



## OPEN ACCESS

## EDITED BY

Dousatsu Sakata,  
Osaka University, Japan

## REVIEWED BY

Alina Santiago,  
University Hospital Essen, Germany  
Akihiro Haga,  
Tokushima University, Japan

## \*CORRESPONDENCE

Alexander Ratke,  
✉ alexander.ratke@udo.edu

RECEIVED 11 September 2023

ACCEPTED 21 November 2023

PUBLISHED 05 December 2023

## CITATION

Ratke A, Darsh E, Heinzelmann F,  
Kröniger K, Timmermann B and  
Bäumer C (2023), Deep-learning-based  
deformable image registration of head CT  
and MRI scans.

*Front. Phys.* 11:1292437.

doi: 10.3389/fphy.2023.1292437

## COPYRIGHT

© 2023 Ratke, Darsh, Heinzelmann,  
Kröniger, Timmermann and Bäumer.  
This is an open-access article distributed  
under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#).  
The use, distribution or reproduction in  
other forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Deep-learning-based deformable image registration of head CT and MRI scans

Alexander Ratke<sup>1\*</sup>, Elena Darsh<sup>1</sup>, Feline Heinzelmann<sup>1,2,3,4</sup>,  
Kevin Kröniger<sup>1</sup>, Beate Timmermann<sup>2,3,4,5,6</sup> and  
Christian Bäumer<sup>1,2,3,5,6</sup>

<sup>1</sup>TU Dortmund University, Department of Physics, Dortmund, Germany, <sup>2</sup>West German Proton Therapy Centre Essen, Essen, Germany, <sup>3</sup>West German Cancer Center, Essen, Germany, <sup>4</sup>University Hospital Essen, Department of Particle Therapy, Essen, Germany, <sup>5</sup>University Hospital Essen, Essen, Germany, <sup>6</sup>German Cancer Consortium, Essen, Germany

This work is motivated by the lack of publications on the direct application of multimodal image registration with deep-learning techniques for the enhancement of treatment planning in particle therapy. An unsupervised workflow, which seeks to improve image alignment, was developed and evaluated for computed tomography and magnetic resonance imaging scans of the head. The scans of 39 paediatric patients with brain tumours were available. The focus of the two-step workflow, including preprocessing of the scans for normalisation, is deformable image registration (DIR) with a deep neural network, which generates deformation vector fields (DVs). To obtain a suitable configuration of the network, parameter tuning is performed by varying its parameters, e.g., layer size, regularisation ( $\lambda$ ) of the DVF and learning rate ( $\alpha$ ). Image similarity was determined with the Dice similarity coefficient,  $m_{DSC}$ , using segmented images and the mutual-information metric,  $m_{MI}$ . The performance of the deep-learning models was assessed with the inverse consistency,  $m_{IC}$ , and the Jacobian determinant,  $m_{JD}$ . Inverse consistency is obtained for  $m_{IC} = 0$  mm, while the determinant of a deformed image is expected to be unity. The deep-learning models passed both performance checks, indicated by the mean values  $\bar{m}_{IC} = (0.57 \pm 1.00)$  mm and  $\bar{m}_{JD} = (1.00 \pm 0.07)$ . Models with  $\lambda \geq 1$  yielded higher  $m_{DSC}$  values than models with lower  $\lambda$  values. A small-architecture model with  $\alpha = 10^{-4}$  was found to be most suitable for DIR, as improvement in image similarity of up to 12% was obtained in terms of  $m_{MI}$ . The direct application of deep-learning models produced registered images improving image alignment between scans of different modalities.

## KEYWORDS

image registration, multimodal, deep learning, deformable transformation, unsupervised

## 1 Introduction

In radiotherapy, medical imaging techniques like computed tomography (CT), magnetic resonance imaging (MRI) or positron emission tomography are used for treatment planning to obtain images of the patient's anatomy [1]. Each modality provides unique image contrast of the tissue, e.g., high soft-tissue distinctness in MRI scans or high bone contrast in CT scans [1]. As the treatment planning includes the contouring of target volumes and healthy tissue, information on the enhanced soft-tissue contrast in MRI scans can improve the quality of radiotherapy by a more precise delineation with images of multiple modalities [1–3].

Treatment-planning systems used in particle therapy, e.g., proton therapy, usually perform a rigid registration to superimpose scans of different modalities [4]. However, the rigid registration does not include the displacement of organs due to patient immobilisation or MRI scans distorted by magnetic fields [1, 4]. Furthermore, physical changes of the patient such as body growth or tumour response are challenging, especially for paediatric patients. Deformable image registration (DIR) with individual voxel displacements can address these discrepancies [3, 4]. Since the integration of MRI into particle-therapy techniques [5, 6] continues to progress, the importance of multimodal DIR is growing as well. In the case of paediatric patients with small geometric scales, the geometrical accuracy is required for the irradiation with steep dose gradients in particle therapy. A current practical limitation is that advanced image processing techniques such as DIR [7, 8], generation of synthetic CT scans [9–11], and automated contouring [12] have been established for adults and, thus, do not cover the wider anatomical range of paediatric cases. This hampers the efficiency gain in treatment planning for the patient group that benefits most from proton therapy.

The research on deep-learning-based medical image analysis has increased in recent years [3]. A variety of publications presented concepts of unsupervised DIR for unimodal use [13–15]. Convolutional neural networks (CNNs), for example, extract the features of images with convolution operations for image deformation or recognition [3, 7]. The U-Net [16] structure, which provides efficient learning of image features on a small-sized data set [7], is often used in combination with CNN. Balakrishnan et al. took advantage of this for fast atlas-based DIR of brain MRI scans [17]. Furthermore, reviews about medical image registration stated that up to 70% of the publications focused on unimodal image registration [7, 8]. Various studies have explored direct multimodal DIR for different regions of interest and modalities. Some studies focused on the registration of MRI and ultrasound scans [18], as well as the registration of CT and MRI scans of the abdomen [18, 19]. In contrast, other research efforts employed indirect deep-learning-based techniques by synthesising CT scans from corresponding MRI scans as a preliminary step, thereby making use of unimodal registration techniques [9]. Nevertheless, an increasing number of recent publications indicate a growing trend towards direct DIR for clinical applications [7, 8].

The investigations of this study contribute to the research field by developing an unsupervised image-registration workflow for head CT and MRI scans using direct deep-learning-based DIR with a CNN. First, the data sets as well as the properties of the scans are described in Section 2.1. After explaining the choice of image-similarity metrics in Section 2.2, the preprocessing, which is required for the deep-learning step, is introduced in Section 2.3. Then, the deep-learning network (Section 2.4) and its parameter tuning (Section 2.5) are presented. The results are rigidly (Section 3.1) and deformably (Section 3.2) registered images generated by the fast and direct application of the registration workflow.

## 2 Material and methods

### 2.1 Data

The data provided by the West German Proton Therapy Centre Essen (WPE) consisted of 39 paired CT and MRI scans of patients with a maximum age of 18. Each pair of images had been acquired to

initially plan the treatment of the patients with brain tumours. For the parameter tuning of the deep-learning-based DIR, the data were subdivided into a training and a testing data set. The former contained the scans of 25 patients. The clinical protocol requires a planning MRI if the most recent diagnostic MRI is older than 30 days or if anatomical changes were likely in this period. For the 25 patients, who were part of a cohort of the KiProReg register study (DRKS00005363) [20], the acquisition of the planning scans was performed subsequent to the respective X-ray CT scan. In addition, the testing data set with 14 paediatric patients was used to assess the workflow with the most suitable parameter setting. The MRI scans of the testing data set had been acquired directly after the CT scans. These patients were part of the KiAPT study [21], which is linked to KiProReg. Both the KiProReg (18-9109-BO) and the KiAPT (18-8320-BO) studies were approved by the Ethics Committee of the University of Duisburg-Essen.

The CT scans were composed of slices containing  $512 \times 512$  pixels. For MRI, the  $T_2$ -weighted scans were used in this study. The aspect ratio of these scans was not constant and varied from  $230 \times 256$  to  $512 \times 600$ . The numbers of CT and MRI slices were in the ranges of 245 to 369 and 33 to 108, respectively. The latter numbers are much smaller than the former ones because of the larger slice thickness of MRI. The MRI scans had been conducted without patient immobilisation. The head width of the patients varied between 120 and 161 mm. Furthermore, the training data set included the contour of the ventricular system for each patient and modality, which had been outlined by a medical physicist. Afterwards, all contours were validated by a senior clinician at WPE.

### 2.2 Image-similarity metrics and evaluation techniques

A major challenge of multimodal approaches in DIR is the variation of the intensity distributions associated with different tissue types. The metrics have to be chosen with the intention of measuring the alignment of image pairs [4]. Two intensity-based metrics assessing image similarity are the normalised cross-correlation [22],  $m_{NCC}$ , and the mutual-information metric [23],  $m_{MI}$ . Another possibility to determine the image alignment is the Dice similarity coefficient [24],  $m_{DSC}$ . This feature-based metric measures the overlap of segmented images. The range of  $m_{NCC}$  and  $m_{DSC}$  is between 0 and 1, where higher and lower values refer to agreement and disagreement, respectively. Moreover, higher  $m_{MI}$  values indicate an increase in image similarity.

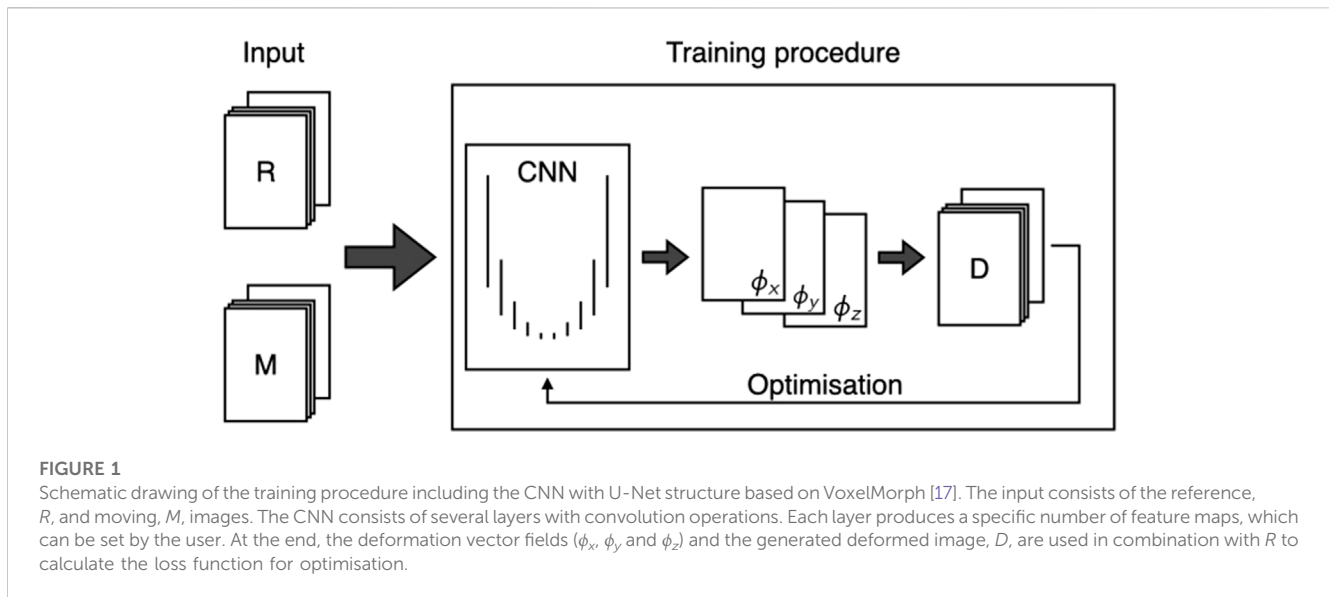
Two approaches were chosen to assess the registration accuracy by comparing the results of deformable and rigid registration. Segmented images were used to calculate the difference in the overlap between deformably and rigidly registered images,

$$\Delta m_{DSC} = m_{DSC,DIR} - m_{DSC,rigid}. \quad (1)$$

The relative deviation,

$$\Delta m_{MI} = (m_{MI,DIR} - m_{MI,rigid}) / m_{MI,rigid}, \quad (2)$$

of the mutual-information metric was additionally calculated. In both cases, positive values indicate an improvement of the deformably registered images with respect to the preprocessed images.



Furthermore, two checks of the results of deep-learning-based DIR were carried out to obtain information on the registration performance [4]. The inverse-consistency method was performed by adding the DVFs of both the MRI-to-CT and CT-to-MRI directions. The sum,  $m_{IC}$ , is expected to be consistent with zero. Another check quantified the change of the voxel volume after registration by calculating the Jacobian determinant,  $m_{JD}$ , for each voxel of the deformation vector field (DVF). The mean value for the deformed image should be unity due to the diffeomorphic transformation.

### 2.3 Preprocessing

A neural-network-based registration requires preprocessing of the input images. The procedure developed in this study runs automatically on the data sets to get rigidly registered images with the same image format. All image-processing steps were devised with the CT scans as the reference images,  $R$ , and the MRI scans as the moving images,  $M$ .

In a first step, the images, which contained many slices without anatomical information and in some cases parts of the shoulder, were removed to preserve memory for further computations. Then, image segmentation was performed to enable computing  $m_{DSC}$  for an image pair. For this, segments of the eyes were derived from intensity restrictions using the Hounsfield scale and thresholding techniques for CT and MRI, respectively. For the training data set, the outlined ventricular system was additionally converted from contour to volume, complementing the segmented image.

The main part of the preprocessing is the reformatting of the images. The adaptation of the image formats is essential for input pairs of a neural network. This includes the pixel spacing, which differs between modalities. Therefore, the scans were scaled to a pixel spacing and a slice thickness of 1 mm, facilitating an equivalent representation in the axial plane. After scaling, the images were cut to an aspect ratio of 3 : 4 with  $192 \times 256$  pixels since this format suits the shape and the size of the head. Then, the spatial discrepancies between CT and MRI scans were reduced by rigid registration. For

the transformation, three translation parameters and two rotation angles were calculated with the centres of mass of the eye segments. The third angle was determined with an iterative procedure of applying rotations in steps of  $1^\circ$  until the highest  $m_{DSC}$  value was measured. Eventually, the number of slices was decreased to 64 for all images.

For the deep-learning training process, the pixel values were manipulated. The thresholds 1024 (0 HU) and 600 were chosen for the CT and MRI scans, respectively. These cuts reduced noise coming from low intensities by clipping lower values to 0. Furthermore, intensities of CT scans higher than 300 HU were subtracted by 300 to maintain the morphology and to soften the highest intensity difference between CT and MRI coming from bone tissue. Also, the intensity ranges were standardised to 256 greyscale values.

### 2.4 Deep learning

This study employed a CNN for image registration of CT and MRI scans. The training procedure is presented in Figure 1. The input of the network consists of at least one pair of preprocessed images to produce deformed images,  $D$ , with deformable registration. If the training data set contains the image pairs of several patients, the input can be grouped into batches,  $b$ , by concatenating various reference and moving images. This increases the possibility to generalise the model [25].

The network is formed according to U-Net [16], signifying an encoder and decoder path for destruction and reconstruction, respectively. Both paths are subdivided into several levels. In the encoder path, each level consists of a specified number of convolutions to gain feature maps depending on the kernel weights. An activation function is then applied to the feature maps. Subsequently, the level is terminated with a pooling operation to reduce the size of the feature maps. Traversing the encoder path yields the structures from the images, but the information of their position is lost. The decoder path is constructed to connect all feature maps in the respective level.

**TABLE 1** Variations of the CNN architecture regarding the number of feature maps in each layer of the encoder and decoder paths.

Name	Encoder	Decoder
Small	[8 – 16 – 16 – 16]	[16 – 16 – 16 – 16 – 16 – 8 – 8]
Medium	[16 – 32 – 32 – 32]	[32 – 32 – 32 – 32 – 32 – 16 – 16]
Large	[24 – 48 – 48 – 48]	[48 – 48 – 48 – 48 – 48 – 24 – 24]

Hence, the feature maps at the end of the encoder path are unfolded to restore the size and to apply convolutions afterwards. Finally, a deformation vector field,  $\phi$ , is generated to predict the spatial transformation of the moving image. To finish one iteration of the training, a loss function is calculated with the aim of minimisation, which necessitates a large number of epochs for the described process to learn and improve the convolution weights from previous iterations.

The CNN, based on VoxelMorph [17, 26], used a kernel size of  $3 \times 3 \times 3$  for the convolution operations and the leaky rectified linear unit [27] with a small gradient of 0.2 as activation function. For pooling operations in the encoder path, the size of the feature maps was halved by passing the maximum value of a  $2 \times 2 \times 2$  grid. A grid with the same size was used to double the format of the feature maps in the decoder path. The computation of  $D$  was performed by exploiting the information of the DVF and applying linear interpolation. The nearest-neighbour method was used for segmented images to maintain the pixel values of the segments. The loss function was defined as

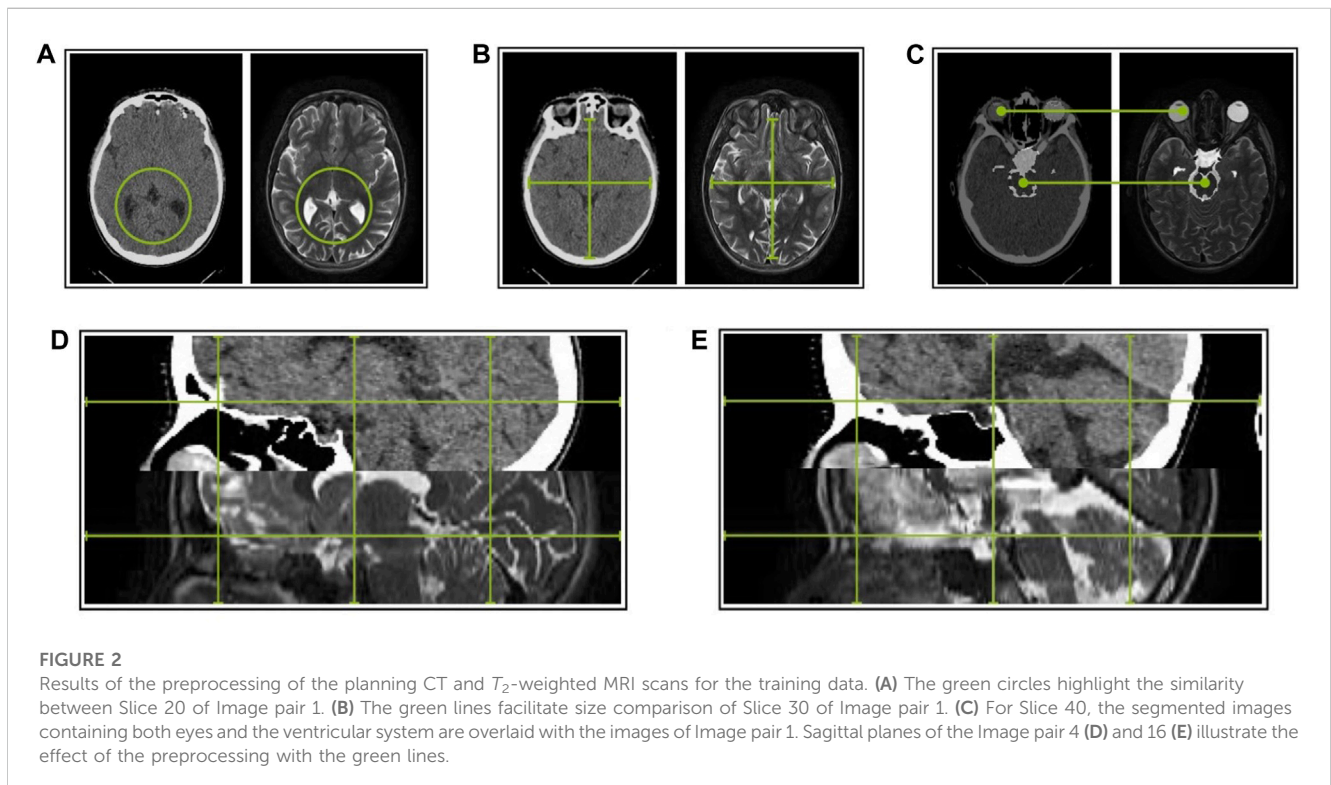
$$\mathcal{L}(R, D, \phi) = \mathcal{L}_{\text{sim}}(R, D) + \lambda \mathcal{L}_{\text{grd}}(\phi). \quad (3)$$

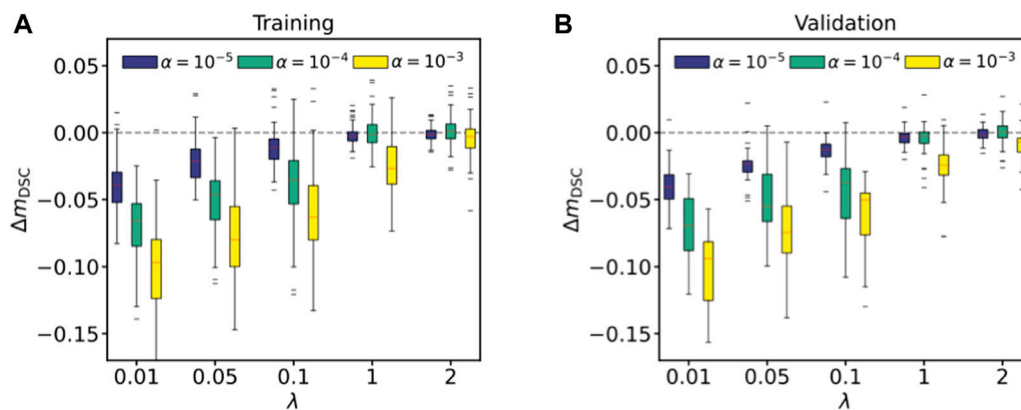
Here, the term  $\mathcal{L}_{\text{sim}}(R, D) = (1 - m_{\text{NCC}})$  was chosen to represent similarity and dissimilarity between the images for values near zero and unity, respectively. The term  $\mathcal{L}_{\text{grd}}(\phi)$ , calculated from the spatial differences of adjacent voxels in the DVF, was scaled with the parameter  $\lambda$  for regularisation [17]. The Adam optimiser [28] was implemented in this deep-learning network with a customisable learning rate,  $\alpha$ .

## 2.5 Parameter tuning

In this study, several models were trained and validated to find an appropriate set of the CNN parameters. For this purpose, the training data set was subdivided to apply a five-fold cross-validation with 80% of the data set for training and 20% for validation. The training of each model ran 200 iterations with a non-randomised training sample and started with the same initial weights to assess the impact of the parameters. The parameters of the most suitable setting were chosen to train a model with the whole training data set and to apply it to the testing data set afterwards.

Regarding the optimisation of the CNN, models were trained independently by varying parameters related to the loss function, the optimiser function, the CNN structure and the batch size. The regularisation parameter in Eq. 3 and the learning rate of the optimiser were part of the model variations, where the settings  $\lambda = \{0.01, 0.05, 0.1, 1, 2\}$  and  $\alpha = \{10^{-5}, 10^{-4}, 10^{-3}\}$  were investigated. The range of  $\lambda$ -values were chosen, since similar variations had been





**FIGURE 3**

Registration accuracy for the small-architecture network and the batch size  $b = 1$  with the Dice similarity coefficient. The difference of  $m_{DSC}$  before and after applying DIR to the segmented images was determined. The results are shown for the training (A) and validation (B) data. Each plot contains the variations of  $\lambda$  and  $\alpha$ , regulating the smoothness of the deformations and the step size of the optimiser, respectively. The red lines inside the boxes represent the medians.

used in unimodal registrations [17]. Moreover, the size of the network architecture was varied three times by means of the number of resulting feature maps in the levels of the encoder and decoder paths. These variants are listed in Table 1. Models with a batch size of  $b = 4$ , which splits the training data set into groups of four image pairs, were trained besides models with  $b = 1$ .

## 3 Results

### 3.1 Preprocessed data

In Figure 2, images with two patient orientations exemplify the outcome of the preprocessing. The unified image format of  $192 \times 256$  pixels is shown for selected slices of one patient in Figures 2A–C, where equivalent structures are located in the same CT and MRI slice with similar positions. The image alignment is also corroborated with the segmented images, which display the eyes and the ventricular system. The MRI slices appear magnified in transversal directions compared to the CT scans. In addition, the selection of the 64 slices regarding the head area can be inferred from the sagittal planes of two preprocessed image pairs in Figures 2D, E.

### 3.2 Deformable registration with deep learning

The parameter tuning, described in Section 2.5, led to 90 registration-model configurations. For each configuration, the results of the cross-validation method were combined in the evaluation process. In the following, the investigation of registration accuracy, performance and qualitative comparison of the registered images, aiming at finding an appropriate parameter setting, is presented.

#### 3.2.1 Accuracy

The evaluation of the registration accuracy was performed with a feature-based metric and an intensity-based metric (Section 2.2).

**TABLE 2** Difference of the Dice similarity coefficients,  $\Delta m_{DSC}$ , for the parameter tuning with the small-architecture model and batch size  $b = 1$ .

$\alpha$	$\lambda = 0.01$	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 2$
Training					
$10^{-5}$	$-0.04 \pm 0.02$	$-0.02 \pm 0.02$	$-0.01 \pm 0.01$	$0.00 \pm 0.01$	$0.00 \pm 0.01$
$10^{-4}$	$-0.07 \pm 0.02$	$-0.05 \pm 0.02$	$-0.04 \pm 0.03$	$0.00 \pm 0.01$	$0.00 \pm 0.01$
$10^{-3}$	$-0.10 \pm 0.03$	$-0.08 \pm 0.03$	$-0.06 \pm 0.03$	$-0.02 \pm 0.02$	$0.00 \pm 0.01$
Validation					
$10^{-5}$	$-0.04 \pm 0.02$	$-0.02 \pm 0.01$	$-0.01 \pm 0.01$	$0.00 \pm 0.01$	$0.00 \pm 0.01$
$10^{-4}$	$-0.07 \pm 0.02$	$-0.05 \pm 0.03$	$-0.04 \pm 0.03$	$-0.01 \pm 0.01$	$0.00 \pm 0.01$
$10^{-3}$	$-0.10 \pm 0.03$	$-0.07 \pm 0.03$	$-0.06 \pm 0.03$	$-0.03 \pm 0.02$	$0.00 \pm 0.01$

Both metrics were determined and compared to the rigidly registered images to highlight improvements or deteriorations achieved by the deep-learning-based DIR.

The  $\Delta m_{DSC}$  values of the small-architecture models with the batch size  $b = 1$  are shown in Figure 3 and listed in Table 2. The distributions of the other configurations showed similar tendencies. In general, a decreasing trend of the accuracy was visible for lower values of the regularisation parameter. Models yielded better accuracy for  $\lambda \geq 1$ . The variation of the learning rate showed that models with  $\alpha = 10^{-3}$  mostly achieved lower accuracy than models with  $\alpha = 10^{-4}$  or  $\alpha = 10^{-5}$ . A comparison between the results of the training data in Figure 3A and the validation data in Figure 3B illustrated the same tendencies, e.g., low values of  $\lambda$  led to a decrease in accuracy.

As the  $\Delta m_{DSC}$  values indicated configurations with higher registration accuracy based on the segmented images, the mutual information is additionally used to assess the image similarity based on statistical information of the images. The results are listed in Table 3. The same tendencies observed with  $\Delta m_{DSC}$  were obtained with mutual information. Consequently, the most suitable CNN

**TABLE 3** Relative deviation,  $\Delta m_{MI}$ , of the mutual-information metric for the parameter tuning with the small-architecture model and batch size  $b = 1$ .

$\alpha$	$\lambda = 0.01$	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 2$
Training					
$10^{-5}$	$-0.09 \pm 0.02$	$-0.05 \pm 0.01$	$-0.01 \pm 0.03$	$0.06 \pm 0.04$	$0.08 \pm 0.05$
$10^{-4}$	$-0.12 \pm 0.02$	$-0.05 \pm 0.02$	$0.01 \pm 0.03$	$0.09 \pm 0.04$	$0.12 \pm 0.05$
$10^{-3}$	$-0.08 \pm 0.02$	$-0.05 \pm 0.02$	$-0.03 \pm 0.02$	$0.03 \pm 0.02$	$0.04 \pm 0.02$
Validation					
$10^{-5}$	$-0.09 \pm 0.02$	$-0.05 \pm 0.01$	$-0.01 \pm 0.02$	$0.06 \pm 0.03$	$0.07 \pm 0.03$
$10^{-4}$	$-0.10 \pm 0.02$	$-0.05 \pm 0.01$	$0.01 \pm 0.02$	$0.07 \pm 0.03$	$0.10 \pm 0.04$
$10^{-3}$	$-0.08 \pm 0.02$	$-0.06 \pm 0.01$	$-0.04 \pm 0.02$	$0.03 \pm 0.02$	$0.03 \pm 0.02$

configuration is the small-architecture model with the parameters  $b = 1$ ,  $\lambda = 2$  and  $\alpha = 10^{-4}$ , achieving the highest improvement in image similarity according to both metrics. The average increase of 12% was yielded with  $\Delta m_{MI}$ .

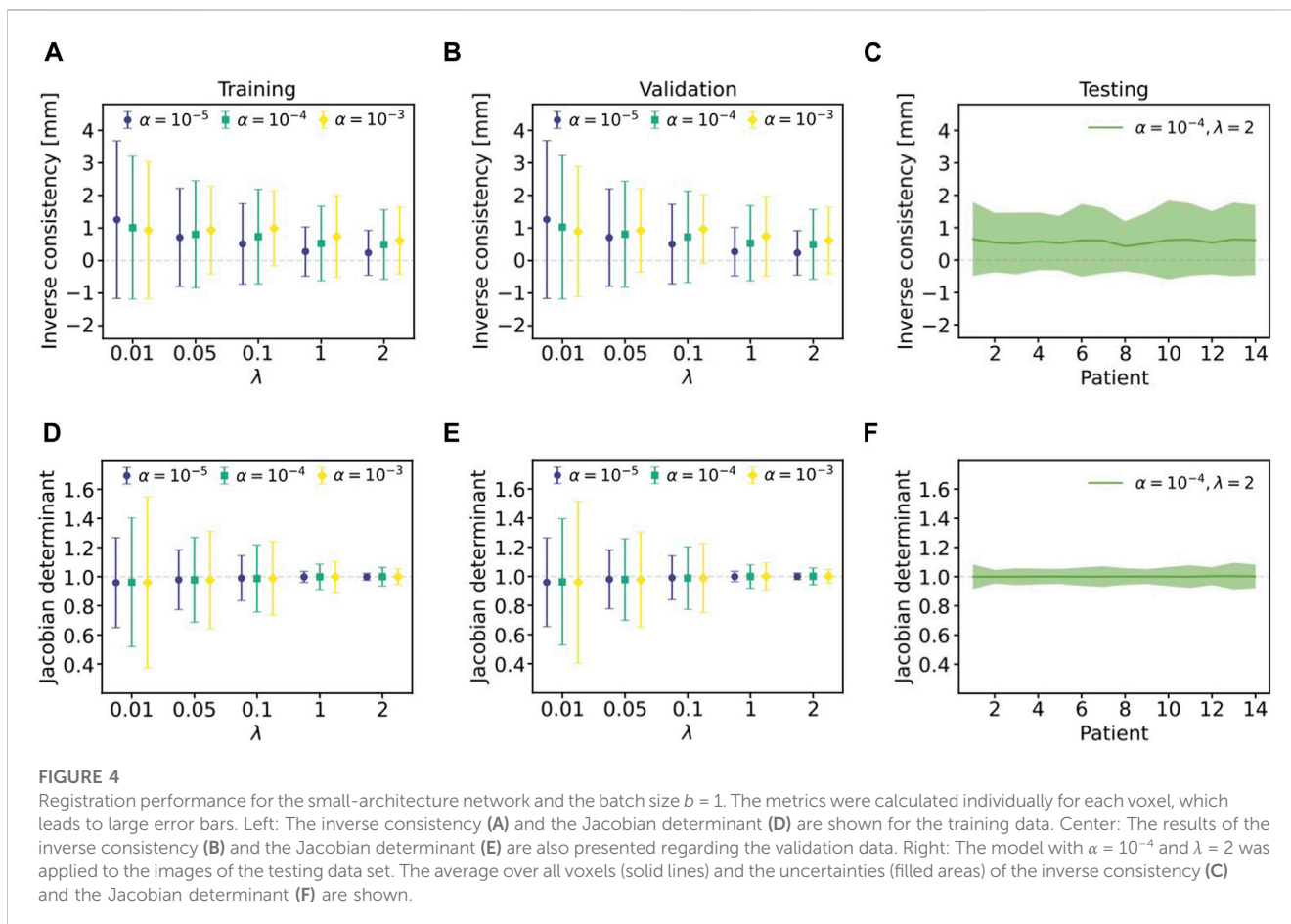
### 3.2.2 Performance

The inverse consistency and the Jacobian determinant were determined individually for each voxel, which led to a high scatter.

Since the results were similar between the variations of the network architectures and the batch sizes, the outcome of the parameter tuning is illustrated in Figure 4 for the small-architecture network and  $b = 1$ . The mean values for each parameter setting were calculated with the 100 and the 20 image pairs of the five-fold training and validation data, respectively. In addition, the model performance of the most suitable configuration was measured with the testing data set.

The inverse consistency indicated that the results of the validation data (Figure 4B) and the testing data (Figure 4C) were compatible with the training data (Figure 4A). Interestingly, a slight shift towards positive values was measured for all parameter settings. This means that the deformations differed regarding the registration direction. However, the results were still consistent with zero, as evidenced by  $\bar{m}_{IC} = (0.57 \pm 1.00)$  mm for the testing data set. The evaluation with the Jacobian determinant pointed out that all parameter variations yielded determinants close to unity, which was indicated by  $\bar{m}_{JD} = 1.00 \pm 0.07$  for the testing data set. The deviation of the Jacobian determinant was larger for low- $\lambda$  models, which hints at unstable registration performance.

Moreover, the values of the Jacobian determinant of one slice are presented in Figure 5A. The strength of the deformations increased towards the outer region of the head. In Figure 5B, the  $|1 - \bar{m}_{JD}|$  distribution was calculated for the radius  $r$ , which started at the isocentre. This indicated deformations of up to 10% on average at the outer region of the head.



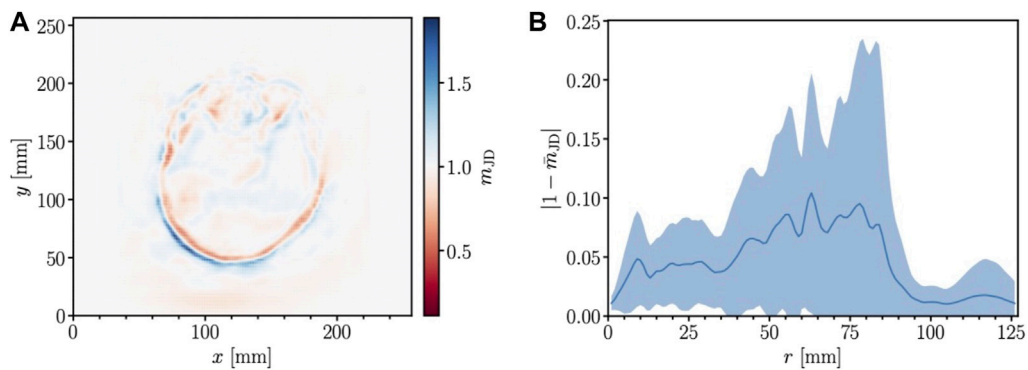


FIGURE 5

Volume changes of one slice of the testing data for the most suitable parameter setting. (A) The Jacobian determinant is shown for each voxel. While  $m_{JD} = 1$  represents no volume change, values above or below unity indicate volume expansion or contraction, respectively. (B) The term  $|1 - \bar{m}_{JD}|$  (blue line) depending on the radius  $r$  is determined from the isocentre. The uncertainties are illustrated as the blue area.

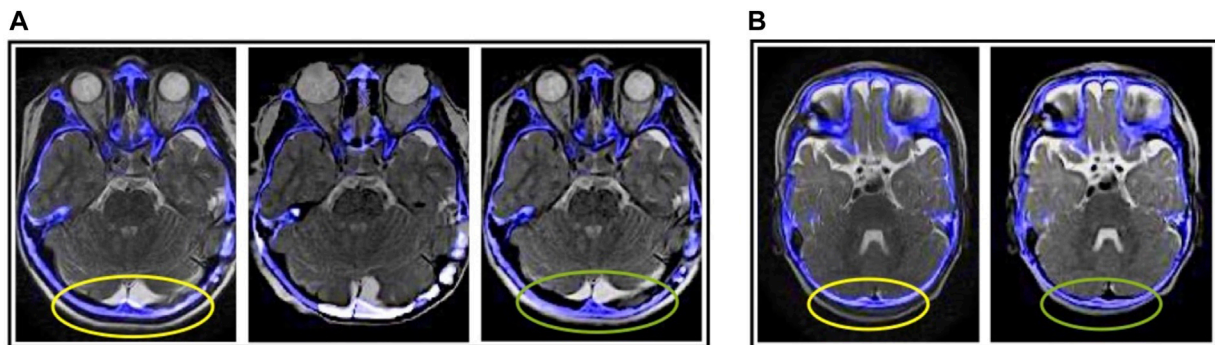


FIGURE 6

Visualisation of overlaid planning CT (blue colours) and deformed  $T_2$ -weighted MRI (greyscale) slices. The blue structures represent the skull in the CT slice. The ellipses indicate distinct differences from low (yellow) to high (green) image alignment. (A) The overlay of the preprocessed CT and MRI slices (left) as well as the results after the DIR (center, right) are depicted for the training data set. Besides the outcome of the most suitable parameter setting (right), MRI slices registered with a low- $\lambda$  model (center), mentioned in Section 3.2.3, are shown. (B) The overlay slices of the preprocessed images (left) are shown to compare the effect of DIR (right) on the testing data set with the most suitable parameter setting.

### 3.2.3 Qualitative evaluation

The images in Figure 6 served to evaluate the outcome of the parameter tuning. The impact of a DIR with  $\lambda < 1$  was studied with the middle image in Figure 6A, which shows the results of the deformation of a medium-architecture model with the parameters  $b = 1$ ,  $\lambda = 0.01$  and  $\alpha = 10^{-5}$ . The registered image includes many spiky distortions, which makes it impractical. The smoothness increases with higher values of  $\lambda$ . This confirms that models should be trained with  $\lambda \geq 1$ . The result of the most suitable setting is presented with the right image in Figure 6A. The largest volume deformation happened at the back of the head as also visible in Figure 5. The position of the tissue in the MRI slice differed from the skull in the CT slice. This misalignment was mainly reduced. Therefore, the model with the most suitable setting was applied to the testing data set, as visualised in Figure 6B. The same effects are noticeable, which means that the application of DIR reduces the spatial discrepancies in the back of the head.

## 4 Discussion

Direct multimodal DIR is a challenging and poorly explored research field due to the difficulties concerning the discrepancies in image representation of different modalities. The unsupervised and fully automated workflow for head CT and MRI scans takes several minutes to produce registered and deformed images based on rigid and deformable registrations. The preprocessing designed as the first workflow step showed the efficacy of generating rigidly aligned images. The eyes were chosen as the reference structure for the rigid registration, which caused larger distortions in the back of the head. These dislocations could be reduced by conventional existing rigid-registration methods. However, anatomical distortions due to patient positioning or distortion effects, which appear in the MRI scans with a radial degradation towards the outer regions of the body [29–32], would still be present. Therefore, deep-learning-based DIR is a promising approach to correct these distortions.

A difficult task in multimodal DIR is the choice of image-similarity metrics to be used in the loss function and to quantitatively validate the outcome. Considering the reduced bone intensities in the CT scans,  $m_{NCC}$  had proven to be most suited. For evaluation, the Dice similarity coefficient, used to assess the registration accuracy, was inefficient due to the uneven distribution of segments. Therefore, the mutual-information metric was used for more accurate evaluation. In addition, the registration performance of the CNN was measured with the Jacobian determinant of the DVF, which also included the volume changes outside the head. Thus, the mean value is mathematically expected to be unity for a proper diffeomorphic registration. Models with the setting  $\lambda < 1$ , for example, yielded  $\bar{m}_{JD} < 1$  (see Figure 4D), indicating registration errors, which are visible in the middle image of Figure 6A. To assess the physical volume changes after the registration, the Jacobian determinant has to be determined within the body contour. Then, the expected value of  $\bar{m}_{JD}$  is supposed to deviate from unity due to distortion effects in the MRI scans [4].

The large variety of deep-learning algorithms for DIR provides the potential for further work since this study only covered one possible CNN application. Compared to the direct application, artificially generated images based on deep learning can avoid multimodal registration problems. A related work [9] showed that the artificial-based procedure achieved results superior to direct multimodal DIR using splines with the software elastix [33]. Therefore, the deep-learning-based results obtained in this study need to be compared with common DIR approaches, which should be investigated in further work. The comparison of different procedures provides the advantages and disadvantages of the deep-learning approach. One advantage, for example, is the fast application of deep neural networks. These can cope with challenges such as physical growth or changes in brain morphology.

Multimodal DIR is an application barely supported in practice. The clinical integration of the proposed registration workflow would require simple application through clear instructions for the clinicians [4]. As the workflow is an unsupervised method, the application would still require supervision by trained clinicians, especially for quality assessment. In this study, the data sets contained scans of patients under the age of 18, which led to a variation in the head width of up to 40 mm. This is useful for DIR with deep learning. As the shape and morphology of the head vary from age to age especially for children, scans of each age group provide more information, such as anatomical structures and intensity distributions, for the CNN. Future work should deal with data expansion and loss-function variation. One option is to vary the similarity part of the loss function by considering other metrics. To the best of our knowledge, little is known about the influence of the high anatomical variability of paediatric cancer patients on the performance of deep-learning-based patient modelling, i.e., registration and contouring. As this impacts the quality and efficiency of treatment planning of a large patient fraction in proton therapy centers, more research and development should be carried out.

Ultimately, this study focused on the implementation of direct multimodal DIR with deep learning for head CT and MRI scans of paediatric patients. The registration method compensated distortions, which remain after rigid registration, by improving image alignment with unique deformation vector fields. The current study is a first step to facilitate treatment planning of paediatric cases with small geometric scales.

## Data availability statement

The data analysed in this study is subject to the following licenses/restrictions: The data sets presented in this article are not readily available because of the ethical reasons regarding patient consent. Requests to access these data sets should be directed to CB, christian.baeumer@uk-essen.de.

## Ethics statement

The studies involving humans were approved by Ethics Committee of the University of Duisburg–Essen. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

## Author contributions

AR: Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. ED: Investigation, Software, Writing–review and editing. FH: Resources, Writing–review and editing. KK: Conceptualisation, Funding acquisition, Project administration, Supervision, Writing–review and editing. BT: Conceptualisation, Funding acquisition, Project administration, Writing–review and editing. CB: Conceptualisation, Funding acquisition, Project administration, Resources, Supervision, Writing–review and editing.

## Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. The research was funded by the Mercator Research Center Ruhr with the grant number St-2019-0007. CB and FH acknowledge support from the Barbara-und-Hubertus-Trettner-Stiftung under project number T0355/36718/2020/sm.

## Acknowledgments

The authors would like to thank Armin Lühr for his scientific advice. Furthermore, the authors would like to thank Sarah Peters for checking the contours of the ventricles. The processing of the KiAPT cohort was supported by a grant from the Brigitte-und-Dr.-Konstanze-Wegener-Stiftung under project number 53.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Pereira G, Traughber M, Muzic R, Jr. The role of imaging in radiation therapy planning: past, present, and future. *Biomed Res Int* (2014) 2014:1–9. doi:10.1155/2014/231090
- Khan F. *The Physics of radiation therapy*. 4 edn. Lippincott Williams & Wilkins (2009).
- Chan HP, Samala R, Hadjiiski L, Zhou C. *Advances in experimental medicine and biology*. Springer (2020). Deep learning in medical image analysis
- Brock K, Mutic S, McNutt T, Li H, Kessler M. Use of image registration and fusion algorithms and techniques in radiotherapy: report of the AAPM radiation therapy committee task group No. 132. *Med Phys* (2017) 44:e43–76. doi:10.1002/mp.12256
- Hoffmann A, Oborn B, Mea M, Yan S, Bortfeld T, Knopf A, et al. MR-guided proton therapy: a review and a preview. *Radiat Oncol* (2020) 15:129. doi:10.1186/s13014-020-01571-x
- Pham T, Whelan B, Oborn B, Delaney G, Vinod S, Brighi C, et al. Magnetic resonance imaging (MRI) guided proton therapy: a review of the clinical challenges, potential benefits and pathway to implementation. *Radiother Oncol* (2022) 170:37–47. doi:10.1016/j.radonc.2022.02.031
- Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. Deep learning in medical image registration: a review. *Phys Med Biol* (2020) 65:20TR01. doi:10.1088/1361-6560/ab843e
- Zou J, Gao B, Song Y, Qin J. A review of deep learning-based deformable medical image registration. *Front Oncol* (2022) 12:1047215. doi:10.3389/fonc.2022.1047215
- McKenzie E, Santhanam A, Ruan D, O'Connor D, Cao M, Sheng K. Multimodality image registration in the head-and-neck using a deep learning-derived synthetic CT as a bridge. *Med Phys* (2020) 47:1094–104. doi:10.1002/mp.13976
- Thummerer A, Oria C, Pea Z, Visser S, Meijers A, Guterres Marmitt G, et al. Deep learning-based 4D-synthetic CTs from sparse-view CBCTs for dose calculations in adaptive proton therapy. *Med Phys* (2022) 49:6824–39. doi:10.1002/mp.15930
- Chang Y, Liang Y, Yang B, Qiu J, Pei X, Xu XG. Dosimetric comparison of deformable image registration and synthetic ct generation based on cbct images for organs at risk in cervical cancer radiotherapy. *Radiat Oncol* (2023) 18:3. doi:10.1186/s13014-022-02191-3
- Trimpl M, Primakov S, Pea L, Stride EPJ, Vallis KA, Gooding MJ. Beyond automatic medical image segmentation—the spectrum between fully manual and fully automatic delineation. *Phys Med Biol* (2022) 67:12TR01. doi:10.1088/1361-6560/ac6d9c
- de Vos B, Berendsen F, Vieregger M, Hea S, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal* (2019) 52:128–43. doi:10.1016/j.media.2018.11.010
- Zhao S, Lau T, Luo J, Chang E, Xu Y. Unsupervised 3D end-to-end medical image registration with volume twinning network. *IEEE J Biomed Health Inform* (2019) 24:1394–404. doi:10.1109/jbhi.2019.2951024
- Fu Y, Lei Y, Wang T, Higgins K, Bradley JD, Curran WJ, et al. LungRegNet: an unsupervised deformable image registration method for 4D-CT lung. *Med Phys* (2020) 47:1763–74. doi:10.1002/mp.14065
- Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *18th international conference on medical image computing and computer-assisted intervention* (2015). p. 234.
- Balakrishnan G, Zhao A, Sabuncu M, Guttag J, Dalca A. VoxelMorph: a learning framework for deformable medical image registration. *IEEE Trans Med Imaging* (2019) 38:1788–800. doi:10.1109/tmi.2019.2897538
- Sun L, Zhang S. Deformable MRI-Ultrasound registration using 3D convolutional neural network. In: *Simulation, image processing, and ultrasound systems for assisted diagnosis and navigation*, 11042. Springer International Publishing (2018).
- Spahr N, Thoduka S, Abolmaali N, Kikinis R, Schenk A. Multimodal image registration for liver radioembolization planning and patient assessment. *Int J Comput Assist Radiol Surg* (2019) 14:215–25. doi:10.1007/s11548-018-1877-5
- Peters S, Frisch S, Stock A, Merta J, Bäumer C, Blase C, et al. Proton beam therapy for pediatric tumors of the central nervous system—experiences of clinical outcome and feasibility from the KiProReg study. *Cancers* (2022) 14:5863. doi:10.3390/cancers14235863
- Bäumer C, Frakulli R, Kohl J, Nagaraja S, Steinmeier T, Worawongsakul R, et al. Adaptive proton therapy of pediatric head and neck cases using MRI-based synthetic CTs: initial experience of the prospective KiAPT study. *Cancers* (2022) 14:2616. doi:10.3390/cancers14112616
- Tsai DM, Lin CT. Fast normalized cross correlation for defect detection. *Pattern Recognit Lett* (2003) 24:2625–31. doi:10.1016/s0167-8655(03)00106-5
- Maes F, Loecckx D, Vandermeulen D, Suetens P. *Handbook of biomedical imaging: methodologies and clinical research*. Springer US (2015). Image registration using mutual information
- Crum W, Camara O, Hill D. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans Med Imaging* (2006) 25:1451–61. doi:10.1109/tmi.2006.880587
- Erdmann M, Glombitza J, Kasieczka G, Klemradt U. *Deep learning for Physics research*. 1 edn. World Scientific Publishing Co. Pte. Ltd. (2021).
- Balakrishnan G, Zhao A, Sabuncu M, Guttag J, Dalca A. An unsupervised learning model for deformable medical image registration. In: *2018 IEEE/CVF conference on computer vision and pattern recognition* (2018). p. 9252.
- Maas A, Hannun A, Ng A. Rectifier nonlinearities improve neural network acoustic models. In: *ICML workshop on deep learning for audio, speech and language processing* (2013).
- Kingma D, Ba J. Adam: a method for stochastic optimization. In: *3rd international conference for learning representations* (2015).
- Putz F, Mengling V, Perrin R, Masitho S, Weissmann T, Rösch J, et al. Magnetic resonance imaging for brain stereotactic radiotherapy. *Strahlenther Onkol* (2020) 196:444–56. doi:10.1007/s00066-020-01604-0
- Slagowski J, Ding Y, Wen Z, Fuller C, Chung C, Kadbi M, et al. Quantification of geometric distortion in magnetic resonance imaging for radiation therapy treatment planning. *Int J Radiat Oncol Biol Phys* (2018) 102:e547. doi:10.1016/j.ijrobp.2018.07.1527
- Pappas E, Alshantiry M, Moutsatsos A, Lababidi H, Alsafi K, Georgiou K, et al. MRI-related geometric distortions in stereotactic radiotherapy treatment planning: evaluation and dosimetric impact. *Technol Cancer Res Treat* (2017) 16:1120–9. doi:10.1177/1533034617735454
- Ulin K, Urie M, Cherlow J. Results of a multi-institutional benchmark test for cranial CT/MR image registration. *Int J Radiat Oncol Biol Phys* (2010) 77:1584–9. doi:10.1016/j.ijrobp.2009.10.017
- Klein S, Staring M, Murphy K, Vieregger M, Pluim J. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging* (2010) 29:196–205. doi:10.1109/tmi.2009.2035616