Clinical Evaluation of AI-Assisted Virtual Contrast Enhanced MRI in Primary Gross Tumor Volume Delineation for Radiotherapy of Nasopharyngeal Carcinoma

Wen Li[1], Dan Zhao[2], Zhi Chen[1], Zhou Huang[2], Saikit Lam[1], Yaoqin Xie[3], Wenjian Qin[3], Andy Lai-Yin Cheung[4], Haonan Xiao[1], Chenyang Liu[1], Francis Kar-Ho Lee[5], Kwok-Hung Au[5], Victor Ho-Fun Lee[4], Jing Cai[1,6(✉)], and Tian Li[1(✉)]

[1] The Hong Kong Polytechnic University, Hong Kong SAR, China
Jing.cai@polyu.eud.hk, litian.li@polyu.edu.hk
[2] Department of Radiation Oncology, Beijing Cancer Hospital & Institute, Beijing, China
[3] Shenzhen Institute of Advanced Technology, Chinese Academy of Science, Shenzhen, China
[4] Department of Clinical Oncology, The University of Hong Kong, Hong Kong SAR, China
[5] Department of Clinical Oncology, Queen Elizabeth Hospital, Hong Kong SAR, China
[6] Research Institute for Smart Ageing, The Hong Kong Polytechnic University, Hong Kong SAR, China

Abstract. This study aims to investigate the clinical efficacy of AI generated virtual contrast-enhanced MRI (VCE-MRI) in primary gross-tumor-volume (GTV) delineation for patients with nasopharyngeal carcinoma (NPC). We retrospectively retrieved 303 biopsy-proven NPC patients from three oncology centers. 288 patients were used for model training and 15 patients were used to synthesize VCE-MRI for clinical evaluation. Two board-certified oncologists were invited for evaluating the VCE-MRI in two aspects: image quality and effectiveness in primary tumor delineation. Image quality of VCE-MRI evaluation includes distinguishability between real contrast-enhanced MRI (CE-MRI) and VCE-MRI, clarity of tumor-to-normal tissue interface, veracity of contrast enhancement in tumor invasion risk areas, and efficacy in primary tumor staging. For primary tumor delineation, the GTV was manually delineated by oncologists. Results showed the mean accuracy to distinguish VCE-MRI from CE-MRI was 53.33%; no significant difference was observed in clarity of tumor-to-normal tissue interface between VCE-MRI and CE-MRI; for the veracity of contrast enhancement in tumor invasion risk areas and efficacy in primary tumor staging, a Jaccard Index of 76.04% and accuracy of 86.67% were obtained, respectively. The image quality evaluation suggests that the quality of VCE-MRI is approximated to real CE-MRI. In tumor delineation evaluation, the Dice Similarity Coefficient and Hausdorff Distance of the GTVs that delineated from VCE-MRI and CE-MRI were 0.762 (0.673–0.859) and 1.932 mm (0.763 mm–2.974 mm) respectively, which were clinically acceptable according to the experience of the radiation oncologists. This study demonstrated the VCE-MRI is highly promising in replacing the use of gadolinium-based CE-MRI for NPC delineation.

Keywords: MRI · Nasopharyngeal Carcinoma · Tumor Delineation

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14226, pp. 541–550, 2023.
https://doi.org/10.1007/978-3-031-43990-2_51

1 Introduction

Nasopharyngeal carcinoma (NPC), also known as lymphoepithelioma, is a highly aggressive malignancy that originates in nasopharynx [1]. NPC is characterized by a distinct geographical distribution in Southeast Asia, North Africa, and Arctic [2]. In China, NPC accounts for up to 50% of all head and neck cancers, while in Southeast Asia, NPC accounts for more than 70% of all head and neck cancers [3]. Radiotherapy (RT) is currently the main treatment remedy, which needs precise tumor delineation to ensure a satisfactory RT outcome. However, accurately delineating the NPC tumor is challenging due to the highly infiltrative nature of NPC and its complex location, which is surrounded by critical organs such as brainstem, spinal cord, temporal lobes, etc. To improve the visibility of NPC tumor for precise gross-tumor-volume (GTV) delineation, contrast-enhanced MRI (CE-MRI) is administrated through injection of gadolinium-based contrast agents (GBCAs) during MRI scanning. Despite the superior tumor-to-normal tissue contrast of CE-MRI, the use of GBCAs during MRI scanning can result in a fatal systemic disease known as nephrogenic systemic fibrosis (NSF) in patients with renal insufficiency [4]. NSF can cause severe physical impairment, such as joint contractures of fingers, elbows, and knees, and can progress to involve critical organs such as the heart, diaphragm, pleura, pericardium, kidney, liver, and lung [5]. It was reported that the incidence rate of NSF is around 4% after GBCA administration in patients with severe renal insufficiency, and the mortality rate can reach 31% [6]. Currently, there is no effective treatment for NSF, making it crucial to find a CE-MRI alternative for patients at risk of NSF.

In recent years, artificial intelligence (AI), especially deep learning, plays a game-changing role in medical imaging [7, 8], which showed great potential to eliminate the use of the toxic GBCAs through synthesizing virtual contrast-enhanced MRI (VCE-MRI) from gadolinium-free sequences, such as T1-weighted (T1w) and T2-weighted (T2w) MRI [9–12]. In 2018, Gong et al. [11] utilized pre-contrast and 10% low-dose T1w MRI to synthesize the VCE-MRI for brain disease diagnosis using a U-shape model, they found that gadolinium dose is able to be reduced by 10-fold by deep learning while the contrast information could be preserved. Followed by their work, Kleesiek et al. [10] proposed a Bayesian model to explore the feasibility of synthesizing VCE-MRI from contrast-free sequences, their study demonstrated that deep learning is highly feasible to totally eliminate the use of GBCAs. In the area of RT, Li et al. [9] developed a multi-input model to synthesize VCE-MRI for NPC RT. In addition to the advantage of eliminating the use of GBCA, VCE-MRI synthesis can also speed up the clinical workflow by eliminating the need for acquiring CE-MRI scan, which saves time for both clinical staff and patients. However, current studies mostly focus on algorithms development while lack comprehensive clinical evaluations to demonstrate the efficacy of the synthetic VCE-MRI in clinical settings.

The clinical evaluation of AI-based techniques is of paramount importance in healthcare. Rigorous clinical evaluations can establish the safety and efficacy of AI-based techniques, identify potential biases and limitations, and facilitate the integration of clinical expertise to ensure accurate and meaningful results [13]. Furthermore, the clinical evaluation of AI-based techniques can help identify areas for improvement and optimization, leading to development of more effective algorithms.

To bridge this bench-to-bedside research gap, in this study, we conducted a series of clinical evaluations to assess the effectiveness of synthetic VCE-MRI in NPC delineation, with a particular focus on assessment in VCE-MRI image quality and primary GTV delineation. This study has two main novelties: (i) To the best of our knowledge, this is the first clinical evaluation study of the VCE-MRI technique in RT; and (ii) multi-institutional MRI data were included in this study to obtain more reliable results. The success of this study would fill the current knowledge gap and provide the medical community with a clinical reference prior to clinical application of the novel VCE-MRI technique in NPC RT.

2 Materials and Methods

2.1 Data Description

Patient data was retrospectively collected from three oncology centers in Hong Kong. This dataset included 303 biopsy-proven (stage I-IVb) NPC patients who received radiation treatment during 2012–2016. The three hospitals were labelled as Institution-1 (110 patients), Institution-2 (58 patients), and Institution-3 (135 patients), respectively. For each patient, T1w MRI, T2w MRI, gadolinium-based CE-MRI, and planning CT were retrieved. MRI images were automatically registered as MRI images for each patient were scanned in the same position. The use of this dataset was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (HKU/HA HKW IRB) with reference number UW21-412, and the Research Ethics Committee (Kowloon Central/Kowloon East) with reference number KC/KE-18-0085/ER-1. Due to the retrospective nature of this study, patient consent was waived. For model development, 288 patients were used for model development and 15 patients were used to synthesize VCE-MRI for clinical evaluation. The details of patient characteristics and the number split for training and testing of each dataset were illustrated in Table 1. Prior to model training, MRI images were resampled to 256*224 by bilinear interpolation [14] due to the inconsistent matrix sizes of the three datasets.

2.2 VCE-MRI Synthesis Network

The multimodality-guided synergistic neural network (MMgSN-Net) was applied to learn the mapping from T1w MRI and T2w MRI to CE-MRI. The MMgSN-Net was a 2D network. The effectiveness of this network in VCE-MRI synthesis for NPC patients has been demonstrated by Li et al. in [9]. T1w MRI and T2w MRI were used as input and corresponding CE-MRI was used as learning target. In this work, we obtained 12806 image pairs for model training. Different from the original study, which used single institutional data for model development and utilized min-max value of the whole dataset for data normalization, in this work, we used mean and standard deviation of each individual patient to normalize MRI intensities due to the heterogeneity of the MRI intensities across institutions [15].

Table 1. Details of the multi-institutional patient characteristics. FS: field strength; TR: repetition time; TE: echo time; No.: Number; Avg: average.

Institution (Vendor-FS)	Patient No. (train/test)	Avg. age	Modality	TR (ms)	TE (ms)
Institution-1 (Siemens-1.5T)	110 (105/5)	56 ± 11	T1w	562–739	13–17
			T2w	7640	97
			CE-MRI	562–739	13–17
Institution-2 (Philips-3T)	58 (53/5)	49 ± 15	T1w	4.8–9.4	2.4–8.0
			T2w	3500–4900	50–80
			CE-MRI	4.8–9.4	2.4–8.0
Institution-3 (Siemens-3T)	135 (130/5)	57 ± 12	T1w	620	9.8
			T2w	2500	74
			CE-MRI	3.42	1.11

2.3 Clinical Evaluations

The evaluation methods used in this study included image quality assessment of VCE-MRI and primary GTV delineation. Two board-certified radiation oncologists (with 8 years' and 6 years' clinical experience, respectively) were invited to perform the VCE-MRI quality assessment and GTV delineation according to their clinical experience. Considering the clinical burden of oncologists, 15 patients were included for clinical evaluations. All clinical evaluations were performed on an Eclipse workstation (V5.0.10411.00, Varian Medical Systems, USA) with the same monitor, and the window/level can be adjusted freely by the oncologists. The results were obtained under the consensus of the two oncologists.

Image Quality Assessment of VCE-MRI. To evaluate the image quality of synthetic VCE-MRI against the real CE-MRI, we conducted four RT-related evaluations: (i) distinguishability between CE-MRI and VCE-MRI; (ii) clarity of tumor-to-normal tissue interface; (iii) veracity of contrast enhancement in tumor invasion risk areas; and (iv) efficacy in primary tumor staging. The VCE-MRI and CE-MRI volumes were imported as individual patients to Eclipse system and randomly and blindly shown to oncologists for evaluation. The MRI volumes were shown in axial view, sagittal view and coronal view, and the oncologists can scroll through the slices to view adjacent image slices.

(i) *Distinguishability between CE-MRI and VCE-MRI.* To evaluate the reality of VCE-MRI, oncologists were invited to differentiate the synthetic patients (i.e., image volumes that generated from synthetic VCE-MRI) from real patients (i.e., image volumes that generated from real CE-MRI). Different from the previous studies that utilized limited number (20-50 slices, axial view) of 2D image slices for reality evaluation [9, 10], we used 3D volumes in this study to help oncologists visualize the inter-slice adjacent information. The judgement results were recorded, and the accuracy of each institution and the overall accuracy were calculated.

(ii) *Clarity of tumor-to-normal tissue interface.* The clarity of tumor-normal tissue interface is critical for tumor delineation, which directly affects the delineation outcomes. Oncologists were asked to use a 5-point Likert scale ranging from 1 (poor) to 5 (excellent) to evaluate the clarity of tumor-to-normal tissue interface. Paired two-tailed t-test (with a significance level of $p = 0.05$) was applied to analyses if the scores obtained from real patients and synthetic patients are significantly different.

(iii) *Veracity of contrast enhancement in tumor invasion risk areas.* In addition to the critical tumor-normal tissue interface, the areas surrounding the NPC tumor will also be considered during delineation. To better evaluate the veracity of contrast enhancement in VCE-MRI, we selected 25 tumor invasion risk areas according to [16], including 13 high-risk areas and 12 medium-risk areas, and asked oncologists to determine whether these areas were at risk of being invaded according to the contrast-enhanced tumor regions. The 13 high-risk areas include: retropharyngeal space, parapharyngeal space, levator veli palatine muscle, prestyloid compartment, Tensor veli palatine muscle, poststyloid compartment, nasal cavity, pterygoid process, basis of sphenoid bone, petrous apex, prevertebral muscle, clivus, and foramen lacerum. The 12 medium-risk areas include foramen ovale, great wing of sphenoid bone, medial pterygoid muscle, oropharynx, cavernous sinus, sphenoidal sinus, pterygopalatine fossa, lateral pterygoid muscle, hypoglossal canal, foramen rotundum, ethmoid sinus, and jugular foramen. The areas considered at risk of tumor invasion were recorded.

The Jaccard index (JI) [17] was utilized to quantitatively evaluate the results of recorded risk areas from CE-MRI and VCE-MRI. The JI could be calculated by:

$$\mathrm{JI} = |R_{CE} \cap R_{VCE}| / |R_{CE} \cup R_{VCE}| \tag{1}$$

where R_{CE} and R_{VCE} represents the set of risk areas that recorded from CE-MRI and corresponding VCE-MRI, respectively. JI measures similarity of two datasets, which ranges from 0% to 100%. Higher JI indicates more similar of the two sets.

(iv) *Efficacy in primary tumor staging.* A critical RT-related application of CE-MRI is tumor staging, which plays a critical role in treatment planning and prognosis prediction [18]. To assess the efficacy of VCE-MRI in NPC tumor staging, oncologists were asked to determine the stage of the primary tumor shown in CE-MRI and VCE-MRI. The staging results from CE-MRI were taken as the ground truth and the staging accuracy of VCE-MRI was calculated.

Primary GTV Delineation. GTV delineation is the foremost prerequisite for a successful RT treatment of NPC tumor, which demands excellent precision [19]. An accurate tumor delineation improves local control and reduce toxicity to surrounding normal tissues, thus potentially improving patient survival [20]. To evaluate the feasibility of eliminating the use of GBCA by replacing CE-MRI with VCE-MRI in tumor delineation, oncologists were asked to contour the primary GTV under assistance of VCE-MRI. For comparison, CE-MRI was also imported to Eclipse for tumor delineation but assigned as a different patient, which were shown to oncologists in a random and blind manner. To mimic the real clinical setting, contrast-free T1w, T2w MRI and corresponding CT

of each patient were imported into the Eclipse system since sometimes T1w and T2w MRI will also be referenced during tumor delineation. Due to both real patients and synthetic patients were involved in delineation, to erase the delineation memory of the same patient, we separated the patients to two datasets, each with the same number of patients, both two datasets with mixed real patients and synthetic patients without overlaps (i.e., the CE-MRI and VCE-MRI from the same patient are not in the same dataset).When finished the first dataset delineation, there was a one-month interval before the delineation of the second dataset. After the delineation of all patients, the Dice similarity coefficient (*DSC*) [21] and Hausdorff distance (*HD*) [22] of the GTVs delineated from real patients and corresponding synthetic patients were calculated to evaluate the accuracy of delineated contours.

Dice Similarity Coefficient (DSC). *DSC* is a broadly used metric to compare the agreement between two segmentations [23]. It measures the spatial overlap between two segmentations, which ranges from 0 (no spatial overlap) to 1 (complete overlap). The *DSC* can be expressed as:

$$DSC = 2 * |C_{CE} \cap C_{VCE}|/(|C_{CE}| + |C_{VCE}|) \tag{2}$$

where C_{CE} and C_{VCE} represent the contours delineated from real patients and synthetic patients, respectively.

Hausdorff Distance (HD). Even though DSC is a well-accepted segmentation comparison metric, it is easily influenced by the size of contours. Small contours typically receive lower *DSC* than larger contours [24].Therefore, *HD* was applied as a supplementary to make a more thorough comparison. *HD* is a metric to measure the maximum distance between two contours. Given two contours C_{CE} and C_{VCE}, the HD could be calculated as:

$$HD = max(max_{x \in C_{CE}} \mathrm{d}(x, C_{VCE}), max_{y \in C_{VCE}} \mathrm{d}(y, C_{CE})) \tag{3}$$

where $\mathrm{d}(x, C_{VCE})$ and $\mathrm{d}(y, C_{CE})$ represent the distance from point x in contour C_{CE} to contour C_{VCE} and the distance from point y in contour C_{VCE} to contour C_{CE}.

3 Results and Discussion

3.1 Image Quality of VCE-MRI

Table 2 summarizes the results of the four VCE-MRI quality evaluation metrics, including: (i) distinguishability between CE-MRI and VCE-MRI; (ii) clarity of tumor-to-normal tissue interface; (iii) veracity of contrast enhancement in tumor invasion risk areas; and (iv) efficacy in primary tumor staging.

(i) **Distinguishability between CE-MRI and VCE-MRI.** The overall judgement accuracy for the MRI volumes was 53.33%, which is close to a random guess accuracy (i.e., 50%). For Institution-1, 2 real patients were judged as synthetic and 1 synthetic patient was considered as real. For Institution-2, 2 real patients

were determined as synthetic and 4 synthetic patients were determined as real. For Institution-3, 2 real patients were judged as synthetic and 3 synthetic patients were considered as real. In total, 6 real patients were judged as synthetic and 8 synthetic patients were judged as real.

(ii) **Clarity of tumor-to-normal tissue interface.** The overall clarity scores of tumor-to-normal tissue interface for real and synthetic patients were 3.67 with a median of 4 and 3.47 with a median of 4, respectively. No significant difference was observed between these two scores ($p = 0.38$). The average scores for real and synthetic patients were 3.6 and 3, 3.6 and 3.8, 3.8 and 3.6 for Institution-1, Institution-2, and Institution-3, respectively. 5 real patients got a higher score than synthetic patients and 3 synthetic patients obtained a higher score than real patients. The scores of the other 7 patient pairs were the same.

(iii) **Veracity of contrast enhancement in tumor invasion risk areas.** The overall *JI* score between the recorded tumor invasion risk areas from CE-MRI and VCE-MRI was 74.06%. The average *JI* obtained from Institution-1, Institution-2, and Institution-3 dataset were similar with a result of 71.54%, 74.78% and 75.85%, respectively. In total, 126 risk areas were recorded from the CE-MRI for all of the evaluation patients, while 10 (7.94%) false positive high risk invasion areas and 9 (7.14%) false negative high risk invasion areas were recorded from VCE-MRI.

(iv) **Efficacy in primary tumor staging.** A T-staging accuracy of 86.67% was obtained using VCE-MRI. 13 patient pairs obtained the same staging results. For the Institution-2 data, all synthetic patients observed the same stages as real patients. For the two T-stage disagreement patients, one synthetic patient was staged as phase IV while the corresponding real patient was staged as phase III, the other synthetic patient was staged as I while corresponding real patient was staged as phase III.

Table 2. Image quality evaluation results of VCE-MRI: (A) Distinguishability between CE-MRI and VCE-MRI; (B) Clarity of tumor-to-normal tissue interface; (C) Veracity of contrast enhancement in risk areas; and (D) T-staging. Abbreviations: Inst: Institution; C.A.: Center-based average; O.A.: Overall average; Syn: Synthetic.

	(A)			(B)					
	Inst-1	Inst-2	Inst-3	Inst-1		Inst-2		Inst-3	
	/	/	/	Real	Syn	Real	Syn	Real	Syn
C.A.	70%	40%	50%	3.6	3	3.6	3.8	3.8	3.6
O.A.	53.33%			Real: 3.67			Syn: 3.47		
	(C)			(D)					
	Inst-1	Inst-2	Inst-3	Inst-1		Inst-2		Inst-3	
C.A.	71.54%	74.78%	75.85%	80%		100%		80%	
O.A.	74.06%			86.67%					

Figure 1 illustrates an example of the synthetic VCE-MRI. Compared to CE-MRI, T1w MRI has the similar tumor shape but with indistinguishable tumor-to-normal tissue interface, while T2w MRI shows superior lesion contrast but with smaller tumor volume (yellow boxes). The deep learning model integrated the complementary information of T1w MRI and T2w MRI, and successfully synthesized VCE-MRI with similar contrast and tumor volume as CE-MRI, with no obvious contrast differences in tumor regions, as shown in difference map between CE-MRI and VCE-MRI.

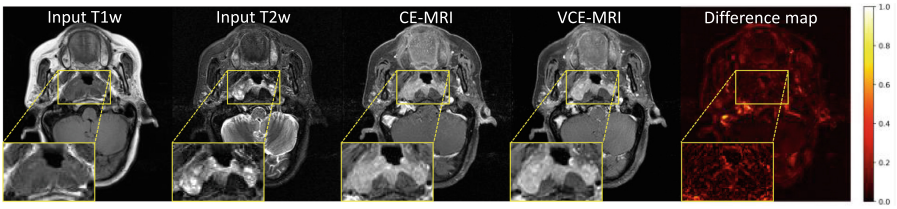

Fig. 1. Illustration of the synthetic VCE-MRI.

3.2 Primary GTV delineation

The average *DSC* and *HD* between the C_{CE} and C_{VCE} was 0.762 (0.673–0.859) with a median of 0.774, and 1.932 mm (0.763 mm–2.974 mm) with a median of 1.913 mm, respectively. For Institution-1, Institution-2, and Institution-3, the average *DSC* were 0.741, 0.794 and 0.751 respectively, while the average *HD* were 2.303 mm, 1.456 mm, and 2.037 mm respectively. Figure 2 illustrated the delineated primary GTV contours from an average patient with the *DSC* of 0.765 and *HD* of 1.938 mm. The green contour shows the primary GTV that delineated form the synthetic patient, while the red contour was delineated from corresponding real GBCA-based patient.

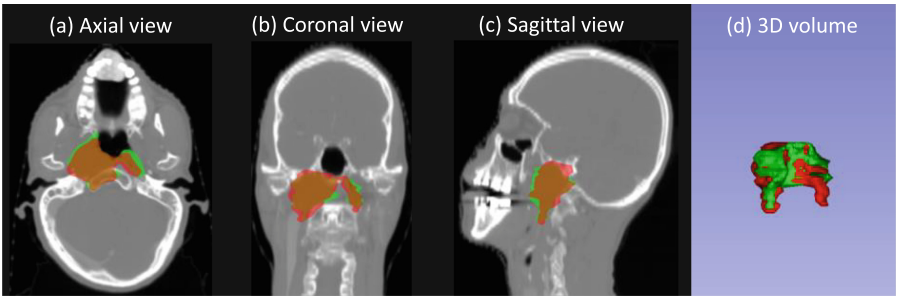

Fig. 2. Illustration of the primary GTVs from a typical patient with an average *DSC* and *HD*.

4 Conclusion

In this study, we conducted a series of clinical evaluations to validate the clinical efficacy of VCE-MRI in RT of NPC patients. Results showed the VCE-MRI has great potential to provide an alternative to GBCA-based CE-MRI for NPC delineation.

Acknowledgement. This research was partly supported by research grants of General Research Fund (GRF 15102219, GRF 15103520), the University Grants Committee, and Project of Strategic Importance Fund (P0035421), Projects of RISA (P0043001), One-line Budget (P0039824, P0044474), The Hong Kong Polytechnic University, and Shenzhen-Hong Kong-Macau S&T Program (Category C) (SGDX20201103095002019), Shenzhen Basic Research Program (R2021A067), Shenzhen Science and Technology Innovation Committee (SZSTI).

Data use Declaration:. The use of this dataset was approved by the Institutional Review Board of University of Hong Kong/Hospital Authority Hong Kong West Cluster (HKU/HA HKW IRB) with reference number UW21-412, and the Research Ethics Committee (Kowloon Central/Kowloon East) with reference number KC/KE-18-0085/ER-1. Due to the retrospective nature of this study, patient consent was waived.

References

1. Chen, Y.-P., et al.: Nasopharyngeal carcinoma. The Lancet **394**(10192), 64–80 (2019)
2. Chang, E.T., et al.: The evolving epidemiology of nasopharyngeal carcinoma. Cancer Epidemiol. Biomarkers Prev. **30**(6), 1035–1047 (2021)
3. Sturgis, E.M., Wei, Q., Spitz, M.R.: Descriptive epidemiology and risk factors for head and neck cancer. In: Seminars in Oncology. Elsevier (2004)
4. Sadowski, E.A., et al.: Nephrogenic systemic fibrosis: risk factors and incidence estimation. Radiology **243**(1), 148–157 (2007)
5. Thomsen, H.S.: Nephrogenic systemic fibrosis: a serious late adverse reaction to gadodiamide. Eur. Radiol. **16**(12), 2619–2621 (2006)
6. Schlaudecker, J.D., Bernheisel, C.R.: Gadolinium-associated nephrogenic systemic fibrosis. Am. Fam. Phys. **80**(7), 711–714 (2009)
7. Qin, W., et al.: Superpixel-based and boundary-sensitive convolutional neural network for automated liver segmentation. Phys. Med. Biol. **63**(9), 095017 (2018)
8. Qin, W., et al.: Automated segmentation of the left ventricle from MR cine imaging based on deep learning architecture. Biomed. Phys. Eng. Express **6**(2), 025009 (2020)
9. Li, W., et al.: Virtual contrast-enhanced magnetic resonance images synthesis for patients with nasopharyngeal carcinoma using multimodality-guided synergistic neural network. Int. J. Radiat. Oncol. Biol. Phys. **112**(4), 1033–1044 (2022)
10. Kleesiek, J., et al.: Can virtual contrast enhancement in brain MRI replace gadolinium? A feasibility study. Invest. Radiol. **54**(10), 653–660 (2019)
11. Gong, E., et al.: Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. J. Magn. Reson. Imaging **48**(2), 330–340 (2018)
12. Li, W., et al. Multi-institutional investigation of model generalizability for virtual contrast-enhanced MRI synthesis. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention–MICCAI 2022. MICCAI 2022. LNCS. Vol. 13437, pp. 765–773. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16449-1_73

13. Topol, E.J.: High-performance medicine: the convergence of human and artificial intelligence. Nat. Med. **25**(1), 44–56 (2019)
14. Gribbon, K.T., Bailey, D.G.: A novel approach to real-time bilinear interpolation. In: Proceedings. DELTA 2004. Second IEEE International Workshop on Electronic Design, Test and Applications. IEEE (2004)
15. Li, W., et al.: Model generalizability investigation for GFCE-MRI synthesis in NPC radiotherapy using multi-institutional patient-based data normalization (2022). TechRxiv.Preprint
16. Liang, S.-B., et al.: Extension of local disease in nasopharyngeal carcinoma detected by magnetic resonance imaging: improvement of clinical target volume delineation. Int. J. Radiat. Oncol. Biol. Phys. **75**(3), 742–750 (2009)
17. Fletcher, S., Islam, M.Z.: Comparing sets of patterns with the Jaccard index. Australas. J. Inf. Syst. **22** (2018)
18. Lee, A.W., et al.: International guideline for the delineation of the clinical target volumes (CTV) for nasopharyngeal carcinoma. Radiother. Oncol. **126**(1), 25–36 (2018)
19. Jager, E.A., et al.: GTV delineation in supraglottic laryngeal carcinoma: interobserver agreement of CT versus CT-MR delineation. Radiat. Oncol. **10**(1), 1–9 (2015)
20. Jameson, M.G., et al.: Correlation of contouring variation with modeled outcome for conformal non-small cell lung cancer radiotherapy. Radiother. Oncol. **112**(3), 332–336 (2014)
21. Balagopal, A., et al.: A deep learning-based framework for segmenting invisible clinical target volumes with estimated uncertainties for post-operative prostate cancer radiotherapy. Med. Image Anal. **72**, 102101 (2021)
22. Yang, J., et al.: A multimodality segmentation framework for automatic target delineation in head and neck radiotherapy. Med. Phys. **42**(9), 5310–5320 (2015)
23. Chang, H.-H., et al.: Performance measure characterization for evaluating neuroimage segmentation algorithms. Neuroimage **47**(1), 122–135 (2009)
24. Schreier, J., et al.: Clinical evaluation of a full-image deep segmentation algorithm for the male pelvis on cone-beam CT and CT. Radiother. Oncol. **145**, 1–6 (2020)