IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 20, 2018, accepted June 17, 2018, date of publication July 9, 2018, date of current version July 30, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2851027

A Novel Deep Learning Framework for Internal Gross Target Volume Definition From 4D Computed Tomography of Lung Cancer Patients

XIADONG LI[1,2], ZIHENG DENG[3], QINGHUA DENG[4], LIDAN ZHANG[5], TIANYE NIU[1,2], AND YU KUANG[6]

[1]Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310000, China
[2]Institute of Translational Medicine, Zhejiang University, Hangzhou 310000, China
[3]School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[4]Department of Radiation Oncology, Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Hangzhou 310000, China
[5]Department of Radiation Oncology, Hangzhou Cancer Hospital, Hangzhou 310000, China
[6]Department of Medical Physics, University of Nevada, Las Vegas, NV 89154, USA

Corresponding authors: Tianye Niu (tyniu@zju.edu.cn) and Yu Kuang (yxk56@case.edu)

This work was supported by the Zhejiang Provincial Natural Science Foundation of China (Grant No. LR16F010001), National High-tech R&D Program for Young Scientists by the Ministry of Science and Technology of China (863 Program, 2015AA020917), Natural Science Foundation of China (NSFC Grant No. 81201091, 51305257), National Key Research Plan by the Ministry of Science and Technology of China (Grant No. 2016YFC0104507).

ABSTRACT In this paper, we study the reliability of a novel deep learning framework for internal gross target volume (IGTV) delineation from 4-D computed tomography (4DCT), which is applied to patients with lung cancer treated by stereotactic body radiation therapy (SBRT). Seventy seven patients who underwent SBRT followed by 4DCT scans were incorporated in this retrospective study. The IGTV_DL was delineated using a novel deep machine learning algorithm with a linear exhaustive optimal combination framework. For the purpose of comparison, three other IGTVs based on common methods was also delineated. We compared the relative volume difference (RVI), matching index (MI), and encompassment index (EI) for the above IGTVs. Then, multiple parameter regression analysis was performed to assess the tumor volume and motion range as clinical influencing factors in the MI variation. The results demonstrated that the deep learning algorithm with linear exhaustive optimal combination framework has a higher probability of achieving optimal MI compared with other currently widely used methods. For patients after simple breathing training by keeping the respiratory frequency in 10 breath per minute (BPM), the four phase combinations of 0%, 30%, 50% and 90% can be considered as a potential solution for an optimal combination to synthesize IGTV in all respiration amplitudes.

INDEX TERMS Deep learning, computed tomography, algorithm, stereotactic ablative radiotherapy, internal gross target volume, lung cancer.

I. INTRODUCTION

Surgical resection has been accepted as a standard of care for early stage of non-small cell lung cancer (NSCLC) with a 5-year survival rate of 50-80% [1], [2]. For patients having NSCLC who are unable or unwilling to receive surgery, SBRT has become one of the best alternative treatment options. Recently, a randomized phase III trial of SBRT for patients with operable stage I NSCLC showed a better overall and recurrence-free survival rate at 3 years with an acceptably reasonable relevant treatment-related adverse events as compared to the surgery group [3]. However, precise delineation of the tumor volume in an efficient way still poses a major challenge in these SBRT cases, especially for institutions that were not equipped with deformable image registration systems. One of the crucial factors that cause problems in the accurate delineation of the target volume is tumor movement-induced geometric uncertainties due to irregular respiration patterns [4].

To compensate the geometric and motion deviation from target, a larger target volume is usually contoured, thus

hindering the delivery of the optimal SBRT dose due to the unnecessary partial volumes of surrounding organs-at-risk involved inside the treatment fields [3]. In the past ten years, four-dimensional computed tomography (4DCT) scanning technique has been used to acquire the information of motion induced organ distortion related to respiration. Various methods to delineate the tumor target on the 4DCT datasets in planning process have been investigated [7]–[11].

The relatively accurate and well-accepted method for determining internal gross target volume (IGTV) is to integrate contours from all 10 phases of the 4DCT dataset. However, this method is time consuming and requires great effort due to the workload incurred on the radiation oncologists. In contrast, the commercial packages offer much faster approaches that are routinely utilized, e.g. contouring on MIP (maximum intensity projection) images and validating the contours with fly-through movie, or contouring on one phase and applying this contour to other phases via deformable registration (based on the MIM_SOFTWARE^{TM}). However, the so-called rapid commercial software processing method cannot be applied to all kinds of respiratory motion patterns, such as in some irregular motion pattern cases even at the expense of the delineation accuracy.

At present, with the development of computer aided image processing technology, the main methods of image segmentation are as follows: (1) Threshold based segmentation method (Threshold) [12]; (2) Edge based segmentation method (edge-detecting) [13]; (3) Region based segmentation method [14]; (4) Image segmentation method based on clustering analysis (Clustering) [15]; (5) Segmentation method based on Wavelet Transform [16]; and (6) Segmentation method based on mathematical morphology [17]. The advantage of the aforementioned non machine learning algorithms are that they can achieve more accurate segmentation by adjusting the computing parameters. Obviously, the drawback is that the segmentation parameters need manual adjustment because of different image contents involved. As such, it is difficult to achieve fully automatic streamlined process. Therefore, we proposed that the transfer learning in a machine learning framework can be adopted to improve the accuracy and efficiency of the tumor delineation in 4DCT. The development of convolutional neural network layers has achieved significant performance in image classification and tumor segmentation for a given DICOM image [18], [19]. In these methods, multiple neural layers with image analysis filters, or convolutions, are applied. The abstracted features of images within each neural layer are extracted by systematically constructing multiple filters across the image. A feature map generated can be used as new input to the following layer. In this study, a convolutional neural network trained on the ImageNet dataset of 20,000 categories was adapted to significantly increase the accuracy of tumor delineation and shorten the training duration of same network trained on the dataset of 4DCT images using the Tensorflow^{TM}. Firstly, we adopted an inception V3 architecture to pre-train on the ImageNet dataset [20]. Secondly, we investigated whether or not the discrete linear exhaustion method can achieve an optimal phase combination for delineation of IGTV in a more efficient and accurate way. In summary, this study demonstrated the reliability and feasibility of this framework from a perspective of clinical practice.

II. MATERIAL AND METHODS

A. PATIENTS DEMOGRAPHICS

A total of 77 patients with NSCLC (42 males and 35 females; average age: 58 years, range: 41-76 years) who received SBRT were included in this retrospective study. According to the TNM staging system of Union for International Cancer Control, 37 patients were diagnosed at stage I and 40 patients were at stage II (Table 1). Patients with tumors having a diameter of the transverse cross section larger than 6 cm were excluded from this study. All clinical treatment protocols were approved by the ethics committee of the hospital and informed consents were obtained from the patients or their authorized family members. Computer aided lung boundary extraction and automatic segmentation was implemented in each transverse section for each patient. This work was done in 10 phases from 0% to 90% phases and the minimum length of the tumor was more than 2 cm in the research cohort. All patients underwent respiratory training before 4D CT Scanning, and the respiratory frequency was controlled at about 10 breaths per minute (BPM). This stems from the fact the frequency is easier to train, whereas amplitude training is more difficult for patients with poor lung function.

TABLE 1. Clinical characteristics of patients in research cohort.

	Characteristics	Number (%)
1	Median age	58 (41-76)
2	Pathological type	
	Adenocarcinoma	37 (48.1%)
	Squamous cell	23 (29.8%)
	Adenosquamous	17 (22.1%)
3	Tumor grade	
	Grade I	37 (48.1%)
	Grade II	40 (51.9%)
4	Tumor position[a]	
	Upper lobe of left lung	19 (24.6%)
	Lower lobe of left lung	15 (19.5%)
	Superior lobe of right lung	17 (22.1%)
	Middle lobe of right lung	10 (12.9%)
	Inferior lobe of right lung	12 (15.6)
	Central type	5 (6.5%)
5	Adjacent tissues	
	Lung tissue	47 (61.0%)
	Mediastina	16 (20.8%)
	Chest wall	14 (18.2%)
6	Tumor diameter(cm)	
	≤2	19 (24.7%)
	2~4	34 (44.2%)
	4~6	24 (31.2%)

Tumor position[a] : one patient with two tumors

B. COMPUTED TOMOGRAPHY (CT) SIMULATION

Prior to treatment planning, all patients underwent 4D CT imaging acquisition. Before scanning, patients were positioned head-first and supine in a customized hot plastic immobilization device. They were subsequently trained to

breathe normally within a relatively consistent respiration frequency under the assistance of our in-house developed Respiration Training and Monitoring System (RTMS). The respiration of the patients was registered with a special bellyband tied around the upper abdomen near the cartilage ossiform, which measured the stretching force changes during inhalation and exhalation. The CT scanner (Brilliance CT Big Bore; Philips Medical Systems, Cleveland, OH, USA) acquired the images in the helical scanning mode with the parameters of 130 kV, 240 mA, 3-mm slice thickness, and 0.5s gantry rotation. The scanning pitch was computed based on a well-established formula [21].

C. TRANSFER LEARNING METHODS

Using Tensorflow^{TM}, we implemented an inception V3 architecture pre-trained on the ImageNet dataset. Retraining process consisted of initializing the convolutional layers with loaded pre-trained weights and retraining the final layers. The newly initialized network then takes the DICOM images of the patients with NSCLC as input and retrains layers in order to delineate the IGTV. Attempts to ''fine-tuning'' the convolutional layers were conducted by unfreezing and updating the pre-trained weights on our medical images using backpropagation tended to decrease model performance due to overfitting. The Inception model was implemented on an Ubuntu 16.04 computer with 28 Intel Xeon Core i7 CPUs, utilizing a 2 NVIDIA Geforce GTX 1080980 Ti 8Gb GPU for training and testing, with 256Gb available in RAM memory. Training of layers was performed by stochastic gradient descent in batches of 20,000 images per step using an Adam Optimizer with a learning rate of 0.001. By interpreting the output of the convolutional network as a model for the distribution over segmentation labels, the pixel of the tumor region is defined as 1, and the other regions are defined as 0, a natural training criterion is to maximize the probability of all labels in our training set or, equivalently, to minimize the negative log-probability $-\log p(Y \mid X) = \sum_{ij} -\log p(Yij \mid X)$ for each labeled lung tumor, where ''ij'' was the coordinate of the pixel, X stands for the golden margin randomly pre-delineated by one of the senior radiologists ($n = 17$). Y is the contour outlined by the trained machine, if the machine learned contour is consistent with the golden margin, the according pixel value, the output of the function is 1, otherwise, if the result is inconsistent, the output is 0. Training on all categories of ROI segmentation was run for 120,000 steps, or 100 epochs.

Iterations were run through the entire training dataset. The process was demonstrated in Fig.1. The training of the final layers were converged for all classes. Hold-out method testing was performed after each step using a partitioning test containing images from subjects independent of the patients, which represented in the training partition by passing each image through the network without performing gradient descent and backpropagation. Finally, the best performing model was retained for analysis.

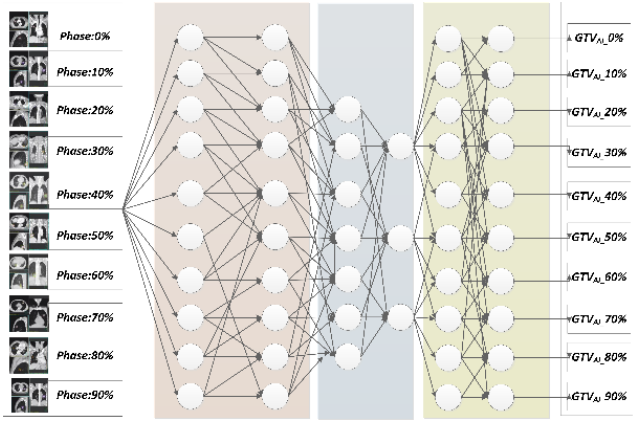


FIGURE 1. Demonstration of the trained neural network for delineation of GTV in 4DCT.

D. TARGET VOLUME DELINEATION

A training set comprising 45 patients is based on the GTVs that have been pre-delineated manually in each phase as the tumor margin by senior radiologists (n=17) with more than 10 years' experience in radiotherapy. Meanwhile, tumor boundary extraction and delineation of GTV_{AI} was performed in the same transverse slice and small pulmonary nodules and bronchial images were excluded by the computer automatically. All the GTV delineation was done in the same window width and level according to the tumor location: Lung window width: 1500-2000HU, window level: 450-600HU (tumor in the lung lobe); mediastinum window width:250-350HU window level: 30-50HU (tumor near the mediastinum or diaphragm and liver). In order to evaluate our trained model in the context of clinical experts, a validation group of 20 NSCLC patients with more than 15,000 images independent of the patients in the training set was also used to compare our machine learning delineation with the contouring delineated by human experts. Note that the experts ''A'' and ''B'' were randomly selected from 17 radiologists, and was assigned to re-delineate the GTV for patients in the validation cohort. MI was also used to reflect the accuracy of trained model by comparing GTVs contoured by the machine with human being (experts ''A'' and ''B''). Only when the MI >= 0.95 was satisfied, the model was considered to meet the requirements of clinical utilities. Finally, four IGTVs were generated for analysis as follows:

(1) IGTV10_DL formed from GTV_{AI} by machine learning with linear exhaustive optimal combination framework. To evaluate the accuracy of IGTV10_DL, three additional IGTVs were also delineated for comparison purpose.

(2) IGTV10_E1 (experts A) was contoured on all 10 respiratory phases of the 4DCT images by a senior radiologist with 10 years of experience in radiotherapy and target delineation.

(3) IGTV10_E2 (experts B) was also delineated in all 10 respiratory phases by another senior radiologist with 20 years of experience in lung tumor diagnosis and target countering.

(4) IGTV10_DEF (deformable) was automatically formed via elastic deformation registration technique using MIM SOFTWARE$^{TM.}$

(5) IGTV10_HB was defined in the following manner: the average volume and tumor boundary of IGTV10_E1 and IGTV10_E2 that are delineated as IGTV10_HB (Human Being) were used as reference volumes.

E. DISCRETE LINEAR EXHAUSTION METHOD (DLEM)

DLEM is a key algorithm in linear exhaustive optimal combination framework, theoretically, the IGTV can be combined with 1 to 10 phases. According to the principle of permutation and combination, a total of 1023 combinations can be formed, so there are 1023 IGTV possibilities for each patient. In order to facilitate the statistical analysis, a 32 × 32 matrix is constructed, and 10 regions were carved up according to the amount of phase combination for each patient. Each element in the matrix is corresponding to a certain IGTV with a specific combination of 10 phases to ensure that each patient's IGTV have the same coding order. The coordinate of each element in a matrix and the region in which it is located are fixed for each patient (Table.2).

TABLE 2. Demonstration of DLEM for IGTV delineation.

Region		Combination	Amount
P1	C_{10}^{01}	*{p0},{ p1},{ p2},{ p3} ,……*	10
P2	C_{10}^{02}	*{p0, p1}, {p0, p2}, {p0, p3},,……*	45
P3	C_{10}^{03}	*{p0, p1, p2},{p0,p1, p3},{p0,p1, p4},……*	120
P4	C_{10}^{04}	*{p0, p1, p2, p3}, {p0, p1, p2, p4}, ,……*	210
P5	C_{10}^{05}	*{p0, p1, p2, p3 , p4},{p0, p1, p2, p3 , p5},……*	252
P6	C_{10}^{06}	*{p0, p1, p2, p3, p4, p5}, {p0, p1, p2, p3, p4, p6},……*	210
P7	C_{10}^{07}	*{p0, p1, p2, p3, p4, p5, p6}, {p0, p1, p2, p3, p4, p5, p7},……*	120
P8	C_{10}^{08}	*{p0, p1, p2, p3, p4, p5, p6, p7},,……*	45
P9	C_{10}^{09}	*{p0, p1, p2, p3, p4, p5, p6, p7, p8} ,……*	10
P10	C_{10}^{10}	*{p0, p1, p2, p3, p4, p5, p6, p7, p8, p9}*	1
SUM			1023

F. TUMOR MOTION RANGE AND THREE DIMENSIONAL TUMOR MOTION RANGE (R_{3D})

The tumor motion range in one direction was designated as the differences between the maximum and minimum coordinate values of the target center in 10 respiration phases. Three directions of the tumor motion ranges were used including left-right (LR; R_{LR}), anterior-posterior (AP, R_{AP}), and cranio-caudal (CC; R_{CC}) orientations; $R_{3D} = (R_{LR}^2 + R_{AP}^2 + R_{CC}^2)^{1/2}$ was computed to evaluate the three-dimensional motion results of the target affected by respiration movements.

G. RELATIVE VOLUME INDEX, ENCOMPASSMENT INDEX AND MATCHING INDEX

Relative volume index (RVI) was defined as the ratio of IGTVs volume to IGTV10_HB volume. The value of RVI equals to 1 if the two volumes were fully matched. Encompassment index (EI) EI = $IGTV_X \cap$ IGTV10_HB/$IGTV_X$ was computed according to Seppenwoolde *et al.* [22]. The concept of EI(x in y) is defined as the percentage of volume x which was encompassed by y. In our study, IGTV10_HB was defined as a reference volume. In a free-breathing respiration status, a lower value of the EI of IGTVs indicates that there is a higher opportunity that the target would be missed during treatment. Accordingly, 1-[EI(x in y)] defines the portions of the target missed during radiotherapy. The demonstration of the EI was illustrated in Figure 2. Matching index (MI) was defined as MI = (IGTVs ∩ IGTV10_HB IGTVs ∪ IGTV10_HB). Computation of the average MI was performed between each IGTVs and IGTV10_HB in order to evaluate the relative volumetric comparability of IGTV10_HB for target delineation and clinical feasibility of individualizing IGTVs.

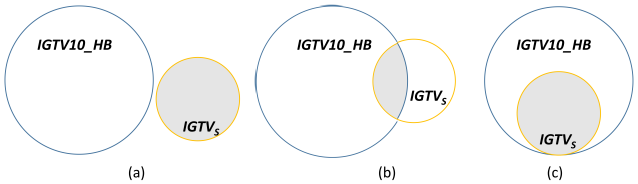


FIGURE 2. Demonstration of the difference between EI and MI, whereby the shadow area represents the intersection of the two circles. For instance, the diameter of the bigger and smeller circle is 3cm and 2 cm, respectively, accordingly, the MI for (c) is 0.44, whereas the EI equals to 1.

H. DATA ANALYSIS AND STATISTICAL ANALYSES

All patients' respiration induced tumor motions were summarized and evaluated, and the volumes of IGTV10_DL, IGTV10_DEF relative to IGTV10_HB were assessed by comparing the MI, EI and RVI. The statistical analyses were performed using SPSS v21 package (IBM, Armonk, NY, USA) in order to estimate any statistically significant differences between the IGTVs determined using each volume (IGTV10_DL, IGTV10_DEF, IGTV10_E1, IGTV10_E2) as compared to the IGTV10_HB. The paired-sample t-test was used in each case, with $P < 0.05$ being considered significant. In addition, multiple parameter regression analysis was used to evaluate whether the tumor volume, location, and motion range are clinical influencing factors for MI differences.

III. RESULTS

A. TUMOR MOTION RANGE AND THREE DIMENSIONAL TUMOR MOTION RANGE (R_{3D})

The mean R_{3D} was 12.3mm (range, 2.5-55.3mm). To study the influence of magnitude of tumor motion based on the accuracy of IGTV delineation, the patients was divided into three groups: those with tumor motion range>= 10mm, those with tumor motion in 5 ∼ 10mm, and those with tumor motion range < 5mm. Tumor motion ranges smaller than 5mm were observed in only 27.3% of the study group in CC direction, whereas ranges greater than or equal to 10mm were found in 33 patients with a maximal range of 53.7mm. However, 45.5% and 50.6% of the patients were observed to have less motion range (R < 5mm) in AP and LR directions, respectively. All measurement results of tumor motion are summarized in Table 3.

TABLE 3. Measurement of tumor motion (mm) in normal respiration pattern.

			Range<5 mm	Range 5~10 mm	Range>10 mm
	Range (mm)	Mean (mm)	No of patients	No of patients	No of patients
R_{CC}	9-53.7	29.5	21	23	33
R_{LR}	1.4-8.4	3.5	39	38	0
R_{AP}	1.1-8.6	4.5	35	42	0
R_{3D}	2.5-55.3	1.23	20	24	33

Abbreviations:
R_{AP}: tumor motion in the anterior-posterior direction;
R_{LR}: tumor motion in the left-right direction;
R_{CC}: tumor motion in the crania–caudal direction;
R_{3D}: defined as $R_{3D}= (R_{LR}^2+ R_{AP}^2+ R_{CC}^2)^{1/2}$.

B. RVI OF IGTVs COMPARED WITH IGTV10_HB

Table 4 shows the RVI values for each of the IGTVs. To investigate RVI of IGTVs compared with IGTV10_HB for all patients, the delineated volume of IGTVs was normalized to IGTV10_HB, and comparisons were performed using paired-sample t-test. In general, we found that the range of IGTV10_E1 (mean ± SD: 0.82 ± 0.51) were smaller than that of IGTV10_HB, whereas the range of IGTV10_E2 was slightly larger than that of IGTV10_HB. On the contrary, the ranges of IGTV10_DEF and IGTV10_DL (mean ± SD: 0.97 ± 0.02 and 0.95 ± 0.04, respectively) were very close to that of the reference volume. The paired sample t-test revealed that the ranges of IGTV10_E1, IGTV10_E2 significantly differed from that of the reference volume (P = 0.00), whereas the range of IGTV10_DEF and IGTV10_DL did not have a significant difference as compared to that of the IGTV10_HB (P = 0.63 and 0.57 respectively).

TABLE 4. Summary of the relative volume index (RVI).

	Range	Mean±SD	Comparison	P value
RVI _ IGTV10_E1	0.71 -0.97	0.82±0.51	IGTV10_HB - IGTV10_E1	0.00
RVI _ IGTV10_E2	0.85 -1.23	1.04±0.06	IGTV10_HB - IGTV10_E2	0.00
RVI _ IGTV10_DEF	0.82 -0.98	0.91±0.02	IGTV10_HB - IGTV10_ DEF	0.63
RVI _ IGTV10_DL	0.89 -0.99	0.95±0.04	IGTV10_HB - IGTV10_ DL	0.57

C. MI and EI COMPARISONS BETWEEN IGTVs AND IGTV10_HB

To compare the differences of MI and EI between IGTVs and IGTV10_HB, all statistical results were collected in Table 5 with the paired-sample t-test performed. For MI, the IGTV10_DL (mean ± SD: 0.92 ± 0.07) closely matched the IGTV10_HB, followed by IGTV10_DEF (mean ± SD: 0.83 ± 0.14), whereas the IGTV10_E1 and IGTV10_E2 have a lower MI (mean ± SD: 0.82 ± 0.07, 0.86 ± 0.11). There were significant differences (P = 0.003) among different IGTV delineation methods. For EI, there was no significant difference between any IGTVs and the reference volume. Also, no statistical significance was found among different IGTV delineation methods (P = 0.224).

The MI between the delineated target volume and reference volume is the main index to determine the accuracy

TABLE 5. Summary of the MI and EI comparisons between IGTVs and IGTV10_HB.

	MI	EI
	Mean ± SD	Mean±SD
IGTV10_E1	0.82 ± 0.07	0.97 ± 0.02
IGTV10_E2	0.86 ± 0.11	1.07 ± 0.04
IGTV10_DEF	0.83 ± 0.14	0.89 ± 0.03
IGTV10_DL	0.92 ± 0.07	0.91 ± 0.05
F	-4.205	-0.804
P	0.003	0.224

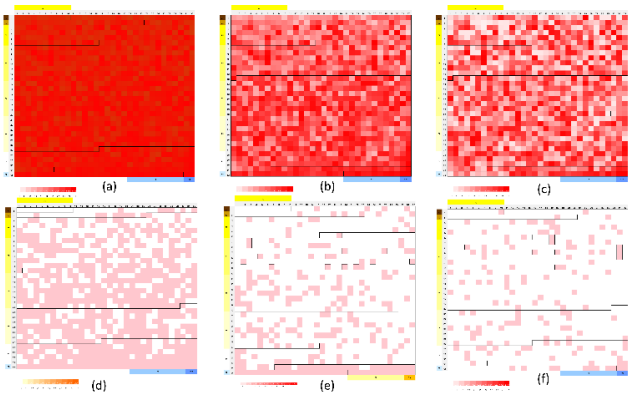


FIGURE 3. MI distribution of IGTV10_DL in three subgroups with different tumor 3D motion range. (a) MI_IGTV10_DL R_{3D}<5mm. (b) MI_IGTV10_DL R_{3D}: 5~10mm. (c) MI_IGTV10_DL R_{3D}>10mm. (d) MI_IGTV10_DL R_{3D}<5mm with optimal MI displayed. (e) MI_IGTV10_DL R_{3D}:5~10mm with optimal MI displayed. (f) MI_IGTV10_DL R_{3D}>10mm with optimal MI displayed.

of targeting during radiotherapy. The EI is the inclusion degree of the newly built volume within the reference volume, which mainly evaluates the missing probability of the newly built volume targeted by treatment beam during radiotherapy. Because there is no statistical difference in the EI between different methods, a further analysis of the correlation between MI and tumor motion range was investigated. A subgroup analysis of different respiratory amplitude of IGTV10_DL and IGTV10_DEF was conducted. An optimal combination of block was defined and displayed with color if the corresponding MI value was more than 0.95 in that block, otherwise, it was suppressed using representation of a white color (Fig.3 ~ 5).

As shown in Fig. 3 ~ 5, there was no difference in the optimal MI ratio between IGTV10_DL and the reference volume as well as between IGTV10_DEF and the reference volume if the breathing amplitude less than 5mm (Fig 5.(a)). When the breathing amplitude is more than 10mm, the probability of optimal MI can reach its maximum point of 28.6% by using 4 phases for IGTV synthetics based on the machine learning method (Fig 5. (c)). Furthermore, through position decoding of the combination sequence, it can be detected that the corresponding phase combination at that point was 0%, 50%, 30%, and 90% (MI=0.971), However, when the breathing amplitude is between 5mm ~ 10mm, the probability of optimal matching index (MI) of IGTV based on machine learning exceeds that of IGTV synthesized by elastic

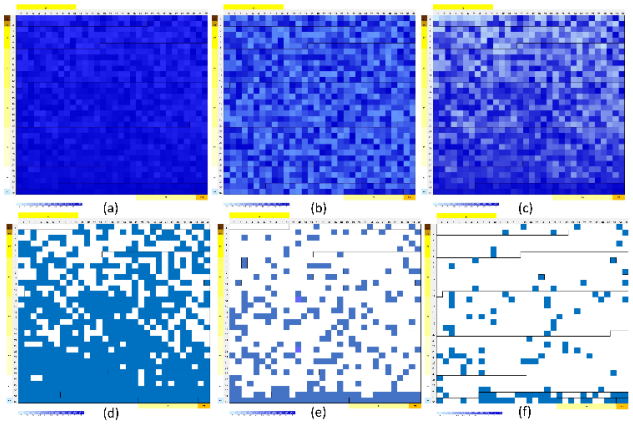


FIGURE 4. MI distribution of IGTV10_DEF in three subgroups with different tumor 3D motion range. (a) MI_IGTV10_DEF R_{3D}<5mm. (b) MI_IGTV10_ DEF R_{3D}: 5∼10mm. (c) MI_IGTV10_DEF R_{3D}>10mm. (d) MI_IGTV10_DEF R_{3D}<5mm with optimal MI displayed. (e) MI_IGTV10_DEF R_{3D}: 5∼10mm with optimal MI displayed. (f) MI_IGTV10_DEF R_{3D}>10mm with optimal MI displayed.

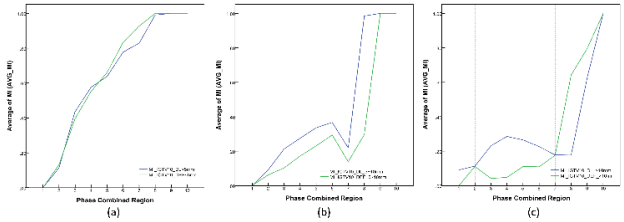


FIGURE 5. Increment of the respiratory amplitude corresponding to the changes in optimal MI ratio of IGTV10_DL and IGTV10_DEF. (a) Illustration of the optimal MI ratio of IGTV10_DL and IGTV10_DEF when R_{3D} < 5mm. (b) Demonstration of the optimal MI ratio of IGTV10_DL and IGTV10_DEF when R_{3D} in range 5 ∼ 10mm. (c) Display of optimal MI ratio of IGTV10_DL and IGTV10_DEF when R_{3D} > 10mm.

deformation registration technology, which comes with an average improvement of 36.1%, When 3 phases were chosen for IGTV combination, the probability of optimal MI was only 21.3%. In contrast, it may need more computing time and the optimal MI probability will not significantly improved if 5 phases were selected to form the IGTV. As such, 4 phases of IGTV combination can balance the optimal MI probability and contouring delineation efficiency (Fig 5. (b)). In the following decoding of the location in the MI matrix, we found that the combination sequence were: 0%, 30%, 50%, and 90% (MI = 0.967); 0%, 30%, 60% and P_90% (MI = 0.955); 0%, 40%, 60%, and 90% (MI = 0.981).

D. CORRELATIONS MI OF IGTVs WITH TUMOR MOTION RANGE AND THE TARGET VOLUME

Generally, larger tumor motion range and smaller tumor volume lead to a larger percentage of difference between IGTVs and IGTV10_HB. To study the correlation between MI of IGTVs and the magnitude of tumor motion, several subgroups were created. All discrete variables were converted into four groups of rank variables for the purpose of comparing the comparability of the magnitude as statistically presented in Table 6. Multivariate linear regression showed that a significant difference was observed in MI_IGTV_E1, MI_IGTV_E2, and MI_IGTV10_DEF (P = 0.013, 0.033 and 0.012, R2 = 0.793, 0.765 and 0.713 respectively, refer to Table 7). The MI has a positive correlation with tumor volume whereas it has a negative correlation with the magnitude of tumor motion in three directions.

TABLE 6. The rule of classification of dispersed data.

R_{LR}	Range(mm)	1.9-3.5	3.6-5.2	5.3-6.9	7.0-
	Grade	1	2	3	4
R_{AP}	Range(mm)	1.1-3.0	3.1-5.0	5.1-7.0	7.1-
	Grade	1	2	3	4
R_{CC}	Range(mm)	0-14.0	15.0-29.0	30.0-44.0	45.0-
	Grade	1	2	3	4
TV[a]	Range(mm^3)	2.0-10.0	11.0-20.0	21.0-40.0	41.0-
	Grade	1	2	3	4
MI	Range	0-0.59	0.6-0.79	0.8-0.89	0.9-1
	Grade	1	2	3	4

TV[a]: target volume.

TABLE 7. The multivariate linear regression results of MI.

	MI_ IGTV_E1	MI_ IGTV_E2	MI_ IGTV10_DL	MI_ IGTV10_DEF
Constant	2.33	3.71	2.57	2.62
B_LR*	-0.39	-0.035	-0.14	-0.12
B_AP*	-0.172	-0.486	-0.218	-0.169
B_CC*	-0.452	-0.103	-0.132	-0.02
B_TV[a]	0.177	0.081	0.357	0.135
R2	0.793	0.765	0.392	0.713
F	5.741	4.861	1.754	6.757
P	**0.013**	**0.033**	0.485	**0.012**

B_TV[a] linear regression coefficien

IV. DISCUSSION

In our study, we implemented an architecture pre-trained on the ImageNet for 4D CT tumor delineation. However, in some cases, the transfer learning method in deep learning has been suspended in the application of medical image analysis. Due to which, we have adopted a re-training process based on the 4DCT DICOM images as a remedial measure. Re-training process consisted of initializing the convolutional layers with loaded pre-trained weights and re-training the final layers. We only use the weights in the pre-training section as an initial optimization starting points in the subsequent re-training process. The newly initialized network then takes the 4D CT DICOM image as input and re-trains layers in order to delineate GTV.

Respiration during CT acquisition is expected to result in distortion of the target volume in scanning process, which may lead to inaccurate target volumetric delineation. This is still a major barrier for the precise delineation of IGTV in SBRT. Various attempts have been made to solve this problem. MIP image has been extensively used in clinical

practice to delineate IGTV when 4DCT was developed in the first instance [23], [24]. However, it has been proven that MIP-based IGTV can underestimate the true tumor volume when the adjacent structures have similar (or higher) densities, which is the case for lesions located near the mediastinum, diaphragm, liver, or chest wall [25], Our study demonstrated that the elastic deformation registration-based IGTV_DEF was slightly similar to pre-reported MIP-based method which underestimated the target volume up to 18% with an average of 9% compared to the IGTV_HB when the lesions were closed to the diaphragm. In all locations, this can potentially result in marginally under-dosing.

Other scholars have investigated the feasibility of a method of using four phases to form IGTVyeo [26], [27]. They reported that the ITV4Phases may be an efficient approach alternative to optimal IGTV10 phases in SBRT for early-stage NSCLC with less tumor motion. However, their research did not include the tumors with significant motion and they could not explain applicability of their method to different motion patterns of the targets. Although it is not an optimal technique, it may still be a reasonably alternative approach when the magnitude of tumor motion was smaller than 10mm. The results deteriorated when tumor motion amplitude was more than 10mm. However, in our study, there were only 8% underestimation of volume with an average of 5% when the IGTV_DL was applied even though the magnitude of tumor motion was more than 10mm. In another study, there have been reported that the uncertainties in individualized IGTVs for SBRT can be further reduced through lumping all the datasets together from 3D-CT, 4D-CT, and MIP [28]. It was suggested that more information about the image sets have to be taken into consideration of the combination of IGTV so that uncertainties can be minimized [29]. Despite the advantages proposed, their approach requires two times CT scanning, thus introducing additional burdens in the equipment abrasion, extra workload for radiation oncologists, and the potential risk induced by the additional radiation dose to the patients.

In this study, we did not investigate the relationship between IGTV_DL, IGTV_DEF and MI in different breathing ranges directly but delivered a more meaningful and scientific for clinical practice. The probability of an optimal MI was computed and evaluated for IGTV_DL and IGTV_DEF, which were based on machine learning with DLEM and elastic deformation registration algorithm, respectively. A higher probability with an optimal MI achieved in this study leads to better clinical applicability [30]. In addition, we found that the higher probability of a better MI is not achievable by using more phases. The best way to improve the probability of optimal MI is to accurately outline the GTV of each slice of the 10 phases in 4D CT. The reason why deep leaning can synthesize IGTV_DL with a higher probability to achieve optimal MI in limited time is due to the fact that it has learned from the huge volume of GTVs delineated by experts in the process of training, which was determined in the validation cohort. This is also the reason why the result of IGTV_DL is better than the two radiologist experts in our work. In contrast, elastic deformation registration technology adopts a nonlinear image registration method. However, image processing process is still based on the CT value and gray information of the CT image [30], [31]. Once the GTV has a density or gray scale near surrounding organs, this technique started to show limitations that are similar to the MIP method. Therefore, for some large breathing range or tumor close to the mediastinum, liver or diaphragm, the machine learning method shows its merits in stability and accuracy of the delineation results.

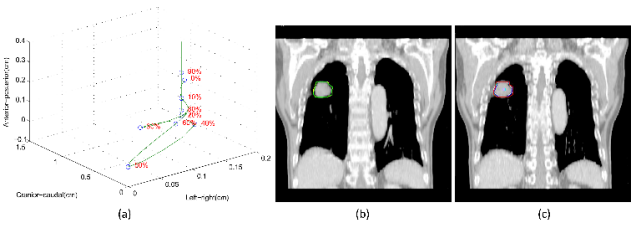


FIGURE 6. The variation trend of R_{3D} and combined IGTVs in group of R_{3D} < 5 mm. (a) the inhomogeneity of tumor centroid with respiratory motion. (b) Illustration of the differences between IGTV_E1 (red line), IGTV_E2 (yellow line), and IGTV10_DL (blue line). (c) Demonstration of the differences between IGTV_HB (light yellow line) and IGTV10_DEF (green soil line).

In this study, due to the lack of careful training for the respiration of patients, some tumor target centroid demonstrated some spatial inhomogeneity with respiratory motion, which may be a reason why the optimal phase combination have multiple solutions rather than a uniform solution (Fig 6 (a)). The 10%, 20%, 30%, 40%, and 50% phases have little change, so the difference between the sequences of these phases will be relatively small. When the breathing amplitude is in the region of 5 $\sim$ 10mm, the probability to achieve optimal MI will decrease significantly when the number of phase combinations is up to 7. This may be caused by coughing of the patient or other abnormal breathing states that is recorded by the 4DCT, which were made as a new phase into the IGTV synthesis. As the data was averaged by the patients in the group, it is more inclined to consider it was caused by systematic deviation of 4DCT. This is shown in Fig.5(c) when the respiratory amplitude is greater than 10mm. The reason why optimal MI probabilities of IGTV_DL and IGTV_DEF methods are quite different after adding new phase information requires further investigation.

In general, patients have received respiratory training before 4DCT scanning. Respiratory frequency remained relatively stable in image acquisition process. For patients without respiratory training, a further validation of our method should be carried. The highlight of this paper is the application of deep machine learning and linear exhaustive method to improve the target delineation accuracy with respiratory induced target motion involved. This framework can be further applied to hepatocellular carcinoma and breast cancer to resolve the problem of moving targets. It can also be used to try to solve the optimal treatment plan selection and dose

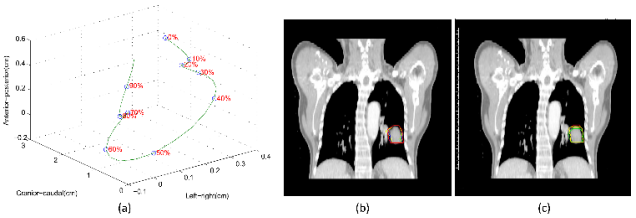


FIGURE 7. The variation trend of R_{3D} and delineated IGTVs in subgroup of $5 \leq R_{3D} < 10$ mm. (a) Demonstration of the tumor centroid of well-trained patient. (b) Illustration of the differences between IGTV10_DL (red line), IGTV10_DEF (yellow line), and IGTV10_E1 (blue line). (c) Demonstration of the differences between IGTV10_DL (red line), IGTV10_E2 (green line), IGTV_HB (light yellow line).

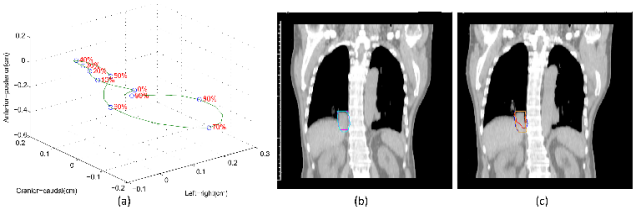


FIGURE 8. The variation trend of R_{3D} and contoured IGTVs in subgroup of $R_{3D} \geq 10$mm. (a) Illustration of the motion of tumor centroid for a patient whose tumor was located near the liver. (b) Illustration of the differences between IGTV10_DL (light blue line), IGTV_HB (light Purple). (c) Demonstration of the differences between IGTV_E1 (yellow line), and IGTV_E2 (blue line) and IGTV_DEF (red line).

matching problem. In the follow-up studies, the framework for arbitrary breathing pattern can be further investigated.

V. CONCLUSION

The deep learning algorithm with linear exhaustive optimal combination framework has a higher probability to achieve optimal MI compared to the other existing methods. For patients after simple breathing training (by keeping the respiratory frequency in 10 BPM), the four-phase combinations of 0%, 30%, 50% and 90% are recommended as an optimal combination to form IGTV in all respiration amplitude. For commercially used elastic deformation registration algorithm, the probability of the optimal MI to the reference volume decreases when the respiration amplitude exceeds 10mm. As such, it should be used with more care in clinical practice.

VI. DISCLOSURES

All authors declare no financial and non-financial competing interests.

REFERENCES

[1] K. Kameyama *et al.*, "Evaluation of the new TNM staging system proposed by the international association for the study of lung cancer at a single institution," *J. Thoracic Cardiovascular Surg.*, vol. 137, pp. 1180–1184, May 2009.

[2] S.-G. Yeo and E. S. Kim, "Efficient approach for determining four-dimensional computed tomography-based internal target volume in stereotactic radiotherapy of lung cancer," *Radiat. Oncol. J.*, vol. 31, no. 4, pp. 247–251, 2013.

[3] J. Y. Chang *et al.*, "Stereotactic ablative radiotherapy versus lobectomy for operable stage I non-small-cell lung cancer: A pooled analysis of two randomised trials," *Lancet Oncol.*, vol. 16, no. 6, pp. 630–637. Jun. 2015.

[4] K. E. Rosenzweig *et al.*, "Results of a phase I dose-escalation study using three-dimensional conformal radiotherapy in the treatment of inoperable nonsmall cell lung carcinoma," *Cancer*, vol. 103, pp. 2118–2127, May 2005.

[5] R. Muirhead, S. G. McNee, K. Moore, S. Muscat, and C. Featherstone, "Use of maximum intensity projections (MIPs) for target outlining in 4DCT radiotherapy planning," *J. Thoracic Oncol.*, vol. 3, pp. 1433–1438, Dec. 2008.

[6] K. E. Sixel, M. Ruschin, P. Tirona, and C. Cheung, "Digital fluoroscopy to quantify lung tumor motion: Potential for patient-specific planning target volumes," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 57, no. 3, pp. 717–723, 2003.

[7] J. D. Hoisak, K. E. Sixel, R. Tirona, P. C. F. Cheung, and J.-P. Pignol, "Correlation of lung tumor motion with external surrogate indicators of respiration," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 60, no. 4, pp. 1298–1306, 2004.

[8] J. Balter, R. K. T. Haken, T. S. Lawrence, and J. M. Robertson, "Uncertainties in CT-based radiation therapy treatment planning associated with patient breathing," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 36, pp. 167–174, Sep. 1996.

[9] M. Ezhil *et al.* "Determination of patient-specific internal gross tumor volumes for lung cancer using four-dimensional computed tomography," *Radiat. Oncol.*, vol 24, pp. 1234–1241, Jan. 2009.

[10] E. Rietzel, G. T. Y. Chen, C. G. Willet, and N. C. Choi, "Four-dimensional image-based treatment planning: Target volume segmentation and dose calculation in the presence of respiratory motion," *Int. J. Radiat. Oncol. Biol. Phys.*, vol 61, pp. 1535–1550, Apr. 2005.

[11] C. Zhang, Y. Xie, D. Liu, and L. Wang, "Fast threshold image segmentation based on 2D fuzzy fisher and random local optimized QPSO," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1355–1362, Mar. 2017.

[12] B. Cui, X. Ma, X. Xie, G. Ren, and Y. Ma, "Classification of visible and infrared hyperspectral images based on image segmentation and edge-preserving filtering," *Infr. Phys. Technol.*, vol. 81, pp. 79–88, Mar. 2017.

[13] M. Shirin, H. S. Rad, P. Ghafarian, A. Akbarzadeh, M. B. Karam, and M. R. Ay, "MR-guided attenuation map for prostate PET-MRI: An intensity and morphologic-based segmentation approach for generating a five-class attenuation map in pelvic region," *Ann. Nucl. Med.*, vol. 31, no. 1, pp. 9–39, 2017.

[14] E. Neu, F. Janser, A. A. Khatibi, A. C. Orifici, "Fully automated operational modal analysis using multi-stage clustering," *Mech. Syst. Signal Process.*, vol. 84, pp. 308–323, Feb. 2017.

[15] R. Timmerman, R. Paulus, and J. Galvin, "Stereotactic body radiation therapy for inoperable early stage lung cancer," *JAMA*, vol. 303, no. 11, pp. 1070–1076, 2010.

[16] F. Li *et al.*, "Geometrical differences in gross target volumes between 3DCT and 4DCT imaging in radiotherapy for non-small-cell lung cancer," *J. Radiat. Res.*, vol. 54, no. 5, pp. 950–956, Sep. 2013, doi: 10.1093/jrr/rrt017.

[17] P. Candela, B. Szegedy, and L. Vena, "On linear configurations in subsets of compact abelian groups, and invariant measurable hypergraphs," *Ann. Combinat.*, vol. 20, no. 3, pp. 487–524, 2016.

[18] D. Wang, S. Fong, R. K. Wong, S. Mohammed, J. Fiaidhi, and K. K. L. Wong, "Robust high-dimensional bioinformatics data streams mining by ODR-ioVFDT," *Sci. Rep.*, vol. 7, Feb. 2017, Art. no. 43167.

[19] L. Simon, P. Giraud, J. C. Rosenwald, and V. Servois, "Initial evaluation of a four-dimensional computed tomography system using a programmable motor," *J Appl. Clin. Med. Phys.*, vol. 7, pp. 50–65, Nov. 2006.

[20] M. Hurwitz, C. Williams, P. Mishra, S. Dhou, and J. Lewis "Applications of volumetric images generated with a respiratory motion model based on an external surrogate signal," *Med. Phys.*, vol. 42, no. 6, pp. 3658–3668, Jun. 2015.

[21] H. H. Liu *et al.*, "Assessing respiration-induced tumor motion and internal target volume using four-dimensional computed tomography for radiotherapy of lung cancer," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 68, pp. 531–540, Jun. 2007.

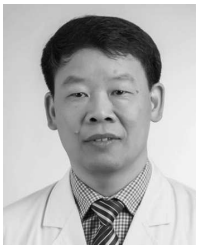
[22] Y. Seppenwoolde *et al.*, "Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 53, pp. 822–834, Jul. 2002.

[23] H. Ge, J. Cai, F. F. Yin, and C. R. Kelsey, "Quantification and minimization of uncertainties of internal target volume for stereotactic body radiation therapy of lung cancer," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 85, no. 2, pp. 438–443, 2013.

[24] S. Mahmood, H. Bilal, R. Shah, and C. Faivre-Finn, "Is stereotactic ablative radiotherapy equivalent to sublobar resection in high-risk surgical patients with stage I non-small-cell lung cancer?" *Interact. CardioVascular Thoracic Surg.*, vol. 17, no. 5, pp. 845–853, Nov. 2013.
[25] Q.-S. Chen, M. S. Weinhous, F. C. Deibel, J. P. Ciezki, and R. M. Macklis, "Fluoroscopic study of tumor motion due to breathing: Facilitating precise radiation therapy for lung cancer patients," *Med. Phys.*, vol. 28, no. 9, pp. 1850–1856, 2015.
[26] X. Yang, S. Han-Oh, M. Gui, Y. Niu, C. X. Yu, and B. Y. Yi "Four-dimensional dose distributions of step-and-shoot IMRT delivered with real-time tumor tracking for patients with irregular breathing: Constant dose rate vs dose rate regulation," *Med. Phys.*, vol. 39, no. 9, pp. 5557–5566, 2012.
[27] C. Gendrin *et al.*, "Monitoring tumor motion by real time 2D/3D registration during radiotherapy," *Radiatherapy Oncol.*, vol. 102, pp. 274–280, Feb. 2012.
[28] M. Ezhil1 *et al.*, "Determination of patient-specific internal gross tumor volumes for lung cancer using four-dimensional computed tomography," *Radiat. Oncol.*, vol. 4, p. 4, Jan. 2009, doi: 10.1186/1748-717X-4-4.
[29] P. Kennel, C. Fiorio, and F. Borne, "Supervised image segmentation using Q-shift dual-tree complex wavelet transform coefficients with a texton approach," *Pattern Anal. Appl.*, vol. 20, no. 1, pp. 227–237, 2017.
[30] L. C. Rodrigues and M. Marengoni, "Segmentation of optic disc and blood vessels in retinal images using wavelets, mathematical morphology and Hessian-based multi-scale filtering," *Biomed. Signal Process. Control*, vol. 36, pp. 39–49, Jul. 2017.
[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

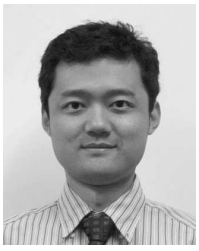
XIADONG LI was born in Zhejiang, China. He received the M.S. degree in engineering physics from Tsinghua University. He is currently pursuing the Ph.D. degree with the Institute of Transformation Medicine, Zhejiang University. For many years, he focused on the technical basis of 4DCT imaging for radiotherapy of lung cancer and the evaluation of radiotherapy and chemotherapy based on clinical characteristics and multi-parameter radiology of esophageal cancer. He holds two invention patents and four copyrights of software works. He has also developed a novel multi-parameters optimization framework based on correlation statistical analysis model applied for prostate radiotherapy planning optimization. He has presided over the Zhejiang Provincial Health Department and one key projects of the Hangzhou health plan. He is also the Vice Chairman of the Radiation Physics and Technique of Anticancer Association of Zhejiang Province. His research interests include application of artificial intelligence technology and machine learning in radiotherapy.

ZIHENG DENG is currently pursuing the bachelor's degree in biomedical engineering with Shanghai Jiao Tong University, Shanghai, China. He is an expert in image processing and data mining applied to biomedical engineering. He is involved in data processing and statistical analysis. He can build machine learning software and hardware platform skillfully and can use MATLAB and R language for secondary exploration. He participated in the Biophysics Modeling Contest during university and achieved second place in mathematical modeling for complex biophysical instance. His current research interests include image fusion, fuzzy pattern recognition, and image processing.

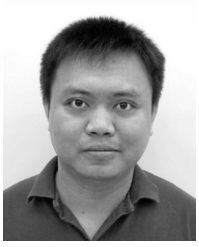


QINGHUA DENG received the bachelor's degree in clinical medicine from the Central South University of China. He is currently an Associate Professor with the Fourth Affiliated Hospital, Zhejiang Chinese Medicine University of Oncology. He was involved in clinical diagnosis and treatment of radiobiology and scientific research for over 25 years. His research fields include radiotherapy for thoracic and abdominal tumors and lymphoma. He has presided over 10 projects at all levels, participated in the National Natural Science Fund and the Jieping Wu Medical Science Foundation. He was also a Secretary General of the Radiation Oncology of Anticancer Association of Zhejiang province and a Committee Member of the Radiation Oncology of the Zhejiang Medical Association.

LIDAN ZHANG was born in Zhejiang, China. After graduating from Wenzhou Medical University, she has been involved in the medical imaging section for nearly 10 years. She is experienced in 4DCT scanning technology, especially in the selection of optimal 4DCT scanning procedures and different exposure parameters in patients with complex respiratory mode. She is currently with Hangzhou Cancer Hospital. She is the Deputy Leader of the CT Group of the Medical Imaging Society of Zhejiang Province.



TIANYE NIU received the Ph.D. degree from the University of Science and Technology of China in 2009. From 2009 to 2013, he held a post-doctoral position in the Medical Physics Program with the Georgia Institute of Technology, USA. He is currently a Faculty Member with the Zhejiang University Institute of Translational Medicine and the Affiliated Sir Run Run Shaw Hospital, Zhejiang University School of Medicine.



YU KUANG received the Ph.D. degree in biomedical engineering from Case Western Reserve University, OH, USA, in 2009. He is currently an Associate Professor of medical physics and the Director of the Professional Doctorate in Medical Physics Program (CAMPEP) at University of Nevada, Las Vegas. He is also an American Board of Radiology Board-Certified Therapeutic Medical Physicist. He completed the medical physics post-doctoral training at the University of Michigan in 2010 and Stanford University in 2012. His research focuses on the development and clinical integration of novel medical imaging devices with medical linear accelerator, proton therapy device, and carbon ion therapy device; real-time image guided and adaptive radiation therapy; combining biological and imaging biomarkers for early detection of cancers and cancer interventions; nanotechnology and its application in imaging and therapeutics; and molecular imaging for radiation biology and clinical applications. His research has been recognized by several notable awards from the American Association of Physicists in Medicine, the American Society for Radiation Oncology, the Radiological Society of North America, and the American Association for Cancer Research.