Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study

Chandrakanth Jayachandran Preetha, Hagen Meredig, Gianluca Brugnara, Mustafa A Mahmutoglu, Martha Foltyn, Fabian Isensee, Tobias Kessler, Irada Pflüger, Marianne Schell, Ulf Neuberger, et al.

Deep-learning-based synthesis of post-contrast T1-weighted MRI for tumour response assessment in neuro-oncology: a multicentre, retrospective cohort study

Chandrakanth Jayachandran Preetha, Hagen Meredig*, Gianluca Brugnara, Mustafa A Mahmutoglu, Martha Foltyn, Fabian Isensee, Tobias Kessler, Irada Pflüger, Marianne Schell, Ulf Neuberger, Jens Petersen, Antje Wick, Sabine Heiland, Jürgen Debus, Michael Platten, Ahmed Idbaih, Alba A Brandes, Frank Winkler, Martin J van den Bent, Burt Nabors, Roger Stupp, Klaus H Maier-Hein, Thierry Gorlia, Jörg-Christian Tonn, Michael Weller, Wolfgang Wick, Martin Bendszus, Philipp Vollmuth*

Summary

Background Gadolinium-based contrast agents (GBCAs) are widely used to enhance tissue contrast during MRI scans and play a crucial role in the management of patients with cancer. However, studies have shown gadolinium deposition in the brain after repeated GBCA administration with yet unknown clinical significance. We aimed to assess the feasibility and diagnostic value of synthetic post-contrast T1-weighted MRI generated from pre-contrast MRI sequences through deep convolutional neural networks (dCNN) for tumour response assessment in neuro-oncology.

Methods In this multicentre, retrospective cohort study, we used MRI examinations to train and validate a dCNN for synthesising post-contrast T1-weighted sequences from pre-contrast T1-weighted, T2-weighted, and fluid-attenuated inversion recovery sequences. We used MRI scans with availability of these sequences from 775 patients with glioblastoma treated at Heidelberg University Hospital, Heidelberg, Germany (775 MRI examinations); 260 patients who participated in the phase 2 CORE trial (1083 MRI examinations, 59 institutions); and 505 patients who participated in the phase 3 CENTRIC trial (3147 MRI examinations, 149 institutions). Separate training runs to rank the importance of individual sequences and (for a subset) diffusion-weighted imaging were conducted. Independent testing was performed on MRI data from the phase 2 and phase 3 EORTC-26101 trial (521 patients, 1924 MRI examinations, 32 institutions). The similarity between synthetic and true contrast enhancement on post-contrast T1-weighted MRI was quantified using the structural similarity index measure (SSIM). Automated tumour segmentation and volumetric tumour response assessment based on synthetic versus true post-contrast T1-weighted sequences was performed in the EORTC-26101 trial and agreement was assessed with Kaplan-Meier plots.

Findings The median SSIM score for predicting contrast enhancement on synthetic post-contrast T1-weighted sequences in the EORTC-26101 test set was 0·818 (95% CI 0·817–0·820). Segmentation of the contrast-enhancing tumour from synthetic post-contrast T1-weighted sequences yielded a median tumour volume of 6·31 cm³ (5·60 to 7·14), thereby underestimating the true tumour volume by a median of –0·48 cm³ (–0·37 to –0·76) with the concordance correlation coefficient suggesting a strong linear association between tumour volumes derived from synthetic versus true post-contrast T1-weighted sequences (0·782, 0·751–0·807, p<0·0001). Volumetric tumour response assessment in the EORTC-26101 trial showed a median time to progression of 4·2 months (95% CI 4·1–5·2) with synthetic post-contrast T1-weighted and 4·3 months (4·1–5·5) with true post-contrast T1-weighted sequences (p=0·33). The strength of the association between the time to progression as a surrogate endpoint for predicting the patients' overall survival in the EORTC-26101 cohort was similar when derived from synthetic post-contrast T1-weighted sequences (hazard ratio of 1·749, 95% CI 1·282–2·387, p=0·0004) and model C-index (0·667, 0·622–0·708) versus true post-contrast T1-weighted MRI (1·799, 95% CI 1·314–2·464, p=0·0003) and model C-index (0·673, 95% CI 0·626–0·711).

Interpretation Generating synthetic post-contrast T1-weighted MRI from pre-contrast MRI using dCNN is feasible and quantification of the contrast-enhancing tumour burden from synthetic post-contrast T1-weighted MRI allows assessment of the patient's response to treatment with no significant difference by comparison with true post-contrast T1-weighted sequences with administration of GBCAs. This finding could guide the application of dCNN in radiology to potentially reduce the necessity of GBCA administration.

Funding Deutsche Forschungsgemeinschaft.

Lancet Digit Health 2021

Published **Online** October 20, 2021 https://doi.org/10.1016/S2589-7500(21)00205-3

See **Online/Comment** https://doi.org/10.1016/S2589-7500(21)00219-3

*Contributed equally

Department of Neuroradiology (C Jayachandran Preetha MSc, H Meredig MD, G Brugnara MD, M A Mahmutoglu MD, M Foltyn MD, I Pflüger MD, M Schell MD, U Neuberger MD, Prof S Heiland PhD, Prof M Bendszus MD, P Vollmuth MD), **Neurology Clinic** (T Kessler MD, A Wick MD, Prof F Winkler MD, Prof W Wick MD), **and Department of Radiation Oncology** (Prof J Debus MD, Prof Klaus H Maier-Hein PhD), **Heidelberg University Hospital, Heidelberg, Germany; Medical Image Computing** (F Isensee PhD, J Petersen PhD, Prof K H Maier-Hein), **Clinical Cooperation Unit Neurooncology** (T Kessler, Prof F Winkler, Prof W Wick), **and Clinical Cooperation Unit Neuroimmunology and Brain Tumor Immunology** (Prof M Platten MD), **German Cancer Research Center, Heidelberg, Germany; Heidelberg Institute of Radiation Oncology, Heidelberg, Germany** (Prof J Debus); **Heidelberg Ion-Beam Therapy Center, Heidelberg, Germany** (Prof J Debus); **Department of Neurology, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany** (Prof M Platten); **Sorbonne Université, Inserm, Institut du Cerveau, Assistance Publique-Hôpitaux de Paris, Hôpitaux Universitaires La Pitié Salpêtrière—Charles Foix, Service de Neurologie**

Introduction

Intravenous administration of gadolinium-based contrast agents (GBCAs) is a standard procedure in diagnostic MRI examinations, in which it allows for the identification of pathological tissue changes through the extravasation of contrast media.[1] In general, GBCAs present a very good

2-Mazarin, Paris, France (A Idbaih MD); **Department of Medical Oncology, Azienda USL of Bologna, Bologna, Italy** (Prof A A Brandes MD); **Brain Tumor Center, Erasmus MC Cancer Institute, Rotterdam, Netherlands** (Prof Martin J van den Bent MD); **Department of Neurology and O'Neal Comprehensive Cancer Center, Division of Neuro-Oncology, University of Alabama at Birmingham, Birmingham, AL, USA** (Prof B Nabors MD); **Malnati Brain Tumor Institute of the Lurie Comprehensive Cancer Center, Department of Neurological Surgery and Department of Neurology, Northwestern Medicine and Northwestern University, Chicago, IL, USA** (Prof R Stupp MD); **European Organisation for Research and Treatment of Cancer, Brussels, Belgium** (T Gorlia PhD); **Department of Neurosurgery, Ludwig-Maximilians-University, Munich, Germany** (Prof J-C Tonn MD); **Department of Neurology, University Hospital and University of Zurich, Zurich,** Switzerland (Prof M Weller MD)

Correspondence to:
Dr Philipp Vollmuth (né Kickingereder), Department of Neuroradiology, Heidelberg University Hospital, 69120 Heidelberg, Germany
philipp.vollmuth@med.uni-heidelberg.de

Research in context

Evidence before this study

Gadolinium-based contrast agents (GBCAs) are widely used to enhance tissue contrast during MRI and are crucial in the management of patients with cancer, enabling accurate identification, characterisation, and staging of tumours. Although GBCAs have a very good safety profile, previous studies have showed gadolinium deposition in the brain after repeated GBCA administration with yet unknown clinical significance. Based on this evidence, the European Medicines Agency restricted the use of some linear GBCAs used in MRI and suspended the authorisation of others in 2017. Studies have explored potential alternatives to reduce or bypass the need of GBCA administration during MRI. In this study, we developed deep convolutional neural networks (dCNN) for generating synthetic post-contrast T1-weighted MRI from the information available within pre-contrast MRI sequences—ie, without administration of GBCAs. We assessed the potential clinical use of synthetic post-contrast T1-weighted MRI in oncological imaging by training, validating, and testing the dCNN on MRI data from three clinical trials in neuro-oncology and institutional data with more than 2000 patients from over 200 institutions and evaluated the diagnostic value of synthetic post-contrast T1-weighted MRI for tumour response assessment as an initial use case. We searched PubMed on Feb 20, 2021, with no date or language restrictions on publications, with the search terms ("synthetic" OR "virtual") AND ("contrast") AND ("oncology" OR "oncological"). Our search did not identify articles that evaluated the diagnostic value of synthetic post-contrast T1-weighted MRI for tumour response assessment in oncology.

Added value of this study

We show the feasibility and clinical potential of dCNN for generating synthetic post-contrast T1-weighted MRI sequences from pre-contrast MRI sequences in neuro-oncology without administration of GBCAs. Specifically, the use of synthetic post-contrast T1-weighted sequences for volumetric quantification of the contrast-enhancing tumour burden and assessment of the treatment response in independent large-scale test sets yielded no significant difference in the time to progression as compared with using true post-contrast T1-weighted sequences with administration of GBCAs. Moreover, the strength of assocation between the time to progression as surrogate endpoint for predicting the patients' overall survival was similar when derived from synthetic versus true post-contrast T1-weighted sequences.

Implications of all the available evidence

Our results show that synthetic post-contrast T1-weighted MRI sequences generated from pre-contrast MRI-sequences using dCNN is feasible and quantification of the contrast-enhancing tumour burden from synthetic post-contrast T1-weighted MRI allows assessment of the patient's response to treatment with no significant difference as compared with using true post-contrast T1-weighted sequences with administration of GBCAs. This approach could serve as a blueprint for application of dCNN in radiology to potentially reduce the necessity of GBCA administration.

safety profile, with mild adverse reactions in only 0·7–2·4% of cases and a very low rate of severe complications in 0·03% of cases,[2] coupled with a low risk of nephrogenic systemic fibrosis even for patients with relatively limited renal functionality.[3] However, studies have shown gadolinium deposition in the brain after repeated GBCA administration,[4] which has sparked debates on the safety profile of contrast agents.[5] Consequently, the potential risks of GBCAs must be weighed against the clinical benefit and the diagnostic information that GBCAs can provide for each individual patient. Previous research efforts have explored potential alternatives to reduce or bypass the need of GBCA administration during MRI. Specifically, for brain MRI, preliminary studies have shown that the dose of GBCA can be reduced 10-fold if artificial neural networks are used to synthesise full-dose post-contrast MRI sequences and preserve the contrast information that would have otherwise only been available with the full-dose GBCA administration.[6] Hypothesis generating studies have also explored the potential of artificial neural networks for synthesising post-contrast MRI sequences from pre-contrast MRI sequences alone, thereby potentially bypassing the need of GBCA for particular applications such as brain tumour imaging.[7,8] Despite this interesting concept of synthesising post-contrast MRI sequences from pre-contrast MRI sequences, there are currently no independent large-scale studies on heterogeneous multi-institutional datasets and assessment of its diagnostic value for clinical decision making.

In this study, we used MRI data from two phase 2 and two phase 3 clinical trials in neuro-oncology[9–12] alongside retrospective institutional data with more than 2000 patients from over 200 institutions to robustly assess the potential of artificial neural networks for synthesising post-contrast MRI sequences from pre-contrast MRI sequences. Specifically, we compared two popular artificial neural network architectures, namely U-Net convolutional neural networks (U-Net)[13] and conditional general adversarial neural networks (CGAN)[14] for generating synthetic post-contrast MRI and assess its diagnostic performance for tumour response assessment in neuro-oncology.

Methods

Study design and participants

In this multicentre, retrospective cohort study, we analysed MRI data from patients with glioblastoma acquired from four cohorts, one cohort from Heidelberg

University Hospital (Heidelberg cohort; Heidelberg, Germany) and three cohorts from the following clinical trials: the multicentre phase 2 CORE trial (CORE cohort; NCT00813943),[11] the multicentre phase 3 CENTRIC trial (CENTRIC cohort; NCT00689221),[12] and the multicentre phase 2 and phase 3 EORTC 26101 trial (EORTC-26101 cohort; NCT01290939; appendix p 2; figure 1).[9,10] MRI examinations for each cohort included T1-weighted images before and after contrast agent administration, fluid-attenuated inversion recovery (FLAIR) and T2-weighted images (all either acquired three dimensional [3D] or with axial orientation) and for a subset of examinations also diffusion-weighted imaging with apparent diffusion coefficient (ADC) maps. Retrospective evaluation of the Heidelberg cohort was approved by the local ethics committee of the University of Heidelberg and informed consent was waived (reference S-784 2018); evaluation of the CENTRIC, CORE, and EORTC-26101 cohorts was granted through an external research project (reference ERP-263 and ERP-362) with the European Organisation for Research and Treatment of Cancer.

The Heidelberg cohort consisted of 775 MRI examinations from a single timepoint from 775 patients (ie, one MRI examination per patient) either preoperatively from initial diagnosis (n=580 [75%]), early postoperatively (<72 h after surgery; n=57 [7%]), or at disease recurrence (n=138 [18%]) in patients with histologically confirmed glioblastoma. For a subset of 358 (46%) MRI examinations, ADC maps were included besides pre-contrast anatomical images.

The CORE[11] and CENTRIC[12] studies were prospective multicentre randomised phase 2 and 3 trials in patients with newly diagnosed glioblastoma comparing standard chemoradiotherapy with or without anti-angiogenic treatment with cilengitide. MRI examinations were acquired at baseline before radiotherapy (ie, postoperatively), 4 weeks after radiotherapy, and at 18, 26, and 34 weeks and every 12 weeks thereafter during the follow-up phase.

The EORTC-26101 study[9,10] was a prospective multicentre randomised phase 2 and 3 trial in patients with first progression of a glioblastoma after standard chemoradiotherapy evaluating the optimal treatment sequence of bevacizumab and lomustine. MRI examinations were acquired at baseline and every 6 weeks until week 24, and every 3 months afterwards.

Overall, 260 patients with 1083 MRI examinations from 58 institutions from the CORE study, 505 patients with 3147 MRI examinations from 139 institutions from the CENTRIC study, and 521 patients with 1924 MRI examinations from 32 institutions from the EORTC-26101 study had complete availability of anatomical MRI sequences (T1-weighted images before and after contrast agent administration, FLAIR and T2-weighted images) following image pre-processing and were included in this analysis (figure 1). A detailed description of the inclusion and exclusion criteria of the MRI data are outlined in the appendix (p 2). A subset of MRI examinations in the CORE cohort (137 [12%] of 1083; 52 patients from 18 institutions), the CENTRIC cohort (548 [17%] of 3147; 170 patients from 66 institutions), and in the EORTC-26101 cohort (1685 [87%] of 1924;

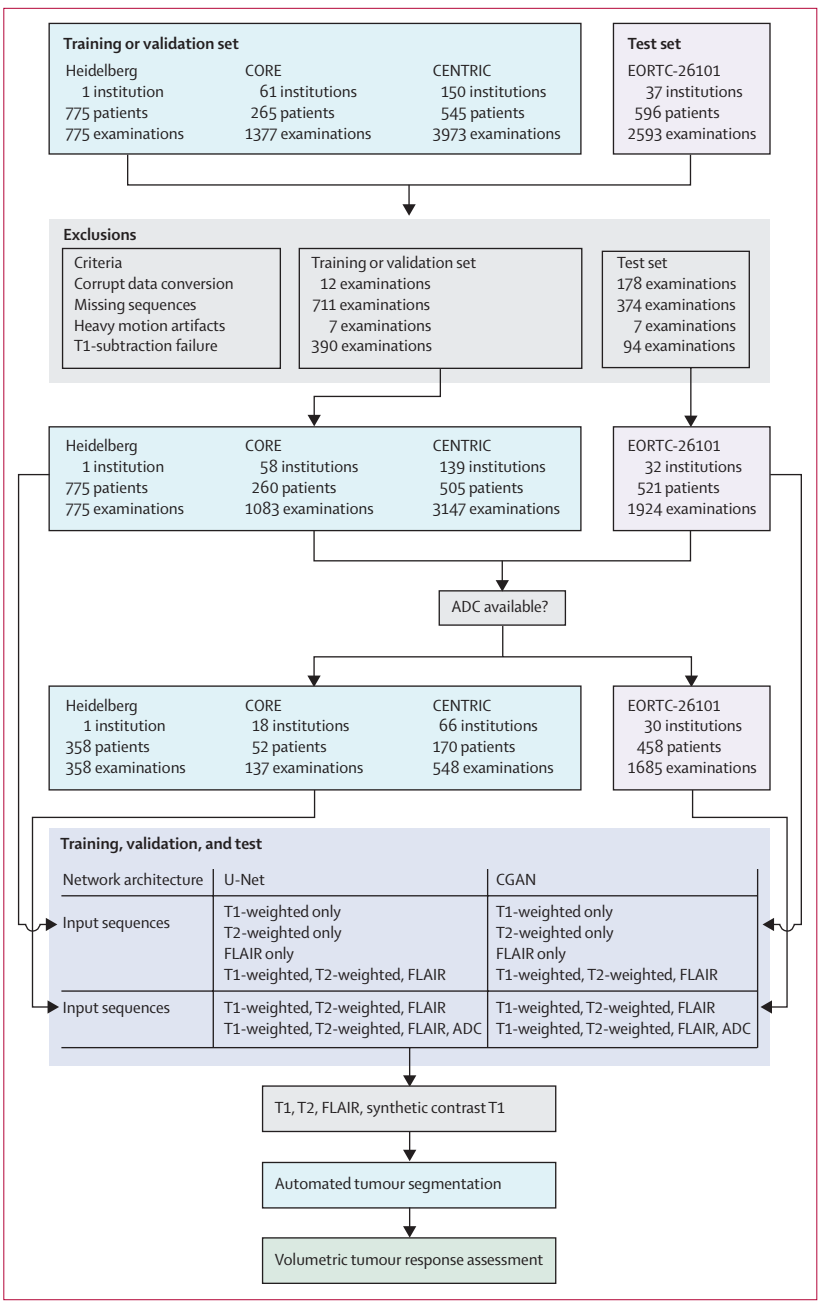

Figure 1: **Flowchart of the procedures for training, validation, independent testing, and statistical analysis for generating synthetic contrast T1-weighted MRI using deep convolutional neural networks**
ADC=apparent diffusion coefficient. CENTRIC=Cilengitide, Temozolomide, and Radiation Therapy in Treating Patients With Newly Diagnosed Glioblastoma and Methylated Gene Promoter Status.[12] CGAN=conditional general adversarial neural networks. CORE=Conventional Care Versus Radioablation (Stereotactic Body Radiotherapy) for Extracranial Oligometastases.[11] EORTC=European Organisation for Research and Treatment of Cancer.[9,10] FLAIR=fluid-attenuated inversion recovery.

See **Online** for appendix

458 patients from 30 institutions) had ADC maps and anatomical sequences.

Cohort size for each of the included datasets in this retrospective study was determined by availability of samples and not derived from a power calculation.

Procedures

A flowchart outlining the analysis workflow is shown in figure 1 and complies with the Checklist for Artificial Intelligence in Medical Imaging criteria.[15] The same image pre-processing steps (appendix p 3) were applied to all MRI examinations from each cohort. Specifically, processing included automated deep-learning-based brain extraction, followed by image registration as described in previous studies.[16,17] All MRI sequences were resampled to an isotropic spacing of 1×1×1 mm³ and independently normalised to zero mean and unit variance. Subsequently, outliers (outside 99th percentile of intensity values) were clipped and image intensities were rescaled to the range (0–1). T1-subtraction volumes were generated by voxel-wise subtraction of the T1-weighted volume from the post-contrast T1-weighted volume.

The study explored two different deep-learning approaches for the generation of synthetic post-contrast T1-weighted imaging (appendix pp 3–4). The first method involved the use of a 3D CNN based on the U-Net architecture.[13] The U-Net model consists of an encoder and a decoder with skip connections between the two sections. The convolutional layers in each block are followed by batch normalisation, dropout, and rectified linear unit activation function. In the encoder section, the network aggregates semantic information at the cost of reduced spatial information. The decoder upsamples the feature maps to recover spatial information. Skip connections enable the decoder to make use of high-resolution features from the contracting path. A combination of mean absolute error (MAE) and the structural similarity index measure (SSIM) loss was used to train the model (appendix p 14).

The second method used a CGAN-based method (appendix p 7) inspired by pix2pix.[14] The CGAN learns a generative model using paired images from source and target domains. The CGAN model consists of a generator and a discriminator. The generator transforms the input images from the source domain to the target domain while the discriminator takes an input image and an unknown image (either generated by the generator or a real image) and attempts to predict if it was produced by the generator. During the discriminator training phase, the generator is kept constant and the weights of the discriminator are updated through backpropagation using the discriminator loss. Similarly, while training the generator, the weights of the discriminator are kept constant. However, the loss from discriminator classification is used to update the weights of the generator network. Additionally, hand-picked loss terms (MAE and SSIM loss) were included to encourage the generator to produce output images that are structurally similar to the target image (appendix p 14).

Image patches (size 160×176×32; channels x, y, z) were fed to both networks during the training process. The pre-contrast anatomical sequences from each MRI examination (T1-weighted, FLAIR, T2-weighted) were used as the input to the models and the corresponding T1-subtraction sequence was used as the ground truth. The output from the models (ie, synthetic T1-subtraction sequence) was added to the corresponding T1-weighted images to generate synthetic post-contrast T1-weighted images. The Heidelberg cohort, and the CENTRIC and CORE cohorts, were allocated for training and validation of the U-Net and CGAN models (with 5-fold cross-validation and with splitting of the samples performed on a patient level to avoid data leakage). The performance of the models was then independently tested on the EORTC-26101 cohort using SSIM as the evaluation metric. Training and evaluation of the models was also performed using individual anatomical sequences (either T1-weighted, FLAIR, or T2-weighted) to determine their effect on synthetic contrast map generation. Moreover, to assess whether the addition of ADC maps increases the performance for generating synthetic post-contrast T1-weighted sequences, both U-Net and CGAN models were also trained and evaluated with and without ADC images as an input alongside pre-contrast anatomical sequences (T1-weighted, FLAIR, and T2-weighted) for the subset of MRI examinations from each cohort with availability of ADC images. Sensitivity of the trained models to the type of MRI image acquisition (two dimensional [2D] *vs* 3D) was studied for the different MRI sequences.

Generation of synthetic post-contrast T1-weighted sequences for the EORTC-26101 test dataset was done using an ensemble approach that combines the output from the networks (by averaging the predictions) trained on different data splits (5-fold cross-validation splits). To determine the agreement in the segmentation of the contrast-enhancing tumour based on synthetic versus true post-contrast T1-weighted sequence, automated tumour segmentation was performed on the EORTC-26101 test set as described previously using a trained 3D U-Net convolutional neural network model.[16] Therefore, volumetric segmentation masks with delineation of the contrast-enhancing tumour and the non-enhancing T2-FLAIR hyperintense abnormality excluding the contrast-enhancing and necrotic portion of the tumour, resection cavity, and obvious leukoaraiosis were generated. The tumour segmentation process for each examination was performed twice: once using synthetic post-contrast T1-weighted sequences and once using true post-contrast T1-weighted sequences alongside pre-contrast T1-weighted, T2-weighted, and FLAIR sequences as an input.

In the EORTC-26101 test set, volumetric tumour response assessment was performed as described

previously.[16] Specifically, calculation of the time to progression was performed by analysing the longitudinal change in the contrast-enhancing tumour volumes for each patient using synthetic versus true post-contrast T1-weighted sequences. The time to progression was calculated from the date of randomisation and censored at the date of last MRI if no progression occurred during follow-up. Progression was defined as an increase in the contrast-enhancing tumour volume by 40% as compared with baseline or best response, which corresponds to an increase of the area of contrast-enhancing tumour by 25% determined by bi-perpendicular tumour diameters as defined by response assessment in neuro-oncology (RANO) criteria (assuming spherical configuration).[18] Moreover, measurable lesions were defined as more than 0·524 cm³, thus equalling the 1×1 cm cutoff (ie, 1 cm diameter of a sphere) as defined by RANO.[18] Additionally, a separate time to progression was calculated by incorporating the longitudinal change in the T2-FLAIR signal abnormality volume besides the change in the contrast-enhancing tumour volume as described previously.[16]

Outcomes

The present study had four main objectives. First, to evaluate the structural similarity of synthetic post-contrast T1-weighted sequences generated from pre-contrast MRI sequences using dCNN with the corresponding true post-contrast T1 sequences. Second, to assess the agreement in contrast-enhancing tumour segmentations and volumes when using synthetic versus true post-contrast T1-weighted sequences in the EORTC-26101 test set. Third, to assess the time to progression when using synthetic versus true post-contrast T1-weighted sequences for tumour response assessment in the EORTC-26101 test set. Fourth, to compare time to progression when using synthetic versus true post-contrast T1-weighted sequences for tumour response assessment as surrogate endpoint for predicting overall survival within the EORTC-26101 test set (information on overall survival taken from the EORTC-26101 trial database; and calculated from the date of randomisation until death or last follow-up).

Statistical analysis

Comparison of SSIM scores (which are a combination of luminance, contrast, and structural comparison functions)[19] between the different models in the validation and test set was performed using the Wilcoxon matched pairs signed-ranks test. Specifically, the SSIM score was derived by comparison of the contrast enhancement on synthetic and true T1-weighted subtraction sequences and quantified the similarity on a scale of 0 (no similarity) to 1 (perfect similarity). The reported 95% CIs were calculated using bootstrapping (with n=1000 iterations) with the bias-adjusted and accelerated bootstrap method. A sensitivity analysis, comparing SSIM scores in the EORTC-26101 test set between different image acquisition protocols (2D *vs* 3D) was performed using the Wilcoxon rank sum test. The agreement between the contrast-enhancing tumour when using synthetic versus true post-contrast T1-weighted sequences was evaluated with the Sørensen–Dice similarity coefficient for segmentation agreement and Bland–Altman plots and concordance correlation coefficients for volume agreement. The reported 95% CIs for the median Sørensen–Dice and concordance correlation coefficients were calculated using bootstrapping (with n=1000 iterations) with the bias-corrected and accelerated bootstrap method. Kaplan-Meier plots and log-rank tests were generated to assess the agreement in the time to progression on a group level when using synthetic versus true post-contrast T1-weighted sequences. The strength of association of the time to progression derived from synthetic versus true post-contrast T1-weighted MRI as surrogate endpoints for predicting overall survival was assessed using time-dependent Cox proportional hazards regression models for overall survival, as described previously.[16] Specifically, we generated Cox proportional hazards regression models with the time to progression (either derived from synthetic *vs* true post-contrast T1-weighted MRI) included as a time-dependent covariate alongside baseline epidemiological (patients' age), clinical (glucocorticoid intake, WHO performance status), molecular characteristics (*MGMT* promoter methylation status), and the treatment regimen[20] (containing bevacizumab *vs* no bevacizumab). We assessed the performance of each Cox proportional hazards regression model with Harrell's concordance index (c-index, with 95% CIs calculated using bootstrapping with n=1000 iterations).

Statistical analysis was performed with R (version 4.0.3) and Python (3.8.3) by PV and CJP.

Role of the funding source

The funders of this study or the three clinical trials from which the data were obtained had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

The compiled MRI data from patients with glioblastoma being treated at Heidelberg University Hospital (775 patients with one MRI examination each) and within the multi-institutional longitudinal phase 2 CORE trial

	Validation SSIM (95% CI)	Test SSIM (95% CI)
CGAN	0·827 (0·826–0·829)	0·818 (0·817–0·820)
U-Net	0·809 (0·808–0·811)	0·809 (0·807–0·810)
p value	<0·0001	<0·0001

p values are from a Wilcoxon signed-rank test. CGAN=conditional general adversarial neural networks. SSIM=structural similarity index measure.

Table 1: **Comparison of the SSIM between U-Net and CGAN architecture on validation and test data**

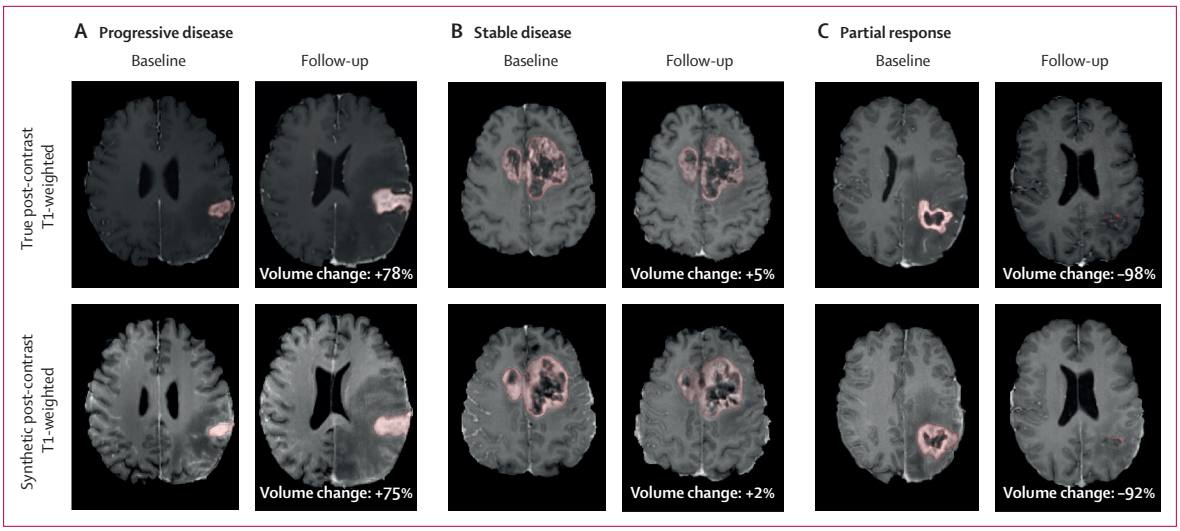

Figure 2: **True and synthetic post-contrast T1-weighted sequences**
True post-contrast T1-weighted sequences (top) and synthetic post-contrast T1-weighted sequences (bottom) from two timepoints (baseline and follow-up MRI) in three illustrative cases showing increase in tumour burden. (A) Progressive disease. (B) Stable disease or decrease in tumour burden. (C) Partial response. Within each MRI examination the automatically generated contrast-enhancing tumour segmentation map is shown in red, with the percentual change in the contrast-enhancing tumour volume shown at the bottom of each follow-up examination.

(206 patients with 1083 MRI examinations from 59 institutions) and phase 3 CENTRIC trial (505 patients with 3147 MRI examinations from 149 institutions) were allocated for training and validation of the dCNN models for generating synthetic post-contrast T1-weighted images from pre-contrast T1-weighted, T2-weighted, and FLAIR images (figure 1). Independent testing of the dCNN models on the multi-institutional longitudinal phase 2 and phase 3 EORTC-26101 trial (521 patients with 1924 MRI examinations from 32 institutions) yielded a median SSIM score of 0·809 (95% CI 0·807–0·810) for predicting contrast enhancement on synthetic post-contrast T1-weighted MRI when developing the dCNN based on the U-Net architecture. The performance of the dCNN improved significantly (p<0·0001) when using a CGAN architecture, yielding a median SSIM score of 0·818 (0·817–0·820) for predicting contrast enhancement on synthetic post-contrast T1-weighted MRI (table 1). Examples of the training epochs and the predicted contrast maps along with the corresponding ground truth and residual images are shown in figure 2 and the appendix (pp 8–9). For the CGAN architecture, pre-contrast T1-weighted, followed by FLAIR and T2-weighted sequences were ranked as most influential for predicting synthetic post-contrast T1-weighted sequences with a median SSIM of 0·811 (0·810–0·813) for T1-weighted, 0·746 (0·741–0·744) for FLAIR, and 0·742 (0·741–0·744) for T2-weighted sequences on the EORTC-26101 test set (appendix p 15). The SSIM for predicting synthetic post-contrast T1-weighted sequences from these individual sequences was significantly lower than the SSIM when using a combination of pre-contrast T1-weighted, FLAIR, and T2-weighted sequences (p<0·0001 each; appendix p 15). The subset analysis for those examinations with availability of ADC images showed no improvement in the SSIM when comparing the inclusion of ADC in addition to pre-contrast anatomical MRI (T1-weighted, FLAIR, and T2-weighted) in the EORTC-26101 test set (p=0·73; appendix p 16). Sensitivity analysis in the EORTC-26101 test set showed that high-resolution 3D image acquisition improved the performance of the CGAN model for predicting synthetic post-contrast T1-weighted MRI, yielding significantly higher SSIM scores as compared with 2D image acquisition (appendix p 17).

Automated segmentation of the contrast-enhancing tumour in the EORTC-26101 test set from synthetic post-contrast T1-weighted sequences yielded a median tumour volume of 6·31 cm³ (95% CI 5·60 to 7·11), thereby underestimating the tumour volume by a median of –0·48 cm³ (–0·37 to –0·76; figures 2, 3A), corresponding to a median underestimation of –7% as compared with true post-contrast T1-weighted sequences (median tumour volume of 6·74 cm³, 6·13 to 7·75). The concordance correlation coefficients, which measures tumour volume agreement, suggested a strong linear relationship (0·782, 0·751 to 0·807, p<0·0001) between tumour volumes derived from synthetic versus true post-contrast T1-weighted sequences (figure 3A; appendix p 10). Spatial agreement in the tumour segmentation when using synthetic versus true post-contrast T1-weighted sequences (median Sørensen–Dice coefficient of 0·28, 0·26 to 0·31]) significantly depended on the volume of the tumour with improved spatial agreement for larger tumours (r=0·438, 0·401 to 0·475, p<0·0001; figure 3B).

Volumetric tumour response assessment in a subset of the EORTC-26101 data with availability of baseline and

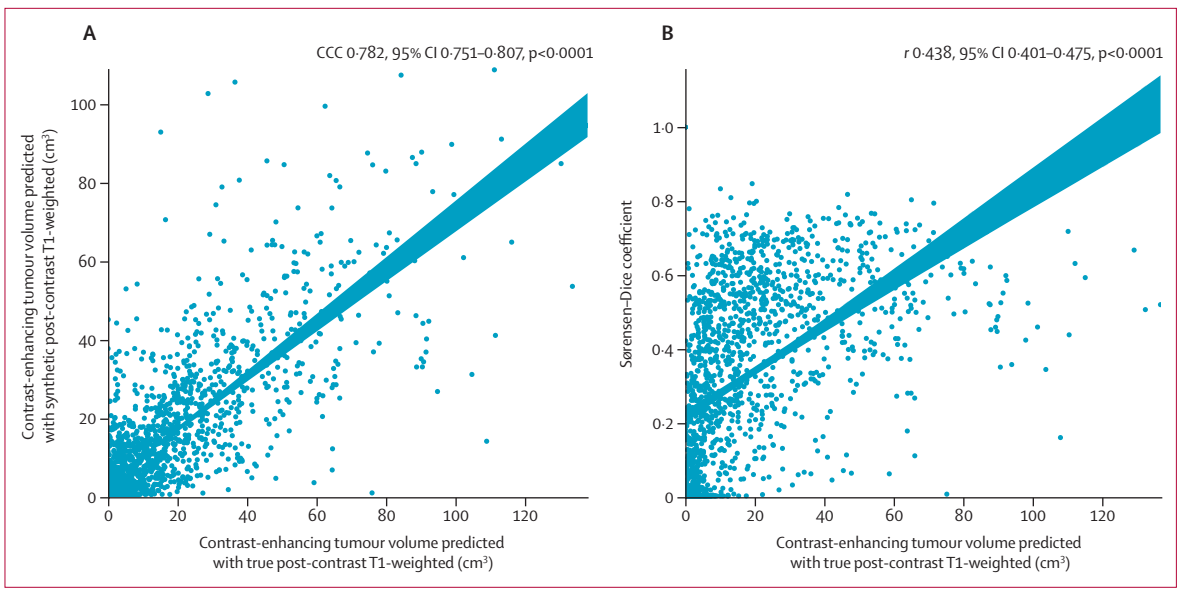

Figure 3: **Contrast-enhancing tumour volumes segmented using true and synthetic post-contrast T1-weighted sequences and Sørensen–Dice coefficients**
(A) Contrast-enhancing tumour volumes segmented using synthetic and true post-contrast T1-weighted sequences. (B) Sørensen–Dice coefficients (spatial agreement in contrast-enhancing tumour segmentation between synthetic and true post-contrast T1-weighted sequences) and contrast-enhancing tumour volumes obtained from true post-contrast T1-weighted sequences. CCC=concordance correlation coefficient.

subsequent follow-up MRI examinations (386 [74%] of 521 patients) showed a median time to progression of 4·2 months (95% CI 4·1–5·2) with synthetic post-contrast T1-weighted and 4·3 months (4·1–5·5) with true post-contrast T1-weighted sequences (p=0·33; figure 4). For the subset of patients with disagreement in the time to progression (either time or censoring status) between synthetic and true post-contrast T1-weighted sequences (181 [47%] of 386 patients) the median time to progression was 4·2 months (4·1–5·5) with synthetic post-contrast T1-weighted and 5·4 months (4·2–5·6) with true post-contrast T1-weighted sequences (p=0·14; figure 4). Specifically, the use of synthetic post-contrast T1-weighted sequences did not detect tumour progression in 53 (14%) of 386 cases that was otherwise evident when using true post-contrast T1-weighted sequences. On the contrary, the use of synthetic post-contrast T1-weighted sequences detected tumour progression in 59 (15%) of 386 cases that was otherwise not evident when using true post-contrast T1-weighted sequences.

By incorporating the longitudinal change in the volume of the T2-FLAIR signal abnormality as an additional response criterion besides the change in the volume of the contrast-enhancing tumour, the agreement in the time to progression between using synthetic or true post-contrast T1-weighted MRI could be improved. Specifically, median time to progression was 4·1 months (4·0–4·4) with synthetic post-contrast T1-weighted and 4·1 months (3·7–4·2) with true post-contrast T1-weighted sequences (p=0·24; appendix p 11). For the subset of patients with disagreement in the time to progression (either time or censoring status) between synthetic and

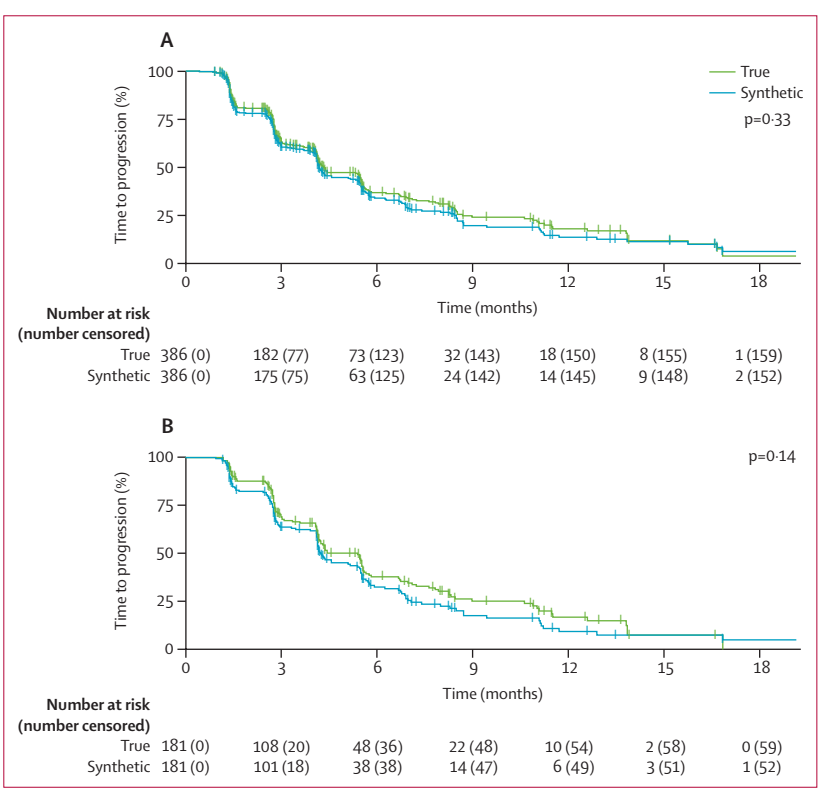

Figure 4: **Time to progression in the EORTC-26101 test set**
Agreement in the time to progression in the EORTC-26101 test set when performing volumetric tumour response assessment using synthetic post-contrast T1-weighted sequences (blue line) and true post-contrast T1-weighted sequences (green line). (A) All patients. (B) The subset of patients with disagreement in the time to progression (either time or censoring status) between synthetic and true post-contrast T1-weighted sequences.

	Point estimate (95% CI)	Z value*	p value
Synthetic post-contrast T1-weighted MRI sequences			
Volumetric time to progression[1]	HR 1·749 (1·282–2·387)	3·529	0·0004
Baseline characteristics			
Treatment regimen (BEV yes or no)[2]	HR 1·068 (0·791–1·441)	0·433	0·67
Glucocorticoid intake (yes or no)[3]	HR 1·710 (0·860–1·361)	3·470	0·00052
WHO Performance Status (>0 or 0)[4]	HR 1·362 (1·263–2·316)	1·864	0·062
Age (years)[5]	HR 1·007 (0·991–1·021)	0·866	0·39
MGMT promoter methylation status (methylated or unmethylated)[6]	HR 0·681 (0·512–0·907)	–2·633	0·0085
C-index of the model	0·667 (0·622–0·708)	..	..
True post-contrast T1-weighted MRI sequences			
Volumetric time to progression†	HR 1·799 (1·314–2·464)	3·665	0·0003
Baseline characteristics			
Treatment regimen (BEV yes or no)[2]	HR 1·016 (0·874–1·388)	0·106	0·92
Glucocorticoid intake (yes or no)[3]	HR 1·081 (1·331–2·448)	3·800	0·00015
WHO Performance Status (>0 or 0)[4]	HR 1·433 (1·033–1·989)	2·152	0·031
Age (years)[5]	HR 1·006 (0·991–1·021)	0·800	0·42
MGMT promoter methylation status (methylated or unmethylated)[6]	HR 0·714 (0·535–0·954)	–2·279	0·023
C-index of the model	0·673 (0·626–0·711)	..	..

The time to progression based on synthetic post-contrast T1-weighted MRI sequences versus true post-contrast T1-weighted MRI sequences was included as a time-dependent covariate to assess the strength of association with overall survival. BEV=bevacizumab. C-index=Harrell's concordance index. HR=hazard ratio. *Z value is the ratio of each regression coefficient to its SE. 1=time to progression is included as a time-dependent covariate; 2=included as binary covariate (initial treatment containing bevacizumab [n=242] *vs* no bevacizumab [n=144]); 3=included as binary covariate (glucocorticoid intake at baseline yes [n=182] *vs* no [n=204]); 4=included as binary covariate (WHO Performance Status at baseline >0 [n=255] *vs* 0 [n=131]); 5=included as continuous covariate (patients' age at baseline [median 59, IQR 51–64 years]); 6=included as binary covariate (*MGMT* promoter methylation status methylated [n=124] *vs* unmethylated [n=122] with information not available for the remaining [n=140] cases).

Table 2: **Cox proportional hazards regression models for overall survival in the EORTC-26101 test set**

true post-contrast T1-weighted sequences (166 [43%] of 386 patients) the median time to progression was 4·1 months (4·1–5·1) with synthetic post-contrast T1-weighted and 4·4 months (4·1–5·5) with true post-contrast T1-weighted sequences (p=0·083; appendix p 11). The use of synthetic post-contrast T1-weighted sequences did not detect tumour progression in 41 (11%) of 386 patients that was otherwise evident when using true post-contrast T1-weighted sequences. On the contrary, the use of synthetic post-contrast T1-weighted sequences detected tumour progression in 48 (12%) of 386 patients that was otherwise not evident when using true post-contrast T1-weighted sequences.

The time to progression when using synthetic or true post-contrast T1-weighted sequences for tumour response assessment as surrogate endpoints for predicting the patient's overall survival was assessed with time-dependent Cox regression models adjusted for baseline epidemiological (patients' age), clinical (glucocorticoid intake, WHO performance status), molecular characteristics (*MGMT* promoter methylation status), and the treatment regimen (table 2). Thereby, the Cox regression model for overall survival with the time to progression derived from synthetic post-contrast T1-weighted MRI as a time-dependent covariate yielded an hazard ratio (HR) of 1·749 (1·282–2·387, p=0·0004) with a Z value of 3·529 and a model c-index of 0·667 (0·622–0·708). Similar results were obtained with the time to progression derived from true post-contrast T1-weighted MRI (HR 1·799, 1·314–2·464, p=0·0003) with a Z value of 3·665 and a model c-index of 0·673 (0·626–0·711).

Discussion

In this study, we show the feasibility of dCNN for generating synthetic post-contrast T1-weighted sequences from pre-contrast MRI sequences without administration of GBCA. By incorporating MRI data from three phase 2 and phase 3 clinical trials in neuro-oncology[9–12] alongside retrospective institutional data with more than 2000 patients from over 200 institutions for training, validation, and independent testing of the dCNN we show the clinical potential of using synthetic post-contrast T1-weighted sequences for tumour response assessment. Specifically, the use of synthetic post-contrast T1-weighted sequences for automated longitudinal quantification of the contrast-enhancing tumour burden and volumetric tumour response assessment in the EORTC-26101 test set yielded no significant difference to when using true post-contrast T1-weighted sequences with a difference of only 0·1 months in the median time to progression across all patients. Moreover, the time to progression derived from synthetic versus true post-contrast T1-weighted sequences yielded equal prognostic surrogate levels for predicting overall survival.

The current consensus recommendations for a standardised brain tumour imaging protocol and the RANO criteria mandate the use of post-contrast T1-weighted MRI for evaluating the contrast-enhancing tumour burden.[20,21] Specifically, the presence and extent of contrast-enhancing tumour is—alongside the non-enhancing T2-FLAIR hyperintense tumour burden and relevant clinical parameters—a key criterion for assessing the response to treatment and for determining the time to progression in neuro-oncology clinical trials and in routine clinical practice.[21] A survey across 220 institutions in Europe underlined the central importance of post-contrast T1-weighted MRI in brain tumour imaging with more than 99% of these institutions routinely administering GBCA for acquiring post-contrast T1-weighted MRI.[22] However, previous studies have found that repeated administration of GBCA can lead to deposition of residual gadolinium in the brain.[4] Specifically, deposition of GBCA in the brain was found to be highly dependent on the chemical properties of the agent, with the administration of linear GBCAs resulting in a 15-times higher concentration of gadolinium in the brain as compared with macrocyclic GBCAs.[23] Although the clinical significance of the retained gadolinium in the brain remains unknown, the European Medicines Agency has subsequently restricted the use of some linear gadolinium agents used in MRI and suspended

the authorisation of others in 2017.[24] Moreover, authorities urged health-care professionals to use GBCA only when essential diagnostic information cannot be obtained with unenhanced MRI.[24] Previous exploratory studies have therefore focused on potential alternatives to reduce or bypass the need of GBCAs during MRI, which have included advanced MRI sequences[25] or alternative contrast agents (eg, measuring the glucose update through chemical exchange saturation transfer MRI).[26] Moreover, preliminary neuroimaging studies have also explored the technical feasibility of using dCNN for synthesising post-contrast T1-weighted MRI from the information included in pre-contrast MRI sequences.[7,8,27] In this study we went beyond the assessment of technical feasibility and specifically explored the clinical use of these synthetic post-contrast T1-weighted sequences by harnessing large-scale MRI data from several previous clinical trials in the field of neuro-oncology[9–12] alongside retrospective institutional data with more than 2000 patients from over 200 institutions. Our study used two popular dCNN architectures, namely an encoder-decoder architecture (U-Net) as a reference benchmark and a GAN architecture, which has gained substantial attention across multiple industries since its first description in 2014 by Goodfellow and colleagues[28] to generate synthetic instances of data that can pass for real data (eg, for image, video, and voice generation).[29] By applying GANs, a generator neural network was trained with the Heidelberg, CENTRIC, and CORE cohorts to produce post-contrast T1-weighted sequences based on pre-contrast MRI sequences. This generator neural network was combined with a discriminator neural network to distinguish the generator's false post-contrast T1-weighted sequences from true post-contrast T1-weighted sequences. For those instances where the discriminator could easily recognise the generator's false post-contrast T1-weighted sequences as implausible, such as a post-contrast T1-weighted sequence that was clearly different from the underlying true post-contrast T1-weighted sequence, the generator was penalised. Over time, the generator learned to generate more plausible synthetic post-contrast T1-weighted sequences. We show that this technique is superior to encoder-decoder dCNN architectures and allows to generate synthetic post-contrast T1-weighted sequences with a high structural similarity as compared with true post-contrast T1-weighted sequences (SSIM 0·818 in the EORTC-26101 test set). One of the initial hypothesis-generating studies[7] suggested that ADC sequences, which measure the magnitude of diffusion of water molecules and thereby approximate the cellularity of the tumour,[30] provide the most valuable information for the dCNN to synthesise the false post-contrast T1-weighted MRI sequences.[7] In our study, including ADC beyond pre-contrast anatomical MRI sequences for creating synthetic post-contrast T1-weighted MRI sequences did not increase the performance; instead we found that pre-contrast T1-weighted sequences followed by FLAIR and T2-weighted sequences were the most relevant. The lack of importance for ADC may reflect the biological basis of contrast enhancement in brain tumours, which occurs because of a non-specific increase in blood–brain barrier permeability and does not necessarily reflect active tumour (ie, high cellularity reflected by low ADC values).[31] By contrast, the central importance of pre-contrast T1-weighted sequences found in our study could be explained by the fact that the output of the dCNN (T1-subtraction sequence) is structurally similar to the pre-contrast T1-weighted sequence, since it is generated by subtracting the T1-weighted signal intensities from the corresponding post-contrast T1-weighted sequence. Moreover, the comparatively lower importance of T2-weighted and FLAIR sequences could additionally be explained by the lower image resolution and slice thickness of these sequences as compared with the T1-weighted sequences, which led to interpolation artifacts on T2-weighted and FLAIR sequences following image registration to the corresponding T1-weighted sequence that was used as a reference.

The use of synthetic post-contrast T1-weighted MRI sequences for assessment of contrast-enhancing tumour burden in our study showed a strong linear relationship of tumour volumes derived from synthetic versus true post-contrast T1-weighted sequences (concordance correlation coefficient 0·782 in the EORTC-26101 test set, p<0·0001) with contrast-enhancing tumour volumes being underestimated by a median of –0·48 cm³ or 7% in the EORTC-26101 test set. Spatial agreement between tumour volumes derived from synthetic versus true post-contrast T1-weighted sequences in the EORTC-26101 test set was moderate (median Sørensen–Dice coefficient of 0·28). However, this effect was largely driven by cases with small or absent contrast-enhancing tumour volumes, which were over-represented in the follow-up of the EORTC-26101 cohort and a negative effect on volumetric tumour response assessment could be mitigated by applying a volumetric cutoff equivalent to the 1×1 cm for measurable lesions as defined by RANO. Importantly, by analysing the longitudinal change in contrast-enhancing tumour volumes derived from synthetic versus true post-contrast T1-weighted sequences, we found that the median time to progression between the two approaches in the EORTC-26101 test set did only differ by 0·1 months. Inclusion of the longitudinal change in the T2-FLAIR signal abnormality volume as an additional response criterion further improved the agreement, resulting in identical median time to progression with the use of synthetic versus true post-contrast T1-weighted sequences. These results suggest that although contrast-enhancing tumour volumes are underestimated when using synthetic post-contrast T1-weighted sequences, quantification of the relative change in the longitudinal tumour burden resembles the information that would otherwise only be

available with administration of GBCA and acquisition of true post-contrast T1-weighted sequences. Moreover, the surrogate level of the time to progression derived from synthetic post-contrast T1-weighted MRI for predicting overall survival (HR 1·749, 95% CI 1·282–2·387, p=0·0004) was similar to the time to progression derived from true post-contrast T1-weighted MRI (1·799, 1·314–2·464, p=0·0003).

Our study has some limitations. First, we acknowledge the retrospective nature of the study and the inclusion of MRI data from various scanners and acquisition protocols. Although the use of heterogeneous data coming from more than 200 institutions acquired over a timeframe of more than 10 years enabled very stable and generalisable results when facing independent test sets, it also intrinsically restricted the maximum possible performance of the dCNNs due to inconsistencies in data acquisition, MRI field strengths, and interpolation artifacts. Visual inspection of the generated synthetic post-contrast T1-weighted images showed that although enhancement patterns were represented realistically, the overall impression of the synthetic post-contrast T1-weighted images was different from the original post-contrast T1-weighted images. Specifically, blurring of contrast-enhancing structures and slight over pronunciation of leptomeningeal vasculature were the main features in the synthetic post-contrast T1-weighted images and were seen more frequently when artifacts were present in the input sequences. Nevertheless, standardisation efforts of MRI acquisition protocols are an ongoing effort in the field of neuro-oncology,[20] and future large-scale studies could improve the performance of the dCNNs for generating synthetic post-contrast T1-weighted images without compromising generalisability. Second, although the clinical use of synthetic post-contrast T1-weighted MRI was shown for tumour response assessment in the EORTC-26101 test set, discrepancies in the time to progression on a per case basis need to be addressed in future studies. The sensitivity analysis in this study, which showed improved performance for 3D image acquisition protocols supports the importance of standardised acquisition protocols, and additional advanced MRI sequences might enable and improve the development of more precise prediction models. Third, methodological assessments should be done. To preserve voxel-level information, we opted for the U-Net architecture, as a baseline model and as a backbone network for the CGAN. However, fully convolutional networks are a suitable replacement for the U-Net and should be considered in the future. A comparison of synthetic post-contrast T1-weighted MRI based on pre-contrast MRI sequences with those based on low-dose contrast media applications needs to be addressed in future studies and an investigation of real-time practicability and validity of the proposed methodology. Finally, to evaluate the required technical effort and practical usefulness in the clinical setting in real time, further prospective evaluations are required, which should also include a thorough investigation of uncertainty estimation methods providing an assessment of reliability to the practitioner. Adding robustness to eventually missing input sequences could be a valuable extension to the practical application of the proposed method, which could be achieved by applying channel dropout during training.

In conclusion, generating synthetic post-contrast T1-weighted MRI from pre-contrast MRI using dCNN is feasible and quantification of the contrast-enhancing tumour burden from synthetic post-contrast T1-weighted MRI allows assessment of the patient's response to treatment with no significant difference as compared with when using true post-contrast T1-weighted sequences with administration of GBCAs. These findings could inform the application of dCNN to radiology to potentially reduce the necessity of GBCA administration.

Contributors

PV, CJP, and HM designed the study. MAM, IP, MS, MF, GB, UN, and PV performed quality control of MRI data. CJP, HM, GB, MAM, MF, FI, JP, and PV performed pre-processing of MRI data. CJP and HM performed development, training, and application of the artificial neural networks. CJP, HM, and PV performed post-processing of the data generated by the artificial neural network. AW, MP, MJvdB, BN, RS, TG, J-CT, MW, WW, and MB contributed to the primary analysis of the relevant data from the CENTRIC, CORE, and EORTC-26101 trials. PV, CJP, and TG performed the statistical analyses. PV, CJP, HM, WW, and MB interpreted the findings with essential input from all coauthors. PV, CJP, HM, and GB prepared the first draft of the manuscript. All authors critically revised and approved the final version of the manuscript. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication. All authors had access to all the data reported in the study. CJP, HM, GB, and PV accessed and verified all the data reported in the study.

Declaration of interests

SH reports grants from the German Research Council and Dietmar-Hopp Foundation, outside of the submitted work. JD reports grants from ViewRay, the Clinical Research Institute, Accuray, RaySearch Laboratories, Vision RT, Merck, Astellas Pharma, AstraZeneca, Siemens Healthcare, Solution Akademie, Egomed, Quintiles, Pharmaceutical Research Association, Boehringer Ingelheim, PTW-Freiburg, and Nanobiotix, outside of the submitted work. MP reports non-financial support from Pfizer, and grants and personal fees from Bayer, outside of the submitted work. MP also has a licensed patent for IDH1 vaccines, a patent H3 vaccine pending, and a patent AHR inhibitor with royalties paid to Bayer. AI reports grants and travel funding from Carthera; research grants from Transgene, Sanofi, Air Liquide, and Nutritheragene; travel funding from Leo Pharma; and is on the advisory board for Novocure and Leo Pharma, outside the submitted work. MjvdB reports personal fees from Roche, Cellgene, Bristol Myers Squibb, Agios, Merck Sharpe & Dohme, and Boehringer Ingelheim; and grants and personal fees from AbbVie, outside of the submitted work. BN is on the scientific advisory board for Karyopharm and BTG Pharmaceuticals and is on the data safety and monitoring board for the University of Pennsylvania (Philadelphia, PA, USA), outside of the submitted work. J-CT reports personal fees from BrainLab and carThera, outside of the submitted work. WW reports grants from Apogenix, Boehringer Ingelheim, and Pfizer; grants and personal fees from Merck Sharp and Dohme and Roche; and personal fees from Bristol Myers Squibb and Celldex, outside of the submitted work. MB reports personal fees from Boehringer Ingelheim, Merck, Bayer, Teva, B Braun, Springer, and Vascular Dynamics; grants and personal fees from Novartis, Codman, and Guerbet; and grants from Siemens, Hopp Foundation, the German Research Council, the EU, Stryker, and Medtronic, outside of the submitted work. All other authors declare no competing interests.

Data sharing

The institutional data used for training and validation are not publicly available, because they contain protected patient health information. The multi-institutional data from the CENTRIC and CORE trials (which were used for training and validation) and the multi-institutional data from the EORTC-26101 trial (which were used for independent testing) are not publicly available and restrictions apply to their use via the EORTC external research collaboration. The deep-learning algorithm used in this study is not publicly available yet, but the authors are investigating ways of disseminating it. The authors agree to apply the deep-learning algorithm to data provided by other academic researchers on their behalf for research purposes only, following completion of a Material Transfer Agreement. Proposals and requests for data access should be directed to the corresponding author via email.

Acknowledgments

We are grateful to Merck and Roche for supporting this study through an educational grant. We acknowledge funding by Deutsche Forschungsgemeinschaft (the German Research Foundation; project identifier 404521405 [SFB 1389—UNITE Glioblastoma, Work Package C02] and project identifier 402688427 [Priority Programme 2177 Radiomics: Next Generation of Biomedical Imaging, KI 2410/1-1, MA 6340/18-1]).

References

1 Hao D, Ai T, Goerner F, Hu X, Runge VM, Tweedle M. MRI contrast agents: basic chemistry and safety. *J Magn Reson Imaging* 2012; **36:** 1060–71.

2 Forghani R. Adverse effects of gadolinium-based contrast agents: changes in practice patterns. *Top Magn Reson Imaging* 2016; **25:** 163–69.

3 Weinreb JC, Rodby RA, Yee J, et al. Use of intravenous gadolinium-based contrast media in patients with kidney disease: consensus statements from the American College of Radiology and the National Kidney Foundation. *Kidney Med* 2020; **3:** 142–50.

4 Gulani V, Calamante F, Shellock FG, Kanal E, Reeder SB. Gadolinium deposition in the brain: summary of evidence and recommendations. *Lancet Neurol* 2017; **16:** 564–70.

5 Runge VM. Critical questions regarding gadolinium deposition in the brain and body after injections of the gadolinium-based contrast agents, safety, and clinical recommendations in consideration of the EMA's pharmacovigilance and risk assessment committee recommendation for suspension of the marketing authorizations for 4 linear agents. *Invest Radiol* 2017; **52:** 317–23.

6 Gong E, Pauly JM, Wintermark M, Zaharchuk G. Deep learning enables reduced gadolinium dose for contrast-enhanced brain MRI. *J Magn Reson Imaging* 2018; **48:** 330–40.

7 Kleesiek J, Morshuis JN, Isensee F, et al. Can virtual contrast enhancement in brain MRI replace gadolinium? A feasibility study. *Invest Radiol* 2019; **54:** 653–60.

8 Liu C, Zhu N, Sikka D, et al. Deep learning substitutes gadolinium in detecting functional and structural brain lesions with MRI. *Res Sq* 2020; published online Aug 26. https://doi.org/10.21203/rs.3.rs-56518/v1 (preprint)

9 Wick W, Gorlia T, Bendszus M, et al. Lomustine and bevacizumab in progressive glioblastoma. *N Engl J Med* 2017; **377:** 1954–63.

10 Wick W, Stupp R, Gorlia T, et al. Phase II part of EORTC study 26101: the sequence of bevacizumab and lomustine in patients with first recurrence of a glioblastoma. *J Clin Oncol* 2016; **34** (suppl)**:** 2019.

11 Nabors LB, Fink KL, Mikkelsen T, et al. Two cilengitide regimens in combination with standard treatment for patients with newly diagnosed glioblastoma and unmethylated MGMT gene promoter: results of the open-label, controlled, randomized phase II CORE study. *Neuro-oncol* 2015; **17:** 708–17.

12 Stupp R, Hegi ME, Gorlia T, et al. Cilengitide combined with standard treatment for patients with newly diagnosed glioblastoma with methylated MGMT promoter (CENTRIC EORTC 26071-22072 study): a multicentre, randomised, open-label, phase 3 trial. *Lancet Oncol* 2014; **15:** 1100–08.

13 Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds). Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234–41.

14 Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. *arXiv* 2016, published online Nov 26. https://ui.adsabs.harvard.edu/abs/2016arXiv161107004I (preprint).

15 Mongan J, Moy L, Kahn CE Jr, Kahn J. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell* 2020; **2:** e200029.

16 Kickingereder P, Isensee F, Tursunova I, et al. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. *Lancet Oncol* 2019; **20:** 728–40.

17 Isensee F, Schell M, Pflueger I, et al. Automated brain extraction of multisequence MRI using artificial neural networks. *Hum Brain Mapp* 2019; **40:** 4952–64.

18 Ellingson BM, Wen PY, Cloughesy TF. Modified criteria for radiographic response assessment in glioblastoma clinical trials. *Neurotherapeutics* 2017; **14:** 307–20.

19 Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process* 2004; **13:** 600–12.

20 Ellingson BM, Bendszus M, Boxerman J, et al. Consensus recommendations for a standardized Brain Tumor Imaging Protocol in clinical trials. *Neuro-oncol* 2015; **17:** 1188–98.

21 Wen PY, Chang SM, Van den Bent MJ, Vogelbaum MA, Macdonald DR, Lee EQ. Response assessment in neuro-oncology clinical trials. *J Clin Oncol* 2017; **35:** 2439–49.

22 Thust SC, Heiland S, Falini A, et al. Glioma imaging in Europe: a survey of 220 centres and recommendations for best clinical practice. *Eur Radiol* 2018; **28:** 3306–17.

23 Choi JW, Moon WJ. Gadolinium deposition in the brain: current updates. *Korean J Radiol* 2019; **20:** 134–47.

24 European Medicines Agency. EMA's final opinion confirms restrictions on use of linear gadolinium agents in body scans. July 21, 2017. https://www.ema.europa.eu/en/news/emas-final-opinion-confirms-restrictions-use-linear-gadolinium-agents-body-scans (accessed April 4, 2021).

25 Falk Delgado A, Van Westen D, Nilsson M, et al. Diagnostic value of alternative techniques to gadolinium-based contrast agents in MR neuroimaging-a comprehensive overview. *Insights Imaging* 2019; **10:** 84.

26 Paech D, Schuenke P, Koehler C, et al. T1ρ-weighted dynamic glucose-enhanced MR imaging in the human brain. *Radiology* 2017; **285:** 914–22.

27 Liu J, Gong E, Christen T, Zaharchuk G. Contrast-free MRI contrast enhancement with deep attention generative adversarial network. International Society for Magnetic Resonance in Medicine 27th Annual Meeting; Montreal; May 10–13, 2019 (abstr 1091).

28 Goodfellow I, Poget-Abadie J, Mirza M, et al. Generative adversarial nets. 28th Conference on Neural Information Processing Systems; Montreal, Canada; Dec 8–13, 2014 (abstr 27).

29 Gui J, Sun Z, Wen Y, Tao D, Ye J. A review on generative adversarial networks: algorithms, theory, and applications. *arXiv* 2020; published online Jan 20. https://ui.adsabs.harvard.edu/abs/2020arXiv200106937G (preprint).

30 Eidel O, Neumann JO, Burth S, et al. Automatic analysis of cellularity in glioblastoma and correlation with ADC using trajectory analysis and automatic nuclei counting. *PLoS One* 2016; **11:** e0160250.

31 Upadhyay N, Waldman AD. Conventional MRI evaluation of gliomas. *Br J Radiol* 2011; **84:** S107–11.