

# Automatic Tumor Segmentation With a Convolutional Neural Network in Multiparametric MRI: Influence of Distortion Correction

Lars Bielak<sup>1,3</sup>, Nicole Wiedenmann<sup>2,3</sup>, Nils Henrik Nicolay<sup>2,3</sup>, Thomas Lottner<sup>1</sup>, Johannes Fischer<sup>1</sup>, Hatice Bunea<sup>2,3</sup>, Anca-Ligia Grosu<sup>2,3</sup>, and Michael Bock<sup>1,3</sup>

Departments of <sup>1</sup>Radiology, Medical Physics; <sup>2</sup>Radiation Oncology, Medical Center University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany; and <sup>3</sup>German Cancer Consortium (DKTK), Partner Site Freiburg, Freiburg, Germany

## Corresponding Author:

Lars Bielak, MSc

Department of Radiology—Medical Physics, University Medical Center Freiburg, Killianstr. 5a, 79106 Freiburg, Germany;

E-mail: lars.bielak@uniklinik-freiburg.de

**Key Words:** Multi-parametric MRI, radiation therapy, automatic tumor segmentation, convolutional neuronal network

**Abbreviations:** Convolutional neural networks (CNNs), apparent diffusion coefficient (ADC), magnetic resonance imaging (MRI), neural networks (NNs), diffusion-weighted imaging (DWI), echo planar imaging (EPI), magnetic resonance (MR), phase-encoding (PE), gross tumor volumes (GTVs)

## ABSTRACT

Precise tumor segmentation is a crucial task in radiation therapy planning. Convolutional neural networks (CNNs) are among the highest scoring automatic approaches for tumor segmentation. We investigate the difference in segmentation performance of geometrically distorted and corrected diffusion-weighted data using data of patients with head and neck tumors; 18 patients with head and neck tumors underwent multiparametric magnetic resonance imaging, including T2w, T1w, T2\*, perfusion ( $k_{\text{trans}}$ ), and apparent diffusion coefficient (ADC) measurements. Owing to strong geometrical distortions in diffusion-weighted echo planar imaging in the head and neck region, ADC data were additionally distortion corrected. To investigate the influence of geometrical correction, first 14 CNNs were trained on data with geometrically corrected ADC and another 14 CNNs were trained using data without the correction on different samples of 13 patients for training and 4 patients for validation each. The different sets were each trained from scratch using randomly initialized weights, but the training data distributions were pairwise equal for corrected and uncorrected data. Segmentation performance was evaluated on the remaining 1 test-patient for each of the 14 sets. The CNN segmentation performance scored an average Dice coefficient of  $0.40 \pm 0.18$  for data including distortion-corrected ADC and  $0.37 \pm 0.21$  for uncorrected data. Paired *t* test revealed that the performance was not significantly different ( $P = .313$ ). Thus, geometrical distortion on diffusion-weighted imaging data in patients with head and neck tumor does not significantly impair CNN segmentation performance in use.

## INTRODUCTION

Precise delineation and segmentation of tumors is an essential step in radiation therapy planning. Good segmentation accuracy is a prerequisite for both effective tumor treatment and preservation of functionality of surrounding healthy tissue and thereby for prolonged patient survival (1, 2). Manual segmentation of lesions is a tedious task, and hence automatic detection methods have been proposed as tools for diagnostics, treatment planning and response evaluation (3). With these automatic segmentation methods, problems such as interobserver variability in target volume definition, definition and assessment of tumor heterogeneity, and tumor classification may be overcome (4, 5).

Early segmentation solutions were focused on image signal intensity-based methods or semiautomatic computer learning algorithms with manually selected or linearly learned image features (6–13). Many of these segmentation methods made use

of multiparametric imaging based on data from multiple cross-sectional imaging modalities (eg, positron emission tomography, magnetic resonance imaging [MRI], computed tomography). A key feature of MRI however is the possibility to create multiparametric imaging data in a single modality and in a single imaging session—thus, physical, functional, and anatomical features can be imaged during the same examination session and in a (nearly) identical patient position, which facilitates the alignment of image data before segmentation.

Today, the highest scoring algorithms for automatic tumor segmentation use (convolutional) neural networks [(CNNs) (14)]. NNs feed a set of input data through a number of processing layers, where each layer consists of a number of neurons that are activated by a nonlinear function depending on a linear combination of input data and a bias. With increasing number of layers, the ability to represent nonlinear relationships between input and output increases, effectively enabling a deep NN to

learn any functional relationship given enough input data is available. In addition, a CNN is capable of implementing contextual information and can therefore learn high representations of the data such as edge information (15).

In multiparametric MRI for tumor segmentation, different anatomical image contrasts (T1- and T2-weighted) are combined with functional information acquired with perfusion and diffusion measurements. Particularly diffusion-weighted imaging (DWI) has proven to contribute valuable additional information for tumor delineation (16-18). For the DWI images, an echo planar imaging (EPI) pulse sequence is commonly used. Despite its advantages, the EPI technique has a major disadvantage: it is very sensitive to off-resonances caused by inhomogeneity in the  $B_0$  magnetic field, which leads to severe geometrical distortions (19). Several groups have worked on solutions on the pulse sequence level, such as readout segmented (rs)EPI (20) which has been shown to dramatically decrease image distortions (21, 22). As these methods alone cannot remove image distortions completely, the necessity to quantify the effect of image distortion on automatic tumor segmentation becomes evident.

Image distortions are especially pronounced in MRI of head and neck tumors, where the complex geometry of head, neck, and shoulders severely limits  $B_0$  shimming. This results in an increased field inhomogeneity and thus stronger image distortions than in other body regions like the brain. In addition, in tumors with hypoxic subareas, [18F]-fluoromisonidazole positron emission tomography can be used as a metabolic marker for hypoxia localization (23-25), which is important for individualized treatment schemes, for example, by dose painting. In these patients, MRI would be a desirable imaging alternative if the effect of geometric distortion on tumor segmentation performance could be controlled.

In this work, CNNs were used for the segmentation of multiparametric MRI data of patients with head and neck tumor, and the effects of geometric distortion of diffusion-weighted input data on the segmentation performance were analyzed.

## MATERIALS AND METHODS

### Head and Neck Tumor Patient Trial

Patient data were taken from a prospective clinical trial in patients with head and neck squamous cell carcinoma, which helped investigate the correlations between tumor response un-

der radiotherapy and hypoxic tumor subvolumes in patients with head and neck squamous cell carcinomas. Written informed consent was obtained from each patient, and the institutional review board approved the study (Approval No. 479/12). Patients received anatomical and functional MRI before undergoing radiochemotherapy and 2 and 5 weeks into treatment. In this work, the pretherapeutic MRI data were used for analysis to avoid therapy-related bias. In total, multiparametric MRI data from 18 patients were available.

For MRI, a clinical 3 T whole-body magnetic resonance (MR) system (Siemens Tim Trio, Erlangen, Germany) was used. Patients were placed in an individually fitted therapy mask, which was fixed at the patient couch of the MR system. A flexible receive coil was wrapped around the anterior part of the neck, which was used in combination with the additional spine array coils for MR signal reception. The MR protocol of the study consisted of anatomical T1w and T2w MRI, T2\* maps from multiecho gradient echo MRI, perfusion MRI including the vascular permeability  $k_{trans}$ , quantified using contrast-enhanced dynamic T1-weighted MRI, and the apparent diffusion coefficient (ADC) that was assessed with diffusion-weighted echo-planar imaging. DWI data were acquired using standard and readout-segmented diffusion-weighted EPI sequences. Conventional EPI was used with an echo time of 69 ms, acquisition time (TA) = 5 min, while the rsEPI (readout segmentation of long variable echo-trains, RESOLVE) sequence used echo time = 51 ms, TA = 7 min, with 7 segments. Both diffusion sequences used a 3-direction trace scan with b-values of 50, 400, and 800 s/mm<sup>2</sup> to quantify the ADC, with phase-encoding (PE) along the anterior-posterior direction. All relevant sequence specifications are listed in Table 1.

### Data Preprocessing

Owing to additional acquisition times, only 12 of 18 patients tolerated the additional rsEPI protocol. If available, rsEPI images were used in the study. For the other patients, conventional EPI images were used.

Perfusion  $k_{trans}$  was determined according to the Tofts model (26). Both  $k_{trans}$  and T2\* were calculated with the software platform SyngoVia (Siemens Healthcare), while monoexponentially fitted ADC-maps were determined with the MR systems' postprocessing software. To improve the performance of the sub-

**Table 1.** List of Input Channels and Corresponding Sequence Details

Sequence	TE [ms]	TR [ms]	Resolution [mm <sup>3</sup> ]	Comments/Other
T1 Fast Spin Echo	11	504	0.7 × 0.7 × 4.0	
T2 Fast Spin Echo	100	5000	0.7 × 0.7 × 4.0	
Multi-Echo GRE	5-33	600	1.1 × 1.1 × 3.0	$n_{Echoes} = 12$ , reconstructed map: T2*
Dynamic T1w Perfusion Measurement	1.56	4.65	1.4 × 1.4 × 3.0	$n_{Timepoints} = 36$ , reconstructed map: $k_{trans}$
DWI (rsEPI)	51	2510	2 × 2 × 3	$b = \{50, 400, 800\}$ s/mm <sup>2</sup> , reconstructed map: ADC, $n_{Segments} = 7$
DWI (Conventional EPI)	69	3500	2 × 2 × 3	$b = \{50, 400, 800\}$ s/mm <sup>2</sup> , reconstructed map: ADC

sequent CNN analysis and to ensure comparability between subjects, T1- and T2-weighted images were normalized to zero mean and unitary standard deviation. Images were then interpolated to a 1 mm isotropic resolution using cubic splines, and image coregistration was performed using standard MATLAB (The MathWorks, Natick, MA; Version 2016b) tools (eg, `imregister`), based on similarity transformations with a mutual information metric.

### Additional DWI Preprocessing

Head and neck regions are especially challenging for DWI as the complex geometry imposes severe limitations to magnetic field shimming (27). To study the influence of the geometric accuracy on the CNN performance, an algorithm was developed to geometrically correct ADC maps.

The problem of geometric distortions between 2 MR images due to field inhomogeneities is well known, but the imaging protocol did not include additional field map measurements so standard correction schemes (19) could not be applied. Instead, a postprocessing method from optical microscopy (28), 2-photon imaging (29, 30) or particle imaging techniques (31), was adapted that has been developed to correct for nonrigid motion in between acquisitions. The distorted DWI and a geometrically more precise T2w image are treated as 2 images of the same region. The distortion field in between the 2 images is then estimated according to the Lucas-Kanade (32) method implemented in a pyramidal layout (33). Our MATLAB 3D implementation of the algorithm makes use of the mutual information metric to account for the different contrasts of the images. As distortions are expected in only the PE direction owing to the low effective PE bandwidth, the spatial degrees of freedom in the distortion field were limited to the PE direction.

The implementation was validated with volunteer data acquired using a 3 T MRI system (Tim Trio, Siemens, Healthineers) using T2w and DWI contrasts together with a  $B_0$  field map. With the correction algorithm, geometrically corrected ADC maps were calculated for all 18 patients as an additional preprocessing step for the CNN analysis. Distortion fields were extracted from the  $b = 50 \text{ s/mm}^2$  images only, as the low  $b$ -value provides optimal signal-to-noise ratio, and the same image distortion is expected at higher  $b$ -values.

### CNN

Finally, a 3D CNN was configured to perform the segmentation task on the patient data. To study the effect of image distortion on the segmentation result, 2 separate NNs were trained: the first network included the original, uncorrected ADC maps, while the second used the geometrically corrected ADC maps.

For the calculations, the DeepMedic (34) CNN architecture was used. DeepMedic is a 3D CNN which uses 2 calculation pathways, a normal one and one with 3 times lower spatial resolution, to combine local fine structure with coarser contextual image information. Each pathway consisted of 8 hidden layers with  $\{40 \ 40 \ 50 \ 50 \ 60 \ 60 \ 70 \ 70\}$  channels using  $3^3$  kernel sizes followed by 2 fully connected layers of 100 channels each, which combine high- and low-resolution pathways. In this layout, the following 5 input channels were used: T1-weighted images, T2-weighted images,  $k_{\text{trans}}$  maps,  $T2^*$  maps, and ADC maps. As ground truth, gross tumor volumes (GTVs) were used that were contoured by a radiation oncologist and a radiologist on the basis of MR data. For contour-

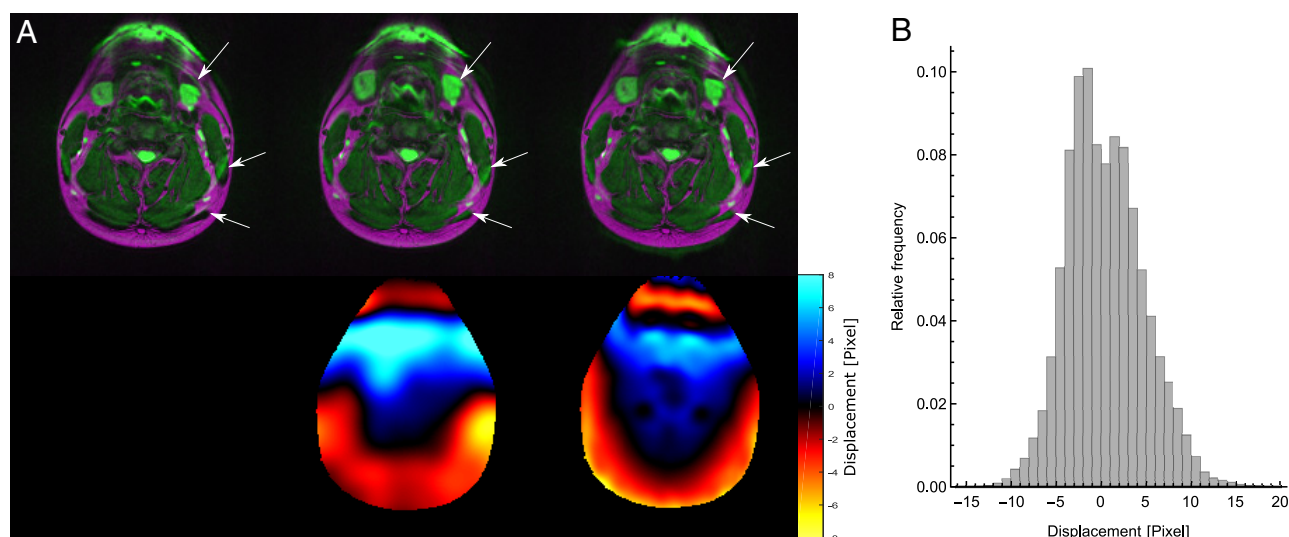
ing, all original MR data were available; however, most volumes were drawn on the basis of T1w imaging and copied to all other contrasts in the process.

The data were divided in groups, with 13 patients in the training set, 4 patients in the validation set, and 1 patient in the testing set. A leave-1-out cross-validation was performed for 14 test patients, both with and without geometrically corrected ADC data. For better comparability, the 14 uncorrected and corrected data samples were chosen to have pairwise equal distributions in validation, training and testing sets. Using this set of networks, a statistical analysis for the 2 cases was used using the Dice coefficient as a measure for segmentation performance. The Dice coefficient is calculated as  $Dice = 2 TP / (2 TP + FN + FP)$  (35), where TP are true positives, FN false negatives, and FP false positives. A paired  $t$  test on the resulting Dice coefficients for the 14 training cases was used to test whether a significant difference could be observed.

### RESULTS

The verification of the distortion correction algorithm on the randomly distorted MR-image showed a substantial decrease in Euclidean image distance from  $0.69 \pm 0.06$  to  $0.21 \pm 0.03$ . The volunteer experiment shows that the algorithm reproduces the general structure of a measured field map with minor deviations in the fine structures (Figure 1A). The Euclidean image distance between the measured field map and the calculated distortion field amounts to  $2.1 \pm 2.3$  pixel. In few regions of strong distortions, for example, on the boundaries of the trachea, distortions are so severe that both registration methods do not deliver clinically acceptable results; however, this was the case in only 6 patients and it equally affected corrected and uncorrected data. As these irreversible distortions affect only parts of an image, the corresponding cases could still be used in the evaluation process. Figure 1 shows the results of subsequent correction—both methods realign anatomical areas well with the corresponding T2w reference image, while severe misalignments are seen without correction. The calculated distortion fields for all patient cases measure a total mean of 0.46 and a standard deviation of 4.24 pixels, which clearly illustrate the need for correction (Figure 1B).

The CNN was trained on the patient data for 35 epochs per sample case. Figure 2 shows the training progress for an exemplary case. The training progress appears to be largely the same for both input cases of corrected and uncorrected ADC data. However, as can be seen in the validation curve, there is a noticeable difference, especially in the sensitivity metric between the two cases. Figure 3 shows the subsequent segmentation result of the corresponding test sample. Both methods, with and without distortion correction, labeled some areas far from the GTV as tumor tissue, but in general, a good overlap between the ground truth (GTV) and the segmentation results with and without distortion correction was found with Dice coefficients up to 0.68 and 0.65, respectively. Figure 4 shows the segmentation performance over all test sessions in a scatter plot. As seen, despite the presence of severe image distortions in the ADC maps, the distortion correction did improve the segmentation performance of the CNN, however, not to a statistically significant degree ( $P = .313$ ). The mean Dice coefficient for segmentation with distortion-corrected ADC-maps was  $0.40 \pm 0.18$ , while for uncorrected data, it amounted to  $0.37 \pm 0.21$ .

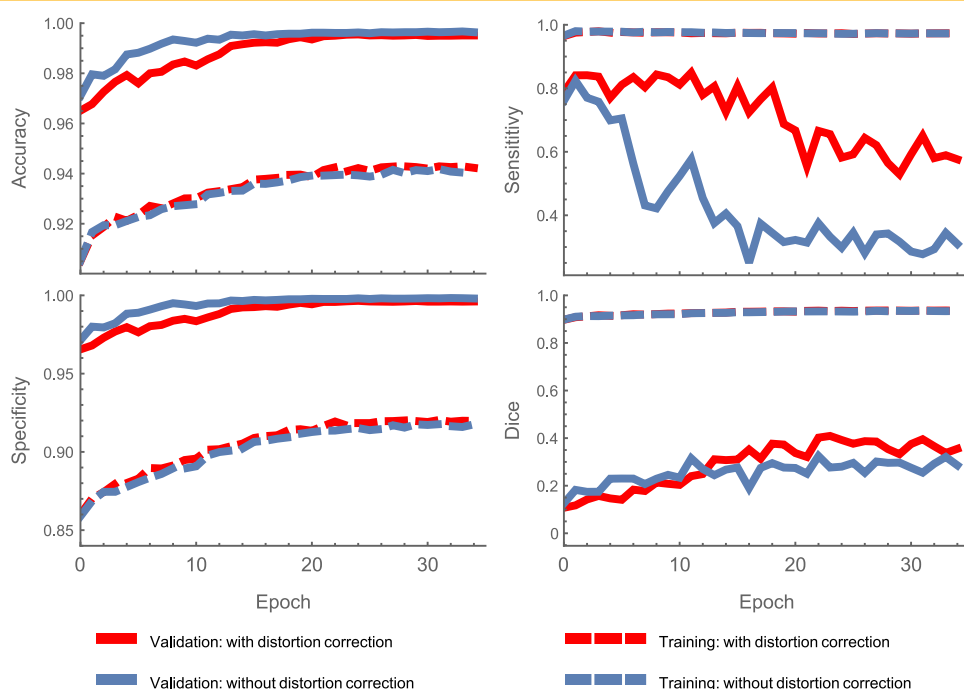


**Figure 1.** (A) Top: Overlay of T2-weighted (T2w) image (purple) and readout segmented echo planar imaging (rsEPI)-image (green). Left: Original image with distortions. Center: Corrected diffusion-weighted imaging (DWI) using the correction algorithm with the T2w image as a reference. Right: Corrected DWI using a measured  $B_0$  field map for correction. Bottom: The corresponding distortion fields used for correction. Both fields show the same general behavior, while some fine structure, especially in regions of strong distortions around the trachea, cannot be resolved using the algorithm. White arrows mark locations where the misalignment of T2w and DWI is clearly seen. (B) A histogram showing the relative amount of displacements within all diffusion images that were included in the study. The standard deviation is 4.2 pixels, which shows the large effect of the distortion correction.

## DISCUSSION

In this work, a CNN was defined and trained to segment head and neck tumors using clinical data from patients undergoing radiation therapy. In particular, 2 input cases were compared with respect to the segmentation performance: 1 with geometric

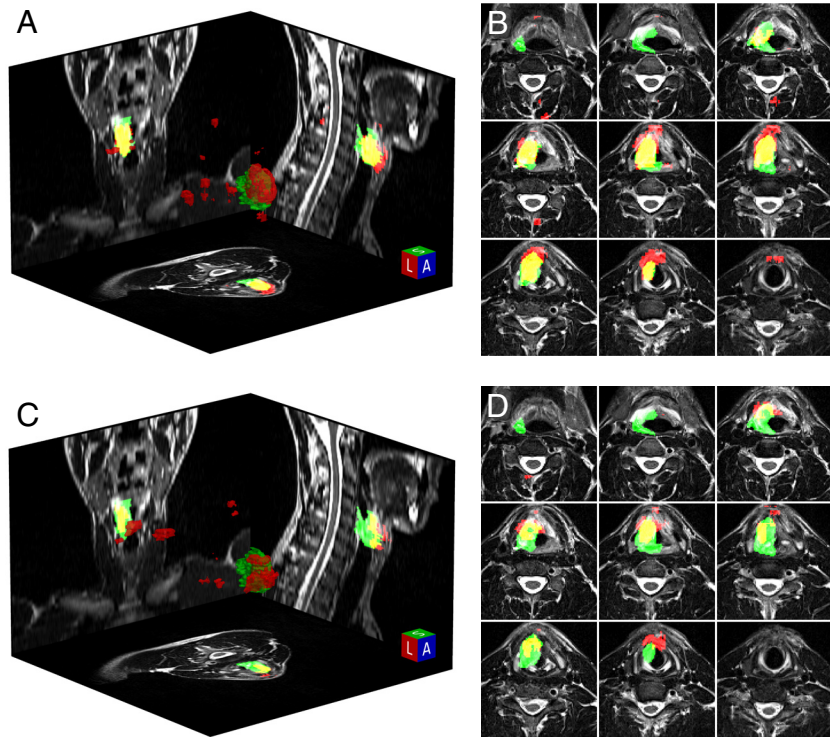
distortion correction of the input DWI data, and 1 without. In this study with 18 patients already a good segmentation could be achieved, and no significant differences between the distortion-corrected and -uncorrected cases were found with regard to the segmentation performance.



**Figure 2.** Training process of the convolutional neural network (CNN) for 1 training example. After training for 35 epochs, the network seemed to have reached peak performance. The plots for corrected and uncorrected training data show great similarity, which is reflected in the comparison of Dice coefficients for testing data.

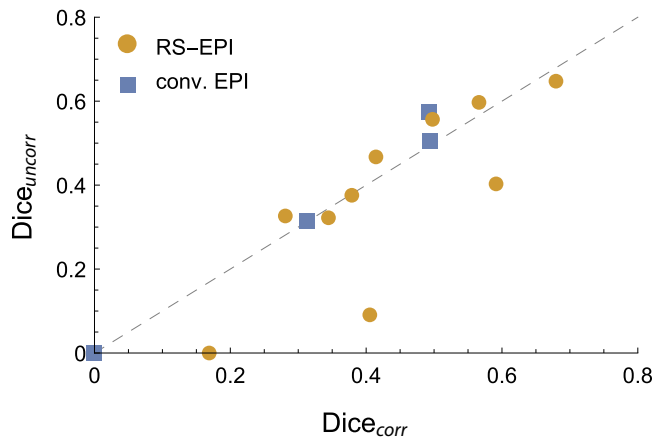


**Figure 3.** 3D visualization of the CNN segmentation with (A) and without (C) distortion correction. In addition, corresponding transverse slices of the region of interest are shown (B, D). The ground truth is shown in green, and the segmentation results are plotted in red. Both segmentations show good overlap with the gross tumor volume (GTV). With a Dice coefficient of 0.59, the overall segmentation of the geometrically corrected data was much higher than that of a Dice of 0.40 in the uncorrected case. However, both segmentations generally included too much tissue on the anterior side, as well as some isolated areas in the neck.



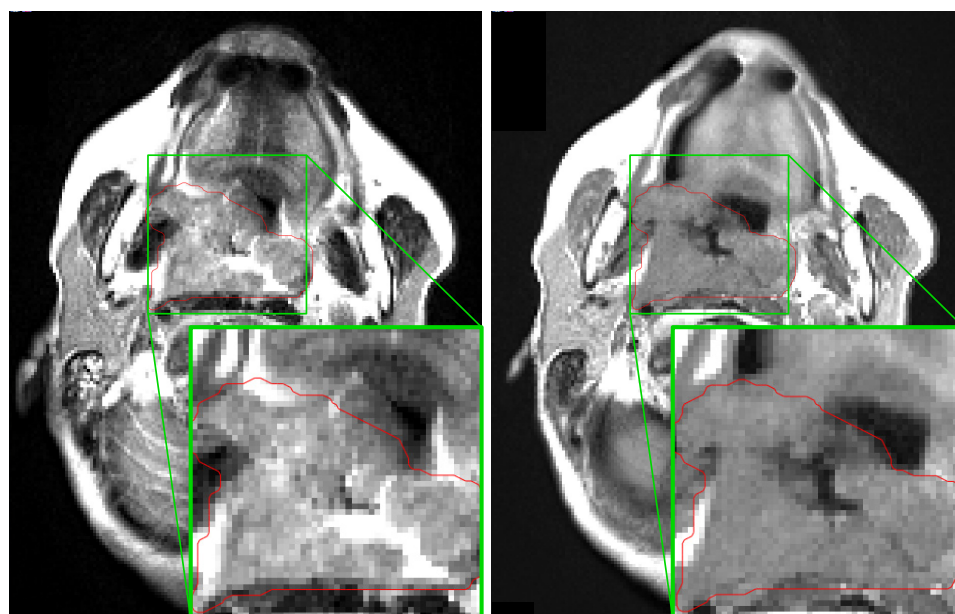
Still, the correction algorithm severely reduced image distortion. The approach is capable of registering different contrasts, such as T2w and DWI image data. Registration could not provide satisfactory results, whenever signals from multiple

voxels were mapped to the same location during the imaging process. Neither method, algorithm- or field map-based, could then recover the original, distortion-free image. This happened on a few sharp tissue-air boundaries and is therefore only a small limitation to the study.



**Figure 4.** Comparison of Dice coefficients with and without geometrically corrected input data for all 14 training rounds. The dashed line marks the line of identity. A paired *t* test on the data did not show a significant difference in Dice coefficient for corrected or uncorrected data. Mean Dice coefficient with distortion correction is  $0.40 \pm 0.18$ , and  $0.37 \pm 0.21$  without correction. Points below the line of identity indicate an improvement in segmentation performance for geometrically corrected ADC data. The 2 different DWI-sequences are shown in yellow and blue.

Owing to the limited number of complete patient data sets, a modified leave-1-out cross-validation method was chosen for statistical analysis. The method is limited by the incomplete number of possible permutations in training, validation, and testing set. A complete leave-1-out cross-validation could not be performed owing to high calculation times for each of the 42 840 possible combinations of the 3 sets. Therefore, 14 permutations with the given numbers of patients in training, validation, and testing categories have been used. Each permutation had a different data sample in the testing category, but the rest was randomly distributed among training and validation sets. This random selection was necessary owing to long calculation times required to completely train a network, taking several days on a Tesla C2075 GPU. To alleviate the challenge of small data sets, additional images after therapy starts could be used for training and testing. However, the tumors often drastically shrink in size, leading to changes in signal intensity for ADC and  $k_{trans}$  (36). Therefore, owing to vanishing tumors, the amount of available during-treatment data is too small for using deep learning techniques. This can already be seen in the present data set, which shows failure of segmentation in 2 of the cross-validation sets (Figure 4). These kinds of statistical fluctuations are to be expected more frequently with a smaller amount of available data. Thorough use of cross-validation must then be applied to extract statistically relevant information. However, there is a lower limit on the amount of data to be used with deep NNs, which can, in most cases for CNNs, be determined only experimentally.



**Figure 5.** T2w (left) and T1-weighted (T1w) (right) images showing the same anatomical area, but acquired 10 minutes after each other. Motion in the trachea leads to slightly differently located tumor borders. This effect introduces errors in the ground truth labels and decreases the maximum achievable segmentation performance.

From other tumor entities such as prostate or breast cancer it is known that DWI plays a vital part in tumor segmentation and definition (37–39), and similar behavior is found in head and neck cancers (40–42). In a preliminary study, we could also show that the overall segmentation performance of head and neck tumors in MRI is critically dependent on diffusion data (43). Therefore, it is surprising that the analysis of the segmentation performance of the CNN with and without distortion correction does not show significant differences. This could be caused by different reasons: In the training process the CNN could have learned a correction scheme to undistort input data within its receptive field. Because each layer consists of a number of convolutions with input data taken from the previous layer, local translation of features can be implemented. In addition, the standard deviation of the displacement map within the primary tumor over all included subjects is 2.29 pixel, while the standard deviation over all other pixels within all subjects is 4.28 pixel. This shows that distortions are far less pronounced within the tumor than in the rest of the FOV, especially in contrast to areas with tissue–air boundaries such as the nasal cavities where high distortions are to be expected in particular for EPI methods.

It is also important to note that the ADC maps constitute only 1 of 5 input channels. The high-resolution T2w images, for example, offer a much higher anatomical contrast and are nearly unaffected by distortion, whereas conventional DWI images can be heavily distorted. Hence, it is to be expected that feature maps linked to the ADC channel will show an effective decrease in feature resolution, while high-resolution information is taken from different input channels such as T2w data. In general, the quality of the ADC data in this study was limited by noise, which reduces the ability to differentiate between tumor and normal tissue. To increase the signal-to-noise ratio, DWI acquisitions can be averaged, which often increase the acquisition times to durations, which are no more compatible with clinical study times. Alternatively, during the ADC calculation,

noise can be explicitly modeled, which has been shown to reduce ADC heterogeneity (44, 45). In addition, the choice of the b-values of the DWI acquisition can be optimized, which requires prior knowledge about the target ADC values (46, 47).

In general, a strong limitation of this study is the size of the training data set. The small size of only 18 patients can lead to a false-positive segmentation far from the GTV owing to geometric distortions (as discussed above) and owing to the selection of the training regions: the algorithm was programmed such that in a statistical mean, the same number of tumor-containing (foreground) and nontumor-containing (background) input patches is selected, leading to an effective underrepresentation of background in the training process. A larger data set could help to train a CNN that can detect more subtle differences in the segmentation performance as the high standard deviation observed in the resulting Dice coefficients in the 14 data samples is expected to converge against a common mean value.

In many studies, data sets with >60 to >250 patients have been used (14, 34). Although other studies showed Dice coefficients in the range of 0.6 up to 0.9 (48), depending on the segmentation target (mostly brain tumors and subregions of brain tumors), our data set focused on a completely different tumor entity. This work offered special insight into the performance of a CNN in a body region with strong imaging challenges, as the head and neck regions show stronger field inhomogeneity than brain regions. Also, in contrast to most brain regions, the head and neck area cannot be assumed to be rigid. Although the head is immobilized using a thermoplastic mask, swallowing and tongue movement lead to intrinsic misalignment of images taken at different time points, as can be seen in Figure 5. Because the CNNs were trained on multiparametric data, some errors—especially on the GTV edges—are present in the ground-truth labels, leading to worse segmentation results than in rigid body areas. In addition, the interobserver variability of head and neck cancer is already much higher than, for example, that of the brain tumors (49, 50). Despite the intrinsic

limitation on image quality, the trained network yielded good tumor segmentations, and it was shown that distortion correction of ADC data does not significantly improve segmentation performance. To reduce the effect of motion-related misalignment, a nonlinear registration method could be applied. However, successful application of these methods is particularly demanding in the head and neck area and thus simultaneous signal acquisition, that is, intrinsic coregistration, would be preferred (51). Simultaneous acquisition of multiple signal parameters could be implemented by MR-fingerprinting, as has been shown in the prostate before (52).

In a next step, the contribution of each CNN input channel (eg, T2w or ADC images) to the segmentation performance needs to be quantified. This will not only allow for better analysis and understanding of the segmentation but also help optimize the imaging protocol with regard to increased patient comfort, that is, gathering more relevant information in less time, and treatment outcome.

## ACKNOWLEDGMENTS

We would like to thank Michael Mix and Arnd Sörensen for their support in the clinical study and many productive discussions.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The trial was registered retrospectively at the German Register for Clinical Studies (DRKS) under register number DRKS00003830 on August 20, 2015. Written informed consent was obtained from each patient, and the institutional review board approved the study (Approval No. 479/12).

## REFERENCES

- Emami B, Lyman J, Brown A, Cola L, Goitein M, Munzenrider JE, Shank B, Solin LJ, Wesson M. Tolerance of normal tissue to therapeutic irradiation. *Int J Radiat Oncol.* 1991;21:109–122.
- Thorwarth D. Biologically adapted radiation therapy. *Z Für Med Phys.* 2018;28:177–183.
- Peeken JC, Bernhofer M, Wiestler R, Goldberg T, Cremers D, Rost B, Wilkens JJ, Combs SE, Nüsslin F. Radiomics in radiooncology—challenging the medical physicist. *Phys Med.* 2018;48:27–36.
- Anderson CM, Sun W, Buatti JM, Maley JE, Policeni B, Mott SL, Bayouth JE. Interobserver and intermodality variability in GTV delineation on simulation CT, FDG-PET, and MR images of head and neck cancer. *J Radiat Oncol.* 2014;1:006.
- Doshi T, Wilson C, Paterson C, Lamb C, James A, MacKenzie K, Soraghan J, Petropoulakis L, Di Caterina G, Grose D. Validation of a magnetic resonance imaging-based auto-contouring software tool for gross tumour delineation in head and neck cancer radiotherapy planning. *Clin Oncol.* 2017;29:60–67.
- Prior FW, Fouke SJ, Benzinger T, Boyd A, Chicoine M, Cholleti S, Kelsey M, Keogh B, Kim L, Milchenko M, Polite DG, Tyree S, Weinberger K, Marcus D. Predicting a multi-parametric probability map of active tumor extent using random forests. In: *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).* 2013:6478–6481.
- Pinto A, Pereira S, Dinis H, Silva CA, Rasteiro DMLD. Random decision forests for automatic brain tumor segmentation on multi-modal MRI images. In: *2015 IEEE 4th Portuguese Meeting on Bioengineering (ENBENG).* 2015:1–5.
- Viswanath S, Bloch BN, Rosen M, Chappelow J, Toth R, Rofsky N, Lenkinski R, Genega E, Kalyanpur A, Madabhushi A. Integrating structural and functional imaging for computer assisted detection of prostate cancer on multi-protocol *in vivo* 3 Tesla MRI. *Proc SPIE—Int Soc Opt Eng.* 2009;7260:726031.
- Shah V, Turkbey B, Mani H, Pang Y, Pohida T, Merino MJ, Pinto PA, Choyke PL, Bernardo M. Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging. *Med Phys.* 2012;39:4093–4103.
- Guo D, Fridriksson J, Fillmore P, Rorden C, Yu H, Zheng K, Wang S. Automated lesion detection on MRI scans using combined unsupervised and supervised methods. *BMC Med Imaging.* 2015;15:50.

In summary our data showed that within the highly challenging anatomic head and neck region, even a CNN trained on nondistortion-corrected data can provide good-quality tumor segmentation. Considering the strong changes in the head and neck anatomy during radiochemotherapy, adaptive replanning strategies may help improve dose coverage of tumors and better sparing of organs at-risk (53, 54). This might ultimately result in better locoregional control rates and decreased treatment-related toxicities (55). The advent of MR-guided radiotherapy concepts, especially using hybrid MR-LINAC systems, facilitates daily MR-based replanning strategies that in turn require swift segmentation tools to allow real-time treatment adaptation (56). To deliver daily imaging-adapted treatment plans, CNN-enabled MR-based autosegmentation strategies are crucial. Our data therefore could provide important information about the design and implementation of CNNs for MR-based autosegmentation.

Informed consent to participate: Informed consent was obtained from all individual participants included in the trial. The institutional ethics review board approved the trial (Approval No. 479/12).

Competing interests: The authors have declared that no competing interest exists.

Funding: This work has been supported in parts by the Joint Funding Project “Joint Imaging Platform” of the German Cancer Consortium (DKTK).

- Gordillo N, Montseny E, Sobrevilla P. State of the art survey on MRI brain tumor segmentation. *Magn Reson Imaging.* 2013;31:1426–1438.
- Wilke M, de Haan B, Juenger H, Karnath H-O. Manual, semi-automated, and automated delineation of chronic brain lesions: a comparison of methods. *Neuroimage.* 2011;56:2038–2046.
- Dupont C, Betrouni N, Reyns N, Vermandel M. On image segmentation methods applied to glioblastoma: state of art and new trends. *IRBM.* 2016;37:131–143.
- Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lencz L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp C, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftikharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SMS, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Leemput KV. The multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans Med Imaging.* 2015;34:1993–2024.
- Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: state of the art and future directions. *J Digit Imaging.* 2017;30:449–459.
- Gkika E, Oehlke O, Bunea H, Wiedenmann N, Adebahr S, Nestle U, Zamboglou C, Kirste S, Fennell J, Brunner T, Gainey M, Baltas D, Langer M, Urbach H, Bock M, Meyer PT, Grosu A-L. Biological imaging for individualized therapy in radiation oncology: part II medical and clinical aspects. *Future Oncol Lond Engl.* 2018;14:751–769.
- Langer DL, van der Kwast TH, Evans AJ, Trachtenberg J, Wilson BC, Haider MA. Prostate cancer detection with multi-parametric MRI: logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. *J Magn Reson Imaging.* 2009;30:327–334.
- Zamboglou C, Drendel V, Jilg CA, Rischke HC, Beck TI, Schultze-Seemann W, Krauss T, Mix M, Schiller F, Wetterauer U, Werner M, Langer M, Bock M, Meyer PT, Grosu AL. Comparison of 68Ga-HBED-CC PSMA-PET/CT and multiparametric MRI for gross tumour volume detection in patients with primary prostate cancer



- based on slice by slice comparison with histopathology. *Theranostics*. 2017;7: 228–237.
19. Jezzard P, Balaban RS. Correction for geometric distortion in echo planar images from B0 field variations. *Magn Reson Med*. 1995;34:65–73.
  20. Porter DA, Heidemann RM. High resolution diffusion-weighted imaging using readout-segmented echo-planar imaging, parallel imaging and a two-dimensional navigator-based reacquisition. *Magn Reson Med*. 2009;62:468–475.
  21. Zhao M, Liu Z, Sha Y, Wang S, Ye X, Pan Y, Wang S. Readout-segmented echo-planar imaging in the evaluation of sinonasal lesions: a comprehensive comparison of image quality in single-shot echo-planar imaging. *Magn Reson Imaging*. 2016;34:166–172.
  22. Foltz WD, Porter DA, Simeonov A, Aleong A, Jaffray D, Chung P, Han K, Ménard C. Readout-segmented echo-planar diffusion-weighted imaging improves geometric performance for image-guided radiation therapy of pelvic tumors. *Radiother Oncol*. 2015;117:525–531.
  23. Hendrickson K, Phillips M, Smith W, Peterson L, Krohn K, Rajendran J. Hypoxia imaging with [F-18] FMISO-PET in head and neck cancer: potential for guiding intensity modulated radiation therapy in overcoming hypoxia-induced treatment resistance. *Radiother Oncol J Eur Soc Ther Radiol Oncol*. 2011;101: 369–375.
  24. Wiedenmann N, Bunea H, Rischke HC, Bunea A, Majerus L, Bielak L, Protopopov A, Ludwig U, Büchert M, Stoykow C, Nicolay NH, Weber WA, Mix M, Meyer PT, Hennig J, Bock M, Grosu AL. Effect of radiochemotherapy on T2\* MRI in HNSCC and its relation to FMISO PET derived hypoxia and FDG PET. *Radiat Oncol*. 2018;13:159.
  25. Bittner M-I, Wiedenmann N, Bucher S, Hentschel M, Mix M, Rücker G, Weber WA, Meyer PT, Werner M, Grosu A-L, Kayser G. Analysis of relation between hypoxia PET imaging and tissue-based biomarkers during head and neck radiochemotherapy. *Acta Oncol*. 2016;55:1299–1304.
  26. Tofts PS, Brix G, Buckley DL, Evelhoch JL, Henderson E, Knopp MV, Larsson HBW, Lee T-Y, Mayr NA, Parker GJM, Port RE, Taylor J, Weisskoff RM. Estimating kinetic parameters from dynamic contrast-enhanced t1-weighted MRI of a diffusable tracer: standardized quantities and symbols. *J Magn Reson Imaging*. 1999;10:223–232.
  27. Oudeman J, Coolen BF, Mazzoli V, Maas M, Verhamme C, Brink WM, Webb AG, Strijkers GJ, Nederveen AJ. Diffusion-prepared neurography of the brachial plexus with a large field-of-view at 3T. *J Magn Reson Imaging*. 2016;43:644–654.
  28. Vinegoni C, Lee S, Feruglio PF, Weissleder R. Advanced motion compensation methods for intravital optical microscopy. *IEEE J Sel Top Quantum Electron*. 2014;20:83–91.
  29. Pnevmatikakis EA, Giovannucci A. NoRMCorre: an online algorithm for piecewise rigid motion correction of calcium imaging data. *J Neurosci Methods*. 2017;291:83–94.
  30. Greenberg DS, Kerr JND. Automated correction of fast motion artifacts for two-photon imaging of awake animals. *J Neurosci Methods*. 2009;176:1–15.
  31. Atcheson B, Heidrich W, Ihrke I. An evaluation of optical flow algorithms for background oriented schlieren imaging. *Exp Fluids*. 2009;46:467–476.
  32. Lucas BD, Kanade T. An iterative image registration technique with an application to stereo vision. In: *Proceedings of the 1981 DARPA Image Understanding Workshop*. 1981:121–130.
  33. Bouguet J. Pyramidal implementation of the Lucas Kanade feature tracker. Intel Corp Microprocess Res Labs. 2000.
  34. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, Rueckert D, Glocker B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal*. 2017;36:61–78.
  35. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26:297–302.
  36. Wiedenmann N, Bunea H, Rischke H, Bunea A, Nicolay NH, Majerus L, Bielak L, Protopopov A, Ludwig U, Büchert M, Stoykow C, Mix M, Bock M, Grosu A. EP-2030 multiparametric MRI and FMISO PET in HNSCC and its relation with outcome. *Radiother Oncol*. 2019;133:S1114–S1115.
  37. Steiger P, Thoeny HC. Prostate MRI based on PI-RADS version 2: how we review and report. *Cancer Imaging*. 2016;16:9.
  38. Yabuuchi H, Matsuo Y, Okafuji T, Kamitani T, Soeda H, Setoguchi T, Sakai S, Hatakenaka M, Kubo M, Sadanaga N, Yamamoto H, Honda H. Enhanced mass on contrast-enhanced breast MR imaging: Lesion characterization using combination of dynamic contrast-enhanced and diffusion-weighted MR images. *J Magn Reson Imaging*. 2008;28:1157–1165.
  39. Burnside ES, Sickles EA, Bassett LW, Rubin DL, Lee CH, Ikeda DM, Mendelson EB, Wilcox PA, Butler PF, D'Orsi CJ. The ACR BI-RADS® Experience: learning from history. *J Am Coll Radiol*. 2009;6:851–860.
  40. Driessen JP, van Kempen PMW, van der Heijden GJ, Philippens MEP, Pameijer FA, Stegeman I, Terhaard CHJ, Janssen LM, Grolman W. Diffusion-weighted imaging in head and neck squamous cell carcinomas: a systematic review. *Head Neck*. 2015;37:440–448.
  41. Connolly M, Srinivasan A. Diffusion-weighted imaging in head and neck cancer: technique, limitations, and applications. *Magn Reson Imaging Clin N Am*. 2018; 26:121–133.
  42. Popp I, Bott S, Mix M, Oehlke O, Schimek-Jasch T, Nieder C, Nestle U, Bock M, Yuh WTC, Meyer PT, Weber WA, Urbach H, Mader I, Grosu A-L. Diffusion-weighted MRI and ADC versus FET-PET and GdT1w-MRI for gross tumor volume (GTV) delineation in re-irradiation of recurrent glioblastoma. *Radiother Oncol*. 2019;130:121–131.
  43. Bielak L, Wiedenmann N, Lottner T, Bunea H, Grosu A-L, Bock M. Quantifying information content of multiparametric MRI data for automatic tumor segmentation using CNNs. In: *Proceedings of the International Society for Magnetic Resonance in Medicine*. Montréal, QC, Canada; 2019:2339.
  44. Jha AK, Rodríguez JJ, Stoeckel AT. A maximum-likelihood method to estimate a single ADC value of lesions using diffusion MRI. *Magn Reson Med*. 2016;76: 1919–1931.
  45. Walker-Samuel S, Orton M, McPhail LD, Robinson SP. Robust estimation of the apparent diffusion coefficient (ADC) in heterogeneous solid tumors. *Magn Reson Med*. 2009;62:420–429.
  46. Saritas EU, Lee JH, Nishimura DG. SNR dependence of optimal parameters for apparent diffusion coefficient measurements. *IEEE Trans Med Imaging*. 2011;30: 424–437.
  47. Bielak L, Bock M. Optimization of diffusion imaging for multiple target regions using maximum likelihood estimation. *Curr Dir Biomed Eng*. 2017;3:203–206.
  48. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin P-M, Larochelle H. Brain tumor segmentation with Deep Neural Networks. *Med Image Anal*. 2017;35:18–31.
  49. Chaput A, Robin P, Podeur F, Ollivier M, Keromnes N, Tissot V, Nonent M, Salaün P-Y, Rousset J, Abgral R. Diagnostic performance of 18fluorodesoxyglucose positron emission/computed tomography and magnetic resonance imaging in detecting T1–T2 head and neck squamous cell carcinoma. *Laryngoscope*. 2018;128:378–385.
  50. Visser M, Müller DMJ, van Duijn RJM, Smits M, Verburg N, Hendriks EJ, Nabuurs RJA, Bot JCJ, Eijgelaar RS, Witte M, van Herk MB, Barkhof F, de Witt Hamer PC, de Munck JC. Inter-rater agreement in glioma segmentations on longitudinal MRI. *Neuroimage Clin*. 2019;22:101727.
  51. Monti S, Cavaliere C, Covello M, Nicolai E, Salvatore M, Aiello M. An evaluation of the benefits of simultaneous acquisition on pet/mr coregistration in head/neck imaging. *J Healthc Eng*. 2017;2017:2634389.
  52. Yu AC, Badve C, Ponsky LE, Pahwa S, Dastmalchian S, Rogers M, Jiang Y, Margevicius S, Schluchter M, Tabayooyong W, Abouassaly R, McGivney D, Griswold MA, Gulani V. Development of a combined MR fingerprinting and diffusion examination for prostate cancer. *Radiology*. 2017;283:729–738.
  53. Surucu M, Shah KK, Roeske JC, Choi M, Small W, Emami B. Adaptive radiotherapy for head and neck cancer. *Technol Cancer Res Treat*. 2017;16:218–223.
  54. Castelli J, Simon A, Lafond C, Perichon N, Rigaud B, Chajon E, Bari BD, Ozsahin M, Bourhis J, de Crevoisier R. Adaptive radiotherapy for head and neck cancer. *Acta Oncol*. 2018;57:1284–1292.
  55. Kataria T, Gupta D, Goyal S, Bisht SS, Basu T, Abhishek A, Narang K, Banerjee S, Nasreen S, Sambasivam S, Dhyani A. Clinical outcomes of adaptive radiotherapy in head and neck cancers. *Br J Radiol*. 2016;89:20160085.
  56. Chuter RW, Pollitt A, Whitehurst P, MacKay RI, van Herk M, McWilliam A. Assessing M. R-linear radiotherapy robustness for anatomical changes in head and neck cancer. *Phys Med Biol*. 2018;63:125020.