010166002, National Natural Science Foundation of China (61472145, 61771007), Science and Technology Planning Project of Guangdong Province (2017B020226004), Applied Science and Technology Research and Development Project of Guangdong Province (2016B010127003), Guangdong Natural Science Foundation (2017A030312008), Fundamental Research Fund for the Central Universities (2017ZD051), and Health & Medical Collaborative Innovation Project of Guangzhou City (201803010021, 202002020049).

Supplementary Materials

The details of the 192 radiomics features that used in this study. *(Supplementary Materials)*

References

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

[2] E. Zhuo, W. Zhang, H. Li et al., "Radiomics on multimodalities MR sequences can subtype patients with non-metastatic nasopharyngeal carcinoma (NPC) into distinct survival subgroups," *European Radiology*, vol. 29, no. 10, pp. 5590–5599, 2019.

[3] H. Yuan, Q. Y. Ai, D. L. W. Kwong et al., "Cervical nodal volume for prognostication and risk stratification of patients with nasopharyngeal carcinoma, and implications on the TNM-staging system," *Scientific Reports*, vol. 7, no. 1, article 10387, 2017.

[4] A. T. C. Chan, V. Grégoire, J. L. Lefebvre et al., "Nasopharyngeal cancer: EHNS-ESMO-ESTRO clinical practice guidelines for diagnosis, treatment and follow-up," *Annals of Oncology*, vol. 23, pp. vii83–vii85, 2012.

[5] H. Huang, J. Lu, J. Wu et al., "Tumor tissue detection using blood-oxygen-level-dependent functional MRI based on independent component analysis," *Scientific Reports*, vol. 8, no. 1, article 1223, 2018.

[6] F. K. Lee, D. K. Yeung, A. D. King, S. F. Leung, and A. Ahuja, "Segmentation of nasopharyngeal carcinoma (NPC) lesions in MR images," *International Journal of Radiation Oncology•Biology•Physics*, vol. 61, pp. 608–620, 2005.

[7] M. A. Mohammed, M. K. Abd Ghani, R. I. Hamed, D. A. Ibrahim, and M. K. Abdullah, "Artificial neural networks for automatic segmentation and identification of nasopharyngeal carcinoma," *Journal of Computational Science*, vol. 21, pp. 263–274, 2017.

[8] M. A. Mohammed, M. K. Abd Ghani, N. Arunkumar, S. A. Mostafa, M. K. Abdullah, and M. A. Burhanuddin, "Trainable model for segmenting and identifying nasopharyngeal carcinoma," *Computers & Electrical Engineering*, vol. 71, pp. 372–387, 2018.

[9] Z. Ma, X. Wu, Q. Song, Y. Luo, Y. Wang, and J. Zhou, "Automated nasopharyngeal carcinoma segmentation in magnetic resonance images by combination of convolutional neural networks and graph cut," *Experimental and Therapeutic Medicine*, vol. 16, no. 3, pp. 2511–2521, 2018.

[10] K. Men, X. Chen, Y. Zhang et al., "Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images," *Frontiers in Oncology*, vol. 7, p. 315, 2017.

[11] Q. Li, Y. Xu, Z. Chen et al., "Tumor segmentation in contrast-enhanced magnetic resonance imaging for nasopharyngeal carcinoma: deep learning with convolutional neural network," *BioMed Research International*, vol. 2018, Article ID 9128527, 7 pages, 2018.

[12] L. Zhao, Z. Lu, J. Jiang, Y. Zhou, Y. Wu, and Q. Feng, "Automatic nasopharyngeal carcinoma segmentation using fully convolutional networks with auxiliary paths on dual-modality PET-CT images," *Journal of Digital Imaging*, vol. 32, no. 3, pp. 462–470, 2019.

[13] S. Li, J. Xiao, L. He, X. Peng, and X. Yuan, "The tumor target segmentation of nasopharyngeal cancer in CT images based on deep learning methods," *Technology in Cancer Research & Treatment*, vol. 18, 2019.

[14] Y. Ye, Z. Cai, B. Huang et al., "Fully-automated segmentation of nasopharyngeal carcinoma on dual-sequence MRI using convolutional neural networks," *Frontiers in Oncology*, vol. 10, p. 166, 2020.

[15] C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 12, pp. 2531–2544, 2015.

[16] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 188–194, 2016.

[17] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," 2013, https://arxiv.org/abs/1304.5634.

[18] P. H. Kuo, E. Kanal, A. K. Abu-Alfa, and S. E. Cowper, "Gadolinium-based MR contrast agents and nephrogenic systemic fibrosis," *Radiology*, vol. 242, no. 3, pp. 647–649, 2007.

[19] L.-L. Zhang, M. Y. Huang, Y. Li et al., "Pretreatment MRI radiomics analysis allows for reliable prediction of local recurrence in non-metastatic T4 nasopharyngeal carcinoma," *eBioMedicine*, vol. 42, pp. 270–280, 2019.

[20] Y. Quan, Y. Xu, Y. Sun, and Y. Huang, "Supervised dictionary learning with multiple classifier integration," *Pattern Recognition*, vol. 55, pp. 247–260, 2016.

[21] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 210–227, 2008.

[22] K. Huang and S. Aviyente, "Sparse representation for signal classification," in *Advances in Neural Information Processing Systems*, pp. 609–616, The Neural Information Processing Systems Foundation, 2007.

[23] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1794–1801, Miami, FL, USA, 2009.

[24] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3360–3367, San Francisco, CA, USA, 2010.

[25] I. Diamant, E. Klang, M. Amitai, E. Konen, J. Goldberger, and H. Greenspan, "Task-driven dictionary learning based on mutual information for medical image classification," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 6, pp. 1380–1392, 2017.

[26] M. J. Gangeh, A. K. Farahat, A. Ghodsi, and M. S. Kamel, "Supervised dictionary learning and sparse representation-a review," 2015, https://arxiv.org/abs/1502.05928.

[27] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[28] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang, "Multi-observation visual recognition via joint dynamic sparse representation," in *2011 International Conference on Computer Vision*, pp. 595–602, Barcelona, Spain, 2011.

[29] F. Nie, G. Cai, J. Li, and X. Li, "Auto-weighted multi-view learning for image clustering and semi-supervised classification," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1501–1511, 2018.

[30] L. Zhang, Y. Hu, C. Li, and J.-C. Yao, "A new linear convergence result for the iterative soft thresholding algorithm," *Optimization*, vol. 66, no. 7, pp. 1177–1189, 2017.

[31] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[32] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, Haines Convention Center in Boston, Massachusetts, 2015.