Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Deep learning-based conversion of MRI to synthetic CT for gamma knife radiosurgery – A large-sample study

Peng Zhou
Daping Hospital, Army Medical University
Jian Li
Chongqing Western Hospital
Wansheng Liu
Chongqing Western Hospital
Enyu Liu
Chongqing Western Hospital
Jia Luo
Daping Hospital, Army Medical University
Jiangfeng Wang
Xi' an Dayi Group Organization
Shuaishuai Xu
Shandong Provincial Hospital Affiliated to Shandong First Medical University
Shengxiu Jiao
Shandong Provincial Hospital Affiliated to Shandong First Medical University
Mingying Geng
380904661@qq.com

Daping Hospital, Army Medical University

Article

Keywords: Gamma knife radiosurgery, Synthetic CT, Deep learning, Large-sample analysis, MRI-only radiotherapy

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11503881/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Purpose

Gamma Knife radiosurgery (GKRS) relies on planning computed tomography (CT) for dose calculation, which introduces inherent CT/MR registration errors, unnecessary radiation exposure, and suboptimal soft-tissue contrast. While deep learning-based synthetic CT (sCT) generation from MRI offers a promising solution, existing studies predominantly utilize small sample sizes (typically fewer than 100 cases), which may limit prediction accuracy. This study aims to investigate the impact of a large-sample training dataset on sCT generation accuracy and evaluate its dosimetric performance in frameless GKRS.

Methods and Materials:

A total of 246 patients treated with GKRS for meningiomas were retrospectively enrolled. Eighteen cases were randomly selected for testing, while the remaining 228 cases were partitioned into a small-sample group (28 cases, M28) and a large-sample group (200 cases, M200) to train a U-Net-based MRI-to-CT translation model. The image synthesis accuracy was evaluated using mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM) across the whole body, soft tissue, bone, and intra-scalp regions. For the M200 model, dosimetric verification was performed by transferring the sCT and corresponding contours into the treatment planning system. Dose-volume histogram (DVH) parameters for the gross tumor volume (GTV) and various organs at risk (OARs) were calculated and statistically compared with the ground-truth CT plans.

Results

The M200 model consistently outperformed the M28 model across all evaluated regions and metrics. Specifically, M200 yielded significantly lower MAE (e.g., bone: 126.55 vs. 175.95; body: 18.53 vs. 25.13) and higher SSIM (e.g., bone: 0.91 vs. 0.84) compared to M28. Visual assessment and difference maps confirmed high anatomical fidelity with minimal HU errors, primarily localized at complex bony boundaries. Dosimetric evaluation demonstrated no statistically significant differences in target coverage (D98, D95, D2) or OAR sparing between the original CT and sCT plans (all $p > 0.05$). The DVH curves for both the GTV and OARs were nearly perfectly superimposed.

Conclusions

Increasing the training sample size substantially improves the accuracy of MRI-based sCT generation, particularly in bony structures. The sCT generated by the large-sample model provides sufficient HU

accuracy and dosimetric reliability for GKRS, offering a viable alternative for MRI-only treatment planning workflows.

1. Introduction

Gamma knife radiosurgery (GKRS) is a modality of radiation therapy. Its main structural component consists of 192 ^{60}Co radioactive sources arranged in a hemispherical array. These sources generate gamma rays, which are precisely focused onto intracranial lesions under image guidance. The relatively low-energy gamma rays from different directions converge at the intracranial target, forming a high-energy radiation volume that conforms closely to the shape of the lesion. This enables effective tumor cell eradication while sparing the surrounding normal tissues with minimal impact [1]. The therapeutic success of GKRS is critically dependent on the fidelity of the prescribed dose distribution, which is conventionally derived from electron density maps obtained via computed tomography (CT). CT provides the hounsfield units (HU) required for stoichiometric calibration to relative electron density and stopping-power ratios, thereby enabling reliable photon attenuation and dose deposition modelling within treatment planning systems [2, 3]. Nevertheless, CT has inherent limitations, including registration errors between CT and magnetic resonance imaging (MRI), unnecessary radiation exposure to patients, suboptimal soft-tissue contrast for target delineation, and reduced efficiency in the clinical workflow [4, 5].

MRI offers superior soft-tissue resolution and constitutes the modality of choice for target and organ-at-risk (OAR) delineation in GKRS, particularly for pathologies such as vestibular schwannomas, meningiomas, pituitary adenomas, and brain metastases [6, 7]. However, the MR signal—primarily governed by proton density and T1/T2 relaxation times—exhibits no intrinsic correlation with electron density or photon attenuation properties, rendering MRI inherently unsuitable for quantitative dose computation [8]. Consequently, current clinical practice necessitates rigid or deformable co-registration of the planning MRI with a dedicated planning CT, followed by dose calculation on the CT dataset [9, 10]. This multimodal registration paradigm introduces additional radiation burden and workflow complexity, but more critically, it introduces geometric misregistration errors that are particularly detrimental in the steep dose-gradient regions characteristic of gamma knife treatments [11].

To circumvent these limitations, the concept of synthetic CT (sCT) generation from MRI has emerged as a compelling alternative. The central premise is to infer a pseudo-CT volume that faithfully encodes HU values from the morphological and tissue-characteristic information embedded in MR images. The advent of deep learning has fundamentally transformed the field of medical image synthesis [12–14]. Data-driven architectures—most notably convolutional neural networks (CNNs), generative adversarial networks (GANs), and more recently diffusion models and transformer-based architectures—have enabled end-to-end learning of high-dimensional transformations without requiring explicit physical priors [15–17]. In the specific context of MRI-to-CT translation, deep learning models have demonstrated remarkable proficiency in producing sCT images with structural fidelity and dosimetric accuracy

comparable to real CT, as evidenced across multiple anatomical sites including the brain, pelvis, head and neck, and thorax [18–21].

Recent feasibility studies have provided encouraging evidence that deep learning-based sCT for GKRS is an attainable goal. Li et al. demonstrated that cycleGAN-generated sCT from MRI could be utilised for convolution-based dose calculation in gamma knife plans, with no statistically significant difference between sCT and reference CT for most dosimetric parameters, although deviations of up to 3.3% were observed when the stereotactic frame was absent from the training set [3]. For gamma knife systems, which typically employ hypofractionated doses of approximately 6–7 Gy per fraction, even small differences in dose calculation are critical for improving treatment efficacy and reducing patient toxicity [22]. Only small CT number differences between the synthetic CT and the original CT can ensure minimal dose discrepancies. Therefore, improving the prediction accuracy of CT values to reduce dose errors is beneficial for gamma knife radiosurgery. On the other hand, to the best of our knowledge, there have been numerous reports on synthetic CT generation in radiotherapy and other medical fields [23–29]; however, most of these studies have employed small-sample models, with total training and testing cases not exceeding 100. Nearly all researchers have acknowledged that small-sample models may be insufficient for improving CT prediction accuracy. Therefore, in the present work, we attempted to investigate the impact of a large-sample model on CT generation accuracy and to evaluate its performance in gamma knife radiosurgery. In this work, both small-sample and large-sample datasets were used to train CT prediction models. The dataset size for the small-sample model was determined with reference [3] to the sample sizes reported in the literature on CT prediction models.

2. Materials and methods

2.1 Training data

This work was approved by Ethics Committee of Chongqing Western Hospital (NO: xqyy-hzxm-2025004). The informed written consent was obtained from patients taking part. All methods were performed in accordance with the relevant guidelines and regulations of Chongqing Western Hospital. Following internal ethics committee approval, this study retrospectively enrolled 246 brain patients treated with gamma knife radiosurgery at Chongqing Western Hospital between January 2024 and June 2025. Eighteen cases were randomly selected for testing, while the remaining 228 cases were partitioned into two groups of 28 and 200 cases for training the two prediction models, respectively.

The baseline demographic and clinical characteristics of the study cohort are summarized in Table 1. The cohort consisted of 147 female patients (59.8%) and 99 male patients (40.2%). Regarding age distribution, 129 patients (52.4%) were under 60 years old, while 117 patients (47.6%) were aged 60 years or older. The treated lesions were distributed across various intracranial locations, with the posterior fossa being the most common site (37.0%), followed by the parietal lobe (18.3%), sellar region (12.6%), frontal lobe (11.8%), temporal lobe (11.8%), occipital lobe (7.3%), and lateral ventricle (1.2%). In terms of lesion size, 66 cases (26.8%) had a maximum diameter of ≤ 2 cm, and 180 cases (73.2%) were

> 2 cm. Additionally, 138 lesions (56.1%) had a treatment volume of ≤ 3 cc, whereas 108 lesions (43.9%) were > 3 cc.

Table 1
Patient characteristics.

Variable	Category	Value	Percentage (%)
Gender	Male	99	40.2
	Female	147	59.8
Age	< 60 years	129	52.4
Treatment site	≥ 60 years	117	47.6
	Parietal lobe	45	18.3
	Frontal lobe	29	11.8
	Temporal lobe	29	11.8
	Lateral ventricle	3	1.2
Lesion size	Posterior fossa	91	37.0
	Sellar region	31	12.6
	Occipital lobe	18	7.3
	≤ 2 cm	66	26.8
Lesion volume	> 2 cm	180	73.2
	≤ 3 cc	138	56.1
	> 3 cc	108	43.9

2.2 Gamma knife treatment plan

All gamma knife treatment plans were designed by senior physicists and evaluated by physicians. The overall planning workflow was as follows: First, patient CT images were imported based on the positioning ID. MR images from T1, T2, and Dark sequences were selected in the Patient module and registered to the CT image as the primary sequence for subsequent contouring of OARs and target volumes. Next, the Patient Origin was set in the Images module, and the Imaging Center was placed near the target or at a location with prominent bony landmarks. Subsequently, contouring of target volumes and OARs was performed in the Contours module, and the body outer contour was modified accordingly. Following these preparatory steps, a gamma knife plan was added in the Plans module with the appropriate device model selected, after which the gamma knife plan design phase was initiated. The prescription dose was set in the Stereotactic module, and the appropriate collimator size (0.4, 0.8, 1.4, or 1.8 cm) was selected based on the coverage of the target volume by the 50% prescription isodose line. Different shot placement strategies were adopted according to target volume size: for larger target

volumes, larger collimators were selected, and the isodose distribution was observed in real time using the RealTime function, on the basis of which subsequent collimator sizes and shot placements were determined; for smaller target volumes, smaller collimators were chosen, and shots were placed under real-time isodose display, with shot positions and weights finely adjusted to effectively improve conformality (defined as the ratio of the target volume encompassed by the prescription isodose line to the total volume encompassed by that isodose line). Each shot could be manually dragged or fine-tuned in the Shot Setting interface along the right–left (R–L), anterior–posterior (A–P), and inferior–superior (I–S) directions, in combination with shot weight optimization, to meet clinical requirements. Throughout the shot placement and weighting adjustment process, the fundamental principle was to ensure that target coverage (the percentage of the target volume encompassed by the prescription isodose line) exceeded 95%, provided that OAR dose constraints were not exceeded, thereby ensuring tumor control probability. Finally, the doses to OARs and target coverage were evaluated in the Evaluation module. The plan could only be accepted and transmitted to the Gamma Knife treatment unit for delivery after it had been approved by the physician.

2.3 CT prediction model and workflow

In this study, a U-Net-based medical image translation network was employed to achieve end-to-end mapping from MR images to sCT images. The network architecture follows the encoder–decoder paradigm, taking MR slices as input and predicting CT value representations from MR intensity distributions through structures including multi-scale feature extraction, downsampling, upsampling, and skip connections. The network architecture is illustrated in Fig. 1. On the encoding side, the network progressively enlarges the receptive field through multi-level feature extraction and downsampling operations, extracting multi-scale features ranging from shallow spatial textures to deep anatomical semantics. The number of feature channels is gradually increased from 16 in the shallow layers to 32, 64, and 128, and further expanded to 256 in the bottleneck layer, thereby enhancing the network's representational capacity for the overall head structure, tissue boundaries, and complex anatomical regions. On the decoding side, the network restores the spatial resolution of the image through gradual upsampling, and integrates shallow detail features with deep semantic features via skip connections from the corresponding encoding layers. This design facilitates the preservation of both global anatomical consistency and local structural details during reconstruction, improving the restoration of critical regions such as bony structures, air cavities, and soft-tissue boundaries. Finally, the network generates a single-channel sCT image through the output mapping layer, achieving end-to-end prediction from MR images to synthetic CT images.

The network was trained using NVIDIA Quadro RTX 8000 GPU with 48*4 GB memory. The training parameters is: learning rate: in the current main training script train_test.py, the default initial learning rate is 1e-4, corresponding to the parameter --lr. The model optimizer is initialized using this learning rate. Learning rate adjustment strategy: the training code uses the ExponentialLR learning rate scheduler, with the default decay coefficient gamma set to 0.9. g_scheduler.step() is called at the end of each epoch. Therefore, the learning rate gradually decays with the training epochs and can be expressed as

lr_epoch=1e-4× 0.9^epoch, which is used to reduce the magnitude of parameter updates in the later stages of training. Optimizer: model training uses the AdamW optimizer, with the parameters set to betas=(0.9, 0.999) and the default weight_decay value set to 1e-4. AdamW introduces decoupled weight decay based on Adam's adaptive gradient update mechanism, which helps improve training stability and reduce the risk of overfitting.

2.4 sCT Plan dose recalculation

Due to limitations of the RT Pro TPS/TaiChi 4960 version, the fusion and registration between the synthetic CT and the original CT were performed in RayStation 9A. After the target volumes and OARs delineated on the original CT images were copied onto the synthetic CT, the synthetic CT images with the complete structure set were transferred back to RT Pro TPS/TaiChi 4960 for subsequent dose calculation. To account for potential minor differences in dose calculation across different versions of the RT Pro TPS, we first recalculated the dose on the original CT using RT Pro TPS/TaiChi 4960 based on the patient's original plan. Subsequently, with the Patient Origin and all shot positions and weights kept strictly consistent, the dose was recalculated on the synthetic CT under identical conditions. Finally, the dose-volume histograms (DVHs) and dose distributions obtained from the original CT and synthetic CT were systematically compared. Figure 2 illustrates the workflow of this study, including key steps such as synthetic CT generation and dose calculation on the synthetic CT.

2.5 Model evaluation

In this study, the mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and structural similarity index measure (SSIM) were employed to evaluate the CT number differences between the original CT and the synthetic CT in the whole-body, bone, soft-tissue and region inside the scalp contour. The definitions of these metrics are as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |sCT_i - CT_i|$$

$$PSNR = 20log_{10}\left(\frac{MAX_{CT}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(sCT_i - CT_i)^2}}\right)$$

$$SSIM = \frac{(2\mu_{sCT}\mu_{CT} + C_1)(2\sigma_{sCT,CT} + C_2)}{(\mu_{sCT}^2 + \mu_{CT}^2 + C_1)(\sigma_{sCT}^2 + \sigma_{CT}^2 + C_2)}$$

Where, sCT_i represents the *ith* voxel CT value predicted by deep learning models and the CT_i represents the *ith* voxel CT value of original CT. MAX_{CT} represents the maximum CT value of original CT. μ_{sCT} denotes the mean of image sCT, μ_{CT} is the mean of image CT; σ_{sCT} denotes the standard deviation of image sCT, σ_{CT} denotes the standard deviation of image CT; $\sigma_{sCT,CT}$ is the covariance of image sCT and CT. In the SSIM formula, C_1 and C_2 are small constants used to stabilize the division

when the denominator approaches zero. A higher PSNR indicates a smaller pixel-wise error between the predicted image and the reference image, corresponding to less distortion and better image quality. For SSIM, a value closer to 1 indicates greater similarity between the two images.

The GTV dose coverage difference was evaluated by using the dose-volume indicators D98, D95 and D2. As for OARs, we used the mean dose, the maximum dose and D50 parameters to evaluate the difference of CT plan and the sCT plan. The statistical significance of differences was tested using SPSS (IBM-SPSS statistics, version 17, Inc, Chicago, IL). Wilcoxon Signed Rank test were used to compare the different parameters. A p-value < 0.05 was considered statistically significant.

We hypothesized that larger errors in the predicted CT values would lead to greater dose discrepancies. Therefore, rather than evaluating the dosimetric performance of the small-sample model, we focused on the dose differences between the predicted CT and the original CT generated by the large-sample model.

3. Results

Table 2 presents the MAE, PSNR, and SSIM values achieved by the two models across different regions in the test cases. Overall, the model trained with the large-sample dataset (M200) consistently outperformed the small-sample model (M28) across all evaluated regions and metrics. As can be seen from the results in the table, for the MAE values, the model trained with the large-sample dataset achieved lower MAE values than the small-sample model in the body (18.53 vs. 25.13), soft tissue (12.72 vs. 17.03), bone (126.55 vs. 175.95), and the region within the scalp (54.34 vs. 78.81). Notably, the bone region exhibited the highest MAE values among all regions for both models, reflecting the inherent difficulty in predicting CT numbers for bony structures. For PSNR, higher values were observed for M200 compared with M28 across all regions: body (30.53 ± 1.08 vs. 27.60 ± 0.84), soft tissue (34.84 ± 2.89 vs. 33.06 ± 2.21), bone (27.00 ± 1.08 vs. 24.08 ± 1.03), and the region within the scalp (27.89 ± 1.94 vs. 25.87 ± 1.49). These results indicate that the large-sample model produced synthetic CT images with reduced pixel-wise error and distortion relative to the reference CT. Similarly, the SSIM values for M200 were consistently higher than those for M28, demonstrating superior structural similarity to the original CT. The SSIM improved from 0.94 ± 0.01 to 0.96 ± 0.01 in both the body and soft tissue regions, from 0.84 ± 0.02 to 0.91 ± 0.02 in bone, and from 0.81 ± 0.03 to 0.88 ± 0.03 in the region within the scalp. Notably, the lowest SSIM values were observed in the bone and intra-scalp regions for both models, suggesting that these regions remain the most challenging for accurate CT synthesis. Taken together, these findings demonstrate that increasing the training sample size substantially improves the accuracy of synthetic CT generation across all evaluated head regions, with the most pronounced improvements observed in bony structures and the region within the scalp.

Table 2
MAE, PSNR, and SSIM values of the model in different head regions for the test cases.

Region	M28 M200		M28	M200	M28	M200
Body	MAE	MAE	PSNR	PSNR	SSIM	SSIM
Soft tissue	25.13 ± 4.74	18.53 ± 4.03	27.60 ± 0.84	30.53 ± 1.08	0.94 ± 0.01	0.96 ± 0.01
Bone	17.03 ± 4.20	12.72 ± 3.69	33.06 ± 2.21	34.84 ± 2.89	0.94 ± 0.01	0.96 ± 0.01
Region within the scalp	175.95 ± 20.63	126.55 ± 17.90	24.08 ± 1.03	27.00 ± 1.08	0.84 ± 0.02	0.91 ± 0.02
	78.81 ± 7.33	54.34 ± 7.37	25.87 ± 1.49	27.89 ± 1.94	0.81 ± 0.03	0.88 ± 0.03

Figure 3 illustrates a visual comparison of the original CT, the sCT generated by the large-sample model, and the corresponding HU difference maps (CT-sCT) at three representative transverse slices of the head. Visually, the sCT images demonstrate a high degree of anatomical similarity to the original CT images. The continuous bony structures of the skull, the intricate bony details of the nasal cavity and paranasal sinuses, as well as the air-tissue interfaces, are accurately reconstructed in the sCT. Furthermore, the soft tissue regions and background exhibit no apparent visual discrepancies compared with the ground-truth CT. The difference maps (third column) further quantify the prediction accuracy. As indicated by the color scale ranging from − 1000 to 1000 HU, the majority of the brain parenchyma and soft tissue regions appear in a uniform teal color, corresponding to an HU error close to zero.

Figure 4 presents a representative case comparing the spatial dose distributions calculated on the original CT and the sCT, alongside their corresponding DVHs. Visual inspection of the dose distributions across the axial, coronal, and sagittal planes reveals a high degree of consistency between the CT and sCT plans. The high-dose region at the target is well-covered, and the steep dose gradients sparing the surrounding normal tissues are accurately maintained in the sCT. The spatial dose difference map (CT-sCT) further confirms this observation; the majority of the brain volume exhibits a dose difference near zero (white color), indicating highly consistent dose deposition. Minor dose discrepancies are only observable at the boundaries of complex heterogeneous tissues (e.g., near the skull), which are clinically negligible. The highly consistent dose distributions are further corroborated by the DVH curves. For the target volume (PTV), the curves for the CT (solid red line) and sCT (dashed red line) are nearly perfectly superimposed, demonstrating that target coverage and dose homogeneity are preserved. Similarly, the DVH curves for all evaluated OARs—including the brainstem, lenses, optic nerves, eyes, optic chiasm, temporal lobes, pituitary, and temporomandibular joints—show excellent agreement between the two datasets. Both the low-dose and high-dose regions of the OARs fall within a nearly identical volume percentage. Overall, these results indicate that the sCT generated by the large-sample model provides comparable dosimetric accuracy to the original CT, ensuring reliable treatment planning for Gamma Knife radiosurgery.

Table 3 presents the dosimetric comparison between the original CT and the sCT generated by the large-sample model across 18 testing patients. The dosimetric parameters, including mean dose, maximum dose, D98, D95, D2, and D50, were evaluated for the GTV and various OARs. For the target volume, the dosimetric differences between the CT and sCT plans were negligible. Specifically, the D98 was 2145.92 ± 143.39 cGy for CT versus 2143.11 ± 146.38 cGy for sCT ($p = 0.23$), the D95 was 2249.22 ± 147.27 cGy versus 2247.37 ± 149.95 cGy ($p = 0.34$), and the D2 was 4107.98 ± 270.77 cGy versus 4108.22 ± 272.64 cGy ($p = 0.81$). These results indicate that target coverage and dose homogeneity were highly preserved using the sCT. Similarly, the dose distributions for all evaluated OARs remained virtually unchanged. For the brain, the D50 and mean dose were highly consistent between the two datasets (36.89 ± 24.93 cGy vs 36.73 ± 24.76 cGy, $p = 0.21$; and 90.68 ± 46.81 cGy vs 90.53 ± 46.79 cGy, $p = 0.30$, respectively). OARs with low dose exposure, such as the lenses and eyes, as well as critical structures like the brainstem, optic nerves, and optical chiasm, demonstrated excellent dosimetric agreement, with all associated p-values well above 0.05. Furthermore, no statistically significant differences were observed for the temporal lobes, pituitary, or temporomandibular joints (TMJs) ($p > 0.05$ for all). Overall, all evaluated dosimetric parameters exhibited no statistically significant differences between the CT and sCT plans. These findings demonstrate that the sCT generated by the large-sample model provides sufficient dosimetric accuracy for Gamma Knife radiosurgery planning, ensuring reliable dose calculation without compromising target coverage or OAR sparing.

Table 3
Data shown are the averages and standard deviations of the respective parameters for the 18 testing patients. SD represents standard deviation. The sCT was achieved by the large training sample model.

Structures	Dosimetry Parameters	CT (Mean ± SD)	sCT (Mean ± SD)	P-Value
GTV	D_{98} (cGy)	2145.92 ± 143.39	2143.11 ± 146.38	0.23
	D_{95} (cGy)	2249.22 ± 147.27	2247.37 ± 149.95	0.34
	D_{2} (cGy)	4107.98 ± 270.77	4108.22 ± 272.64	0.81
Brain	D_{50} (cGy)	36.89 ± 24.93	36.73 ± 24.76	0.21
	Mean dose (cGy)	90.68 ± 46.81	90.53 ± 46.79	0.30
Brainstem	Max dose (cGy)	110.55 ± 116.54	109.38 ± 114.51	0.28
Left eye	Mean dose (cGy)	9.41 ± 14.63	9.51 ± 14.67	0.81
Right eye	Mean dose (cGy)	12.04 ± 21.60	12.04 ± 21.55	0.35
Left optic nerve	Max dose (cGy)	57.99 ± 140.06	58.29 ± 141.99	0.27
Right optic nerve	Max dose (cGy)	108.50 ± 294.22	108.15 ± 292.14	0.09
Optical chiasm	Max dose (cGy)	59.27 ± 86.98	59.17 ± 86.74	0.25
Left lens	Max dose (cGy)	12.51 ± 22.36	12.68 ± 22.51	0.38
Right lens	Max dose (cGy)	14.34 ± 31.23	14.21 ± 31.02	0.27
Left inner ear	Max dose (cGy)	108.36 ± 329.31	107.78 ± 327.91	0.28
Right inner ear	Max dose (cGy)	61.93 ± 92.21	61.25 ± 90.91	0.07
Left temporal lobe	D_{50} (cGy)	49.97 ± 77.91	49.99 ± 78.77	0.21
Right temporal lobe	D_{50} (cGy)	31.06 ± 31.42	31.04 ± 31.73	0.11
Pituitary	Max dose (cGy)	51.34 ± 91.78	51.53 ± 92.53	0.46
Left TMJ	Mean dose (cGy)	15.52 ± 36.29	15.92 ± 37.42	0.57
Right TMJ	Mean dose (cGy)	14.90 ± 23.30	14.62 ± 23.05	0.64

4. Discussion

In this study, we developed a U-Net-based deep learning model to synthesize CT images from MRI for frameless gamma knife radiosurgery and evaluated the impact of training dataset size on image prediction accuracy and subsequent dose calculation. Our results demonstrate that increasing the

training sample size significantly enhances the synthesis fidelity across various anatomical regions of the head. Crucially, the dosimetric evaluation confirmed that the sCT generated by our large-sample model provides highly accurate dose calculations, with no statistically significant deviations from the ground-truth CT plans.

In Fig. 3, pronounced HU differences are predominantly localized at the boundaries of complex bony structures and air cavities (e.g., the ethmoid sinus and the skull base), appearing as localized yellow and blue regions. These artifacts are typical in MRI-to-CT synthesis due to the lack of signal in cortical bone and air on conventional MRI, as well as potential minor registration inaccuracies. Despite these localized boundary discrepancies, the overall HU error across the entire image remains minimal. These visual findings confirm that the large-sample model is capable of generating high-fidelity sCT images with accurate tissue attenuation characteristics, which is essential for reliable dose calculation in gamma knife radiosurgery.

The synthesis of CT from MRI has been widely explored; however, a notable limitation in the current literature is the reliance on relatively small datasets, typically comprising 50 to 100 cases. Consistent with previous findings, our study confirms that MRI-to-CT synthesis is particularly challenging in regions with low proton density or absent MR signal, such as the cortical bone and air cavities. In our comparative analysis, the M28 model exhibited substantial prediction errors in the bone region (MAE = 175.95), which may reflect the limited capacity of small datasets to capture the highly heterogeneous spatial distribution of bony structures. By expanding the training cohort to 200 cases, the M200 model achieved a pronounced reduction in MAE (126.55) and a superior SSIM (0.91) in the bone region. This suggests that large-sample datasets enable the network to learn more generalized and robust morphological features, mitigating the inherent ambiguities of bone signal void on conventional MRI.

In evaluating image synthesis quality, while MAE and PSNR provide valuable voxel-wise error metrics, SSIM offers a distinct advantage by assessing structural information, luminance, and contrast. Given that GKRS demands high spatial precision, the high SSIM values observed in our study—particularly in soft tissue (0.96) and the body region (0.96)—indicate that the sCT maintains excellent anatomical structural integrity, a prerequisite for reliable stereotactic targeting. Although slight HU discrepancies inevitably persist at tissue boundaries (e.g., the skull base and ethmoid sinus), these localized errors did not translate into clinically meaningful dose deviations.

The dosimetric evaluation across 18 test patients further validated the clinical applicability of the large-sample model. Because GKRS typically delivers high-dose, hypofractionated radiation (e.g., 6–7 Gy per fraction) in close proximity to critical neurological structures, even minor CT value inaccuracies could theoretically alter dose deposition. In our study, all evaluated dosimetric parameters, including D98, D95, and D2 for the GTV, as well as the maximum and mean doses to various OARs (e.g., brainstem, optic nerves, and lenses), exhibited no statistically significant differences between the sCT and original CT plans (all $p > 0.05$). The near-perfect superimposition of the DVH curves and the negligible differences in

spatial dose distribution maps corroborate that the sCT accurately models electron density for photon attenuation, thereby ensuring that target coverage and OAR sparing are uncompromised.

Despite these promising results, this study has several limitations. First, it is a single-center retrospective study; multi-center validation with diverse MRI scanners and acquisition protocols is necessary to confirm the generalizability of the large-sample model. Second, while the dataset of 200 patients represents a significant increase compared to prior literature, the optimal sample size threshold and the potential for further performance gains with even larger datasets remain to be determined. Finally, the current study focused on meningioma patients; future work should extend this large-sample approach to other intracranial pathologies, such as brain metastases or vestibular schwannomas.

5. Conclusions

This study demonstrates that a large-sample deep learning model can generate high-quality sCT images from MRI with excellent anatomical fidelity. The resulting sCT provides highly reliable dosimetric accuracy for gamma knife radiosurgery, MRI-only stereotactic treatment planning workflows that eliminate the need for unnecessary CT radiation exposure and reduce registration uncertainties.

Authors' contributions

Peng Zhou and Jian Li: Methodology, Software, Supervision, Statistical analyses, Validation, Writing-original draft. Jia Luo, Wansheng Liu, Enyu Liu, Shuaishuai Xu: Data curation and Formal Analysis. Jiangfeng Wang: Image Processing, Image Analysis. Shengxiu Jiao and Mingying Geng: Conceptualization, Project administration, Writing-review and editing and Investigation.

Declarations

Conflict of interest

The authors declare that they have no conflict of interest related to this work.

Internal Review Board (IRB) statement

This work was approved by the Ethics Committee of Chongqing Western Hospital (NO: xqyy-hzxm-2025004). The informed written consent was obtained from patients taking part.

Funding

No Funding

Author Contribution

Peng Zhou and Jian Li: Methodology, Software, Supervision, Statistical analyses, Validation, Writing-original draft. Jia Luo, Wansheng Liu, Enyu Liu, Shuaishuai Xu: Data curation and Formal Analysis. Jiangfeng Wang: Image Processing Image Analysis. Shengxiu Jiao and Mingying Geng: Conceptualization, Project administration, Writing-review and editing and Investigation

Data Availability

All data supporting the findings of this study are available at the following URL: https://doi.org/10.5281/zenodo.23786061.

References

1. Xu, A. et al. Gamma Knife radiosurgery with CT imagebased dose calculation. *J. Appl. Clin. Med. Phys.* ;**16**(6). (2015).
2. Razinskas, G. et al. Validation and optimization of dual-energy CT for accurate dose calculation in photon radiotherapy of brain metastases. *Phys. Med.* **141**, 105702 (2026).
3. Li, F. et al. Synthetic CT for gamma knife radiosurgery dose calculation: A feasibility study. *Phys. Med.* **125**, 104504 (2024).
4. Zhao, Y. et al. Compensation cycle consistent generative adversarial networks (Comp-GAN) for synthetic CT generation from MR scans with truncated anatomy. *Med. Phys.* **50** (7), 4399–4414 (2023).
5. Nijskens, L., van den Berg, C. A. T., Verhoeff, J. J. C. & Maspero, M. Exploring contrast generalisation in deep learning-based brain MRI-to-CT synthesis. *Phys. Med.* **112**, 102642 (2023).
6. Michalet, M., Riou, O., Azria, D., Decoene, C. & Crop, F. News in magnetic resonance imaging use for radiation oncology. *Cancer Radiother*. **26** (6–7), 784–788 (2022).
7. Chetty, I. J. et al. MRI-guided Radiotherapy (MRgRT) for Treatment of Oligometastases: Review of Clinical Applications and Challenges. *Int. J. Radiat. Oncol. Biol. Phys.* **114** (5), 950–967 (2022).
8. Lei, Y. et al. MRI-based synthetic CT generation using semantic random forest with iterative refinement. *Phys. Med. Biol.* **64** (8), 085001 (2019).
9. Jiao, S., Zhao, X., Zhou, P. & Geng, M. Technical note: MR image-based synthesis CT for CyberKnife robotic stereotactic radiosurgery. *Biomed. Phys. Eng. Express* ;**10**(5) (2024).
10. Ma, X. et al. MRI-Only Radiotherapy Planning for Nasopharyngeal Carcinoma Using Deep Learning. *Front. Oncol.* **11**, 713617 (2021).
11. Xu, A. Y. et al. Gamma Knife radiosurgery with CT image-based dose calculation. *J. Appl. Clin. Med. Phys.* **16** (6), 119–129 (2015).

12. Han, X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med. Phys.* **44** (4), 1408–1419 (2017).
13. Liu, F., Yadav, P., Baschnagel, A. M. & McMillan, A. B. MR-based treatment planning in radiation therapy using a deep learning approach. *J. Appl. Clin. Med. Phys.* **20** (3), 105–114 (2019).
14. Boulanger, M. et al. Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review. *Phys. Med.* **89**, 265–281 (2021).
15. Maspero, M. et al. Deep learning-based synthetic CT generation for paediatric brain MR-only photon and proton radiotherapy. *Radiother Oncol.* **153**, 197–204 (2020).
16. Ranjan, A., Lalwani, D. & Misra, R. GAN for synthesizing CT from T2-weighted MRI data towards MR-guided radiation treatment. *MAGMA* **35** (3), 449–457 (2022).
17. Yang, X. et al. CNN-based multi-modal radiomics analysis of pseudo-CT utilization in MRI-only brain stereotactic radiotherapy: a feasibility study. *BMC Cancer*. **24** (1), 59 (2024).
18. Yip, T. T. Y., Li, Z. & Li, T. Clinical validation of MR-generated synthetic CT by MRCAT for brain tumor radiotherapy. *J. Appl. Clin. Med. Phys.* **26** (1), e14494 (2025).
19. Shu, X. et al. Knowledge-based deep residual U-Net (DRU) for synthetic CT generation using a single MR volume for frameless radiosurgery. *J. Appl. Clin. Med. Phys.* **27** (1), e70343 (2026).
20. Palmér, E. et al. Head and neck cancer patient positioning using synthetic CT data in MRI-only radiation therapy. *J. Appl. Clin. Med. Phys.* **23** (4), e13525 (2022).
21. Longuefosse, A. et al. Generating High-Resolution Synthetic CT from Lung MRI with Ultrashort Echo Times: Initial Evaluation in Cystic Fibrosis. *Radiology* **308** (1), e230052 (2023).
22. Samanci, Y. et al. Hypofractionated frameless gamma knife radiosurgery for large metastatic brain tumors. *Clin. Exp. Metastasis*. **38** (1), 31–46 (2021).
23. Parrella, G. et al. Synthetic CT in Carbon Ion Radiotherapy of the Abdominal Site. *Bioeng. (Basel)*. **10** (2), 250 (2023).
24. Hsu, S. H. et al. Synthetic CT generation for MRI-guided adaptive radiotherapy in prostate cancer. *Front. Oncol.* **12**, 969463 (2022).
25. Pan, S. et al. Synthetic CT generation from MRI using 3D transformer-based denoising diffusion model. *Med. Phys.* **51** (4), 2538–2548 (2024).
26. Chen, L., Liang, X., Shen, C., Jiang, S. & Wang, J. Synthetic CT generation from CBCT images via deep learning. *Med. Phys.* **47** (3), 1115–1125 (2020).
27. Moglioni, M. et al. Synthetic CT imaging for PET monitoring in proton therapy: a simulation study. *Phys. Med. Biol.* ;**69**(6) (2024).
28. Lu, Q., Luo, F., Shi, J. & Xu, K. Synthetic CT generation from CBCT based on structural constraint cycle-EEM-GAN. *Biomed. Phys. Eng. Express* ;**10**(6) (2024).
29. Arbabi, S. et al. MRI-based synthetic CT in the detection of knee osteoarthritis: Comparison with CT. *J. Orthop. Res.* **41** (11), 2530–2539 (2023).

Figures

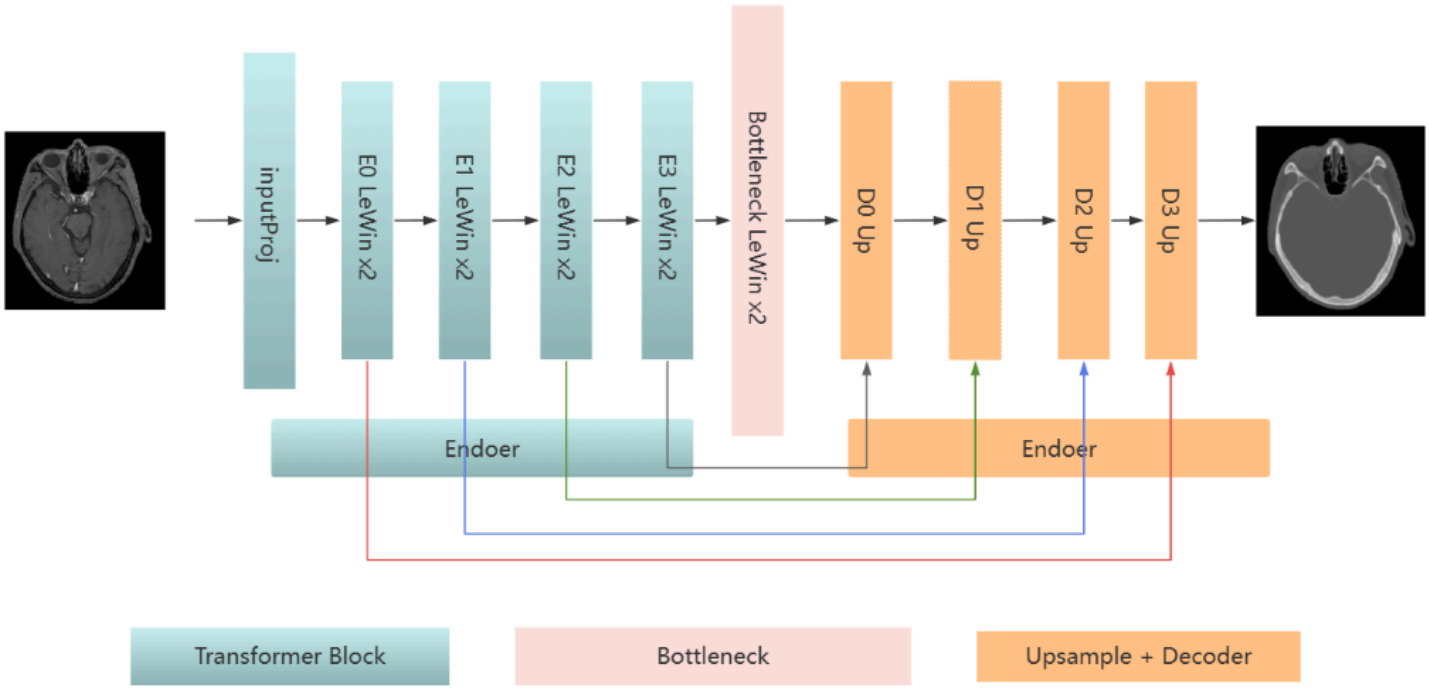


Figure 1

The basic framework of the model.

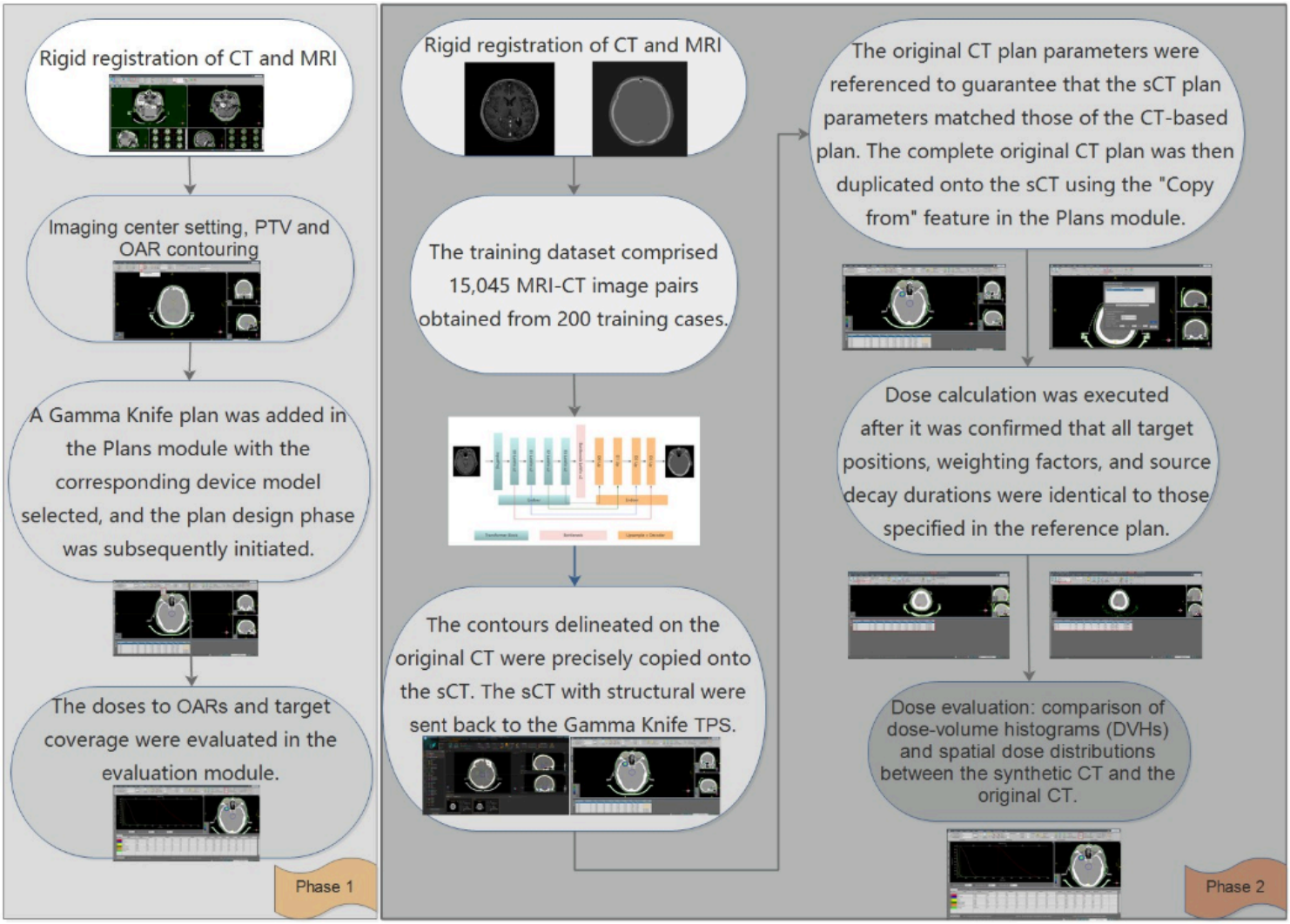


Figure 2

The workflow of this study

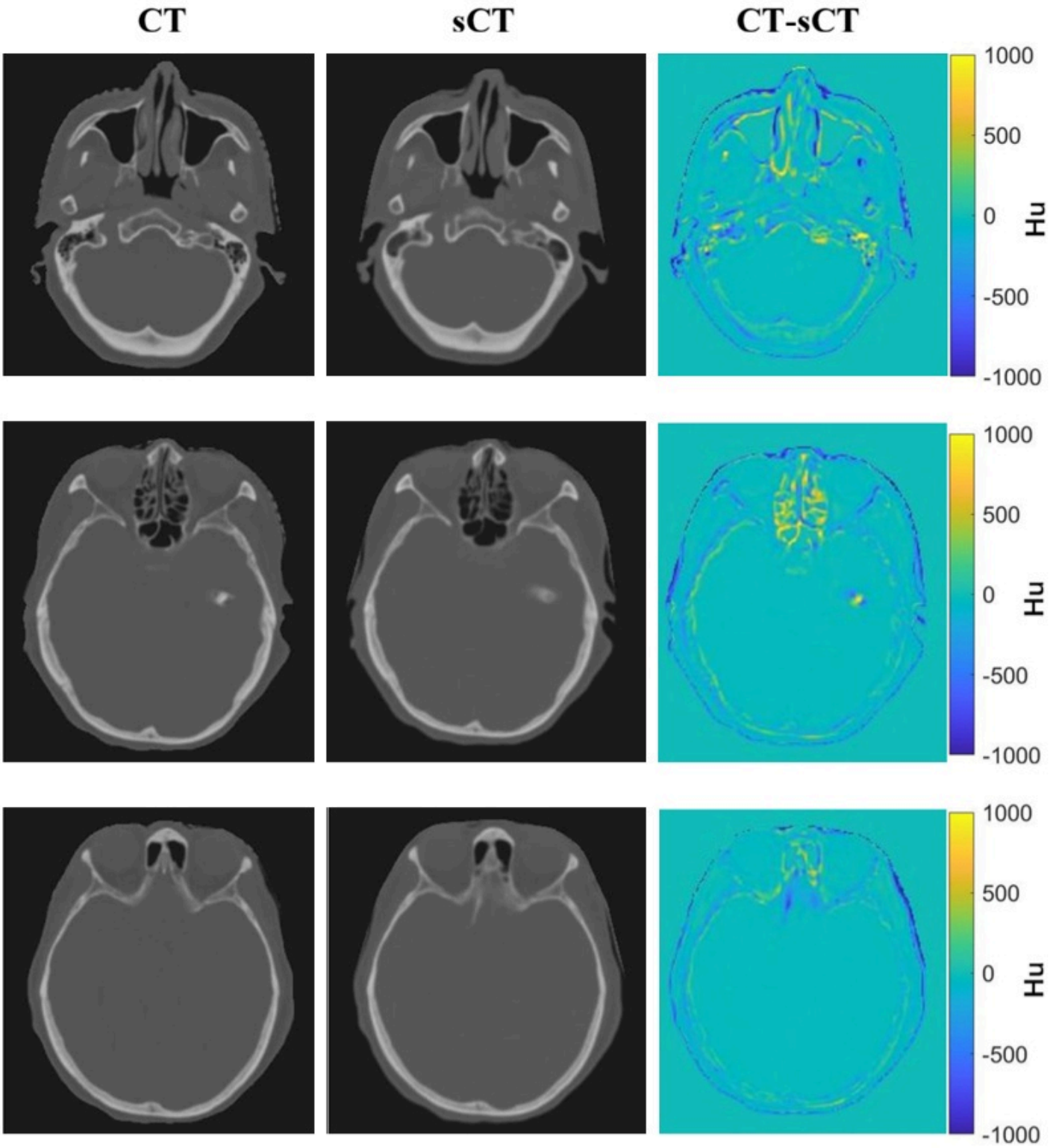


Figure 3

The original CT, predicted sCT and the HU difference between CT and sCT. The sCT was achieved by the large training sample model.

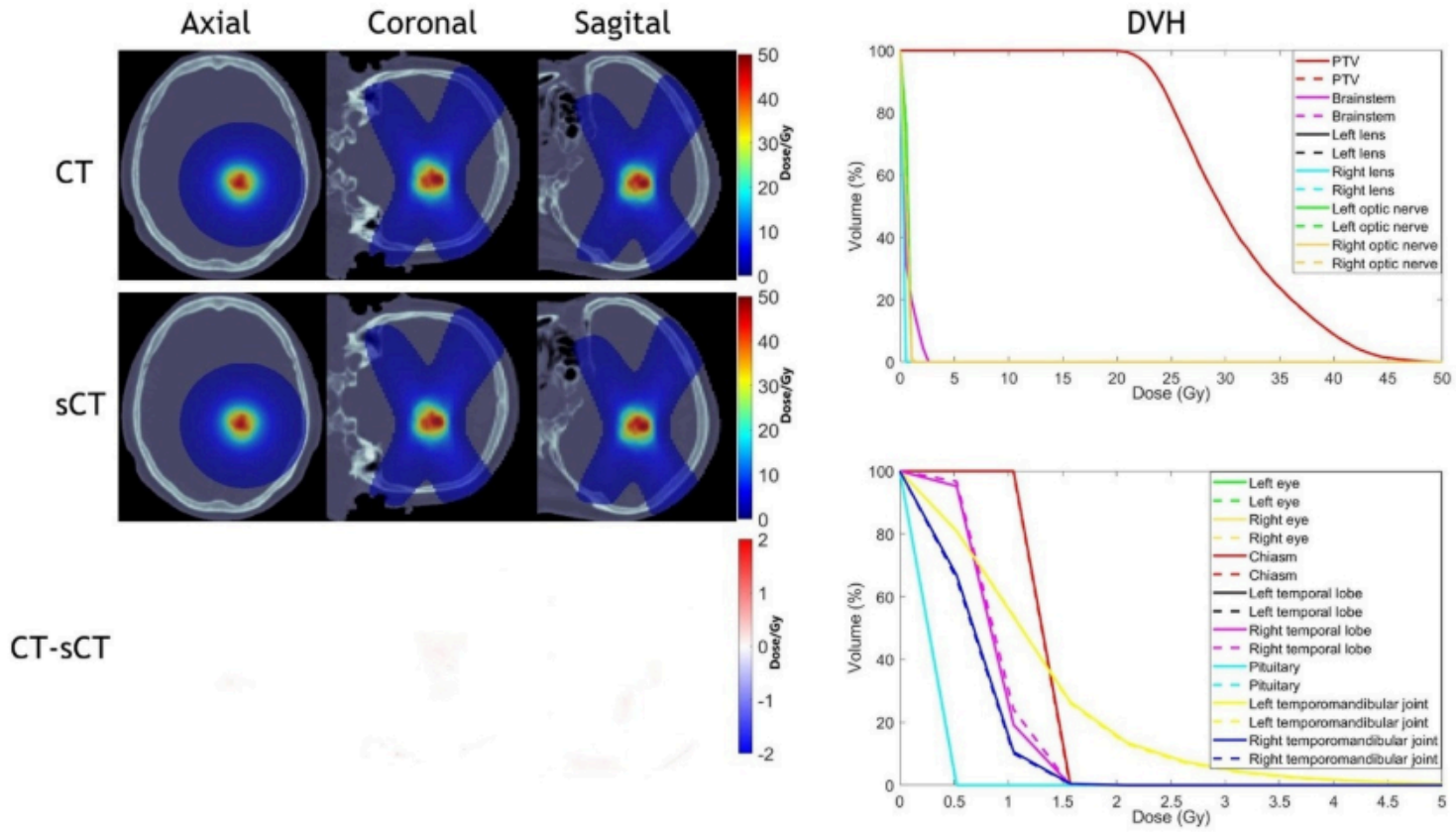


Figure 4

Dose and dose difference distribution calculated on CT, sCT in the axial, coronal and sagittal planes. The DVH curves of CT (the solid line) and sCT (the dashed line) were also given. The sCT was achieved by the large training sample model.