Title

Feasibility of Simulated Postcontrast MRI of Glioblastomas and Lower-Grade Gliomas by Using Three-dimensional Fully Convolutional Neural Networks.

Journal

Radiology Artificial Intelligence, 3(5)

ISSN

2638-6100

Authors

Calabrese, Evan
Rudie, Jeffrey D
Rauschecker, Andreas M
et al.

Publication Date

2021-09-01

DOI

10.1148/ryai.2021200276

Peer reviewed

Radiology: Artificial Intelligence

ORIGINAL RESEARCH

Feasibility of Simulated Postcontrast MRI of Glioblastomas and Lower-Grade Gliomas by Using Three-dimensional Fully Convolutional Neural Networks

Evan Calabrese, MD, PhD • Jeffrey D. Rudie, MD, PhD • Andreas M. Rauschecker, MD, PhD • Javier E. Villanueva-Meyer, MD • Soonmee Cha, MD

From the Department of Radiology and Biomedical Imaging (E.C., J.D.R., A.M.R., J.E.V.M., S.C.) and Center for Intelligent Imaging (E.C.), University of California at San Francisco, 350 Parnassus Ave, Suite 307H, San Francisco, CA 94143-0628. Received November 15, 2020; revision requested January 20, 2021; revision received April 20; accepted April 29. **Address correspondence to** E.C. (e-mail: *evan.calabrese@ucsf.edu*).

Supported by the National Institutes of Health Ruth L. Kirschstein Institutional National Research Service Award under award number T32EB001631. This project was also indirectly supported by the Radiological Society of North America (RSNA) Research and Education (R&E) Foundation, through grant number RR2011. The content is solely the responsibility of the authors and does not necessarily represent the official views of the RSNA R&E Foundation. The University of California San Francisco resident research fund provided financial support for some graphics processing hardware used in this study.

Conflicts of interest are listed at the end of this article.

Radiology: Artificial Intelligence 2021; 3(5):e200276 • https://doi.org/10.1148/ryai.2021200276 • Content codes: AI NR OI

Purpose: To evaluate the feasibility and accuracy of simulated postcontrast T1-weighted brain MR images generated by using precontrast MR images in patients with brain glioma.

Materials and Methods: In this retrospective study, a three-dimensional deep convolutional neural network was developed to simulate T1-weighted postcontrast images from eight precontrast sequences in 400 patients (mean age, 57 years; 239 men; from 2015 to 2020), including 332 with glioblastoma and 68 with lower-grade gliomas. Performance was evaluated by using quantitative image similarity and error metrics and enhancing tumor overlap analysis. Performance was also assessed on a multicenter external dataset (n = 286 from the 2019 Multimodal Brain Tumor Segmentation Challenge; mean age, 60 years; ratio of men to women unknown) by using transfer learning. A subset of cases was reviewed by neuroradiologist readers to assess whether simulated images affected the ability to determine the tumor grade.

Results: Simulated whole-brain postcontrast images were both qualitatively and quantitatively similar to the real postcontrast images in terms of quantitative image similarity (structural similarity index of 0.84 ± 0.05), pixelwise error (symmetric mean absolute percent error of 3.65%), and enhancing tumor compartment overlap (Dice coefficient, 0.65 ± 0.25). Similar results were achieved with the external dataset (Dice coefficient, 0.62 ± 0.27). There was no difference in the ability of the neuroradiologist readers to determine the tumor grade in real versus simulated images (accuracy, 87.7% vs 90.6%; P = .87).

Conclusion: The developed model was capable of producing simulated postcontrast T1-weighted MR images that were similar to real acquired images as determined by both quantitative analysis and radiologist assessment.

Supplemental material is available for this article.

Three-dimensional (3D), gadolinium-enhanced, T1-weighted MRI is essential for evaluation of a variety of pathologic conditions that alter brain vascularity and vascular permeability, including inflammatory, infectious, and neoplastic processes. Unfortunately, gadolinium-based contrast agents have several potential drawbacks that limit their usability. First, there are costs associated with administering contrast agents, patient monitoring, and acquiring additional postcontrast images. Second, there is potential for a variety of adverse reactions, ranging from nausea or headache to anaphylaxis and nephrogenic systemic fibrosis (1). Third, gadolinium-based contrast agents are contraindicated in certain patient groups, including those with advanced renal disease and pregnant patients. Finally, there remains uncertainty surrounding the possible long-term effects of gadolinium deposition in the brain and bones (2,3). For these reasons, it would be particularly appealing to develop simulated images that would closely approximate pathologic changes depicted on real postcontrast images if they could be used without compromising diagnostic accuracy.

Modern brain MRI protocols comprise several different sequences used to generate images encompassing the same anatomy with different contrast weightings. Most common brain MRI series are both physically and mathematically related to each other, which raises the possibility that a specific image series could be computationally simulated from others. In this context, image *simulation* refers to computational synthesis of artificial (simulated) images by using different images of the same anatomy as that of the inputs. Several recent prior studies have demonstrated both intramodality (eg, MRI to MRI) and intermodality (eg, MRI to CT or PET) computational image simulation (4–7). Computational simulation of postcontrast images presents an additional challenge because it requires inference of the effect of an exogenous agent that is not present in the input images. Despite these challenges, a small number of prior studies have demonstrated both the ability

Abbreviations

dCNN = deep convolutional neural network, FLAIR = fluid-attenuated inversion recovery, GPU = graphics processing unit, 3D = three-dimensional

Summary

Simulated postcontrast T1-weighted brain MR images derived from precontrast images by using a three-dimensional deep convolutional neural network closely resembled real postcontrast images of patients with glioblastoma and lower-grade gliomas.

Key Points

- Simulated postcontrast images were quantitatively similar to real postcontrast images across the whole brain and, to a lesser but still significant degree, within the tumor region.
- Automated enhancing tumor segmentations from simulated postcontrast images showed moderate overlap with enhancing tumor segmentations from real postcontrast images (mean Dice coefficient, 0.65 ± 0.25).
- There was no difference in the ability of neuroradiologist readers to determine the tumor grade in real versus simulated images (accuracy, 87.7% vs 90.6%; P = .87).

Keywords

MR-Contrast Agent, MR-Imaging, CNS, Brain/Brain Stem, Contrast Agents-Intravenous, Neoplasms-Primary, Experimental Investigations, Technology Assessment, Supervised Learning, Transfer Learning, Convolutional Neural Network, Deep Learning Algorithms, Machine Learning Algorithms

to infer contrast-enhancing lesions and to simulate postcontrast images from precontrast inputs (8–10).

The purpose of our study was to evaluate the feasibility and accuracy of a deep learning method for simulating postcontrast T1-weighted brain images by using precontrast MR images of patients with glioblastoma and lower-grade gliomas. Compared with prior work focused on postcontrast image simulation, this study uses a larger patient sample (n = 400) and includes validation on an external dataset (n = 286). We further build on prior work by using the following: incorporation of advanced precontrast image series, including arterial spin labeling perfusion images and 55-direction diffusion images; 3D, fully convolutional deep neural networks; and examination of results for both glioblastomas and lower-grade gliomas.

Materials and Methods

Study Group

In this retrospective study, all study protocols were Health Insurance Portability and Accountability Act compliant and were approved by the institutional review board, which provided a waiver for consent. Inclusion criteria were patients 18 years of age or older who underwent a complete preoperative brain MRI and subsequent brain tumor resection at the University of California San Francisco and who had subsequent pathologic results revealing grade II–IV diffuse glioma. Exclusion criteria were any prior history of brain tumor resection or severe MRI artifacts that obscured the tumor region. Eligible patients were identified by using a combined radiologic and pathologic report search tool (mPower Clinical Analytics; Nuance Communications). All eligible cases from January 2015 onward (in sequential order) were included until 400 cases were reached, which occurred in April of 2020. We analyzed preoperative MR images in a total of 400 adult patients with either glioblastoma (n = 332) or lower-grade diffuse gliomas (n = 68), which were each confirmed by using histopathologic assessment. All patients were imaged prior to their first surgical resection. Although a history of prior chemotherapy or radiation was not systematically assessed, surgical resection is typically the first treatment modality for patients with diffuse glioma at the study institution. A total of 199 patients were also part of an unrelated prior study focused on the prediction of genetic alterations in glioblastoma (11). Table 1 provides details on patient characteristics.

MR Image Acquisition

All patients underwent a standardized clinical preoperative primary brain tumor MRI protocol that used a 3.0-T imager (Discovery 750; GE Healthcare) and a dedicated eight-channel head coil (Invivo). The imaging protocol included T2-weighted imaging; T2-weighted, fluid-attenuated inversion recovery (FLAIR) imaging; susceptibility-weighted imaging; diffusion-weighted imaging; pre- and postcontrast T1-weighted imaging; arterial spin labeling; and 55-direction, high angular resolution diffusion imaging. Acquisition parameters are included in Appendix E1 (supplement).

Image Preprocessing

An automated image preprocessing pipeline was implemented in Python version 3.8 (Python Software Foundation) by using Nipype version 1.6.0 *(https://nipype.readthedocs.io/en/latest/)* (12). Diffusion data were processed by using the Functional Magnetic Resonance Imaging of the Brain Software Library (Oxford University), yielding standard derived images that included mean diffusivity and fractional anisotropy (13–15). Patient images were coregistered to the Montreal Neurological Institute Brain Atlas space by using Advanced Normalization Tools (University of Pennsylvania) (16,17). Additional steps included brain masking, coil-bias correction, and intensity normalization (18). Complete preprocessing details are included in Appendix E1 (supplement).

Deep Convolutional Neural Network Architecture

A 3D, fully convolutional, deep convolutional neural network (dCNN) model for simulating postcontrast images from precontrast series was implemented with Python version 3.8 and TensorFlow version 2.3.0 *(https://www.tensorflow.org/)*. The model architecture was based on the U-Net architecture (University of Freiburg) (19) with 3D convolutions, convolutional up- and downsampling, long-range skip connections, bottleneck residual blocks (20), per-layer batch normalization, leaky rectified linear unit activation, and feature dropout. Additional model details are included in Appendix E1 (supplement), and code is provided at *https://github.com/ecalabr/gadnet/*. A single example study from the internal dataset is included for testing purposes.

Table 1: Patient Characteristics

Parameter	Internal Training and Testing	BraTS External Test
No. of patients	400	286
Mean age (y)	57	60
Sex	239 men, 161 women	Unknown
Glioma type		
Glioblastoma	332	286
Lower-grade glioma	68	0
Glioma grade		
1	0	0
2	27	0
3	41	0
4	332	286
4*	23	Unknown

Note.—Except where otherwise noted, data are numbers. BraTS = Multimodal Brain Tumor Segmentation.
* Indicates a grade 4 glioma with the R132H gain-of-function mutation of isocitrate dehydrogenase 1.

Model Hyperparameters and Hyperparameter Optimization

Several model hyperparameters were optimized by using a 16-iteration random search of the following parameter space: 8, 16, 32, and 64 base filters; batch sizes of 4, 8, 16, and 32; and dropout rates of 0.2, 0.3, 0.4, and 0.5. Final hyperparameters used for training included 32 base filters per layer, a batch size of 16, and a dropout rate of 0.4. The loss function for model training and hyperparameter optimization was the mean squared error between the intensity-normalized real and simulated images. The loss optimizer was Adam (OpenAI and University of Toronto), which had an initial learning rate of 0.001 (21).

Model Inputs and Training

The model was trained by using images from eight precontrast imaging sequences: T1-weighted, T2-weighted, T2-weighted FLAIR, susceptibility-weighted, diffusion-weighted, arterial spin labeling, mean diffusivity, and fractional anisotropy. A 10-fold cross-validation approach was employed, with a split of 90% training and 10% testing, ensuring independent training and testing sets. Training inputs consisted of 80 × 80 × 80–voxel patches from each of the eight precontrast image series. Data augmentation included random three-axis rotations and dimension swaps. Each training session comprised 25 epochs with four Tesla V100, 32-GB graphics processing units (GPUs) (NVIDIA). Additional details are included in Appendix E1 (supplement).

Image Simulation and Quantitative Similarity Comparison

Simulated postcontrast images were compared with real postcontrast T1-weighted images by using three image similarity metrics (neighborhood cross-correlation, histogram mutual information, and the structural similarity index), as well as four error metrics (normalized root mean square error, symmetric mean absolute percent error, log accuracy ratio, and median symmetric accuracy) (16,22,23). Metrics were computed across the whole brain, as well as within the whole tumor and tumor core. The metric choice was based on prior theoretical and applied studies (8,23). Details of metric implementation are included in Appendix E1 (supplement), and a visual reference for interpreting metric values is presented as Figure E1 (supplement).

Enhancing Tumor Overlap Analysis

Volumetric overlap between the simulated and real enhancing tumor components was assessed by first performing automated deep learning–based tumor subcompartment segmentation of both real and simulated images by using a previously published method (11,24). The resulting real and simulated enhancing tumor compartment automated segmentations were compared by using the Dice overlap. Nonenhancing tumors were included in the analysis to capture potential false-positive results. Of note, the automated tumor segmentation algorithm was applied to the whole brain but does not segment normal intracranial enhancement, such as arteries and veins.

Relative Contribution of Input Series

To determine the relative contribution of each of the precontrast input series to the final simulated result, the dCNN model was trained on a single cross-validation fold with each combination of a T1-weighted series plus one other series. Model performance was assessed by using previously described image similarity and error metrics.

Radiologists' Assessment of Simulated Images

Four neuroradiologist readers (authors J.D.R., A.M.R., J.E.V.M., and S.C. with 1, 1, 5, and 23 years of experience, respectively, as an attending radiologist) were shown a random,

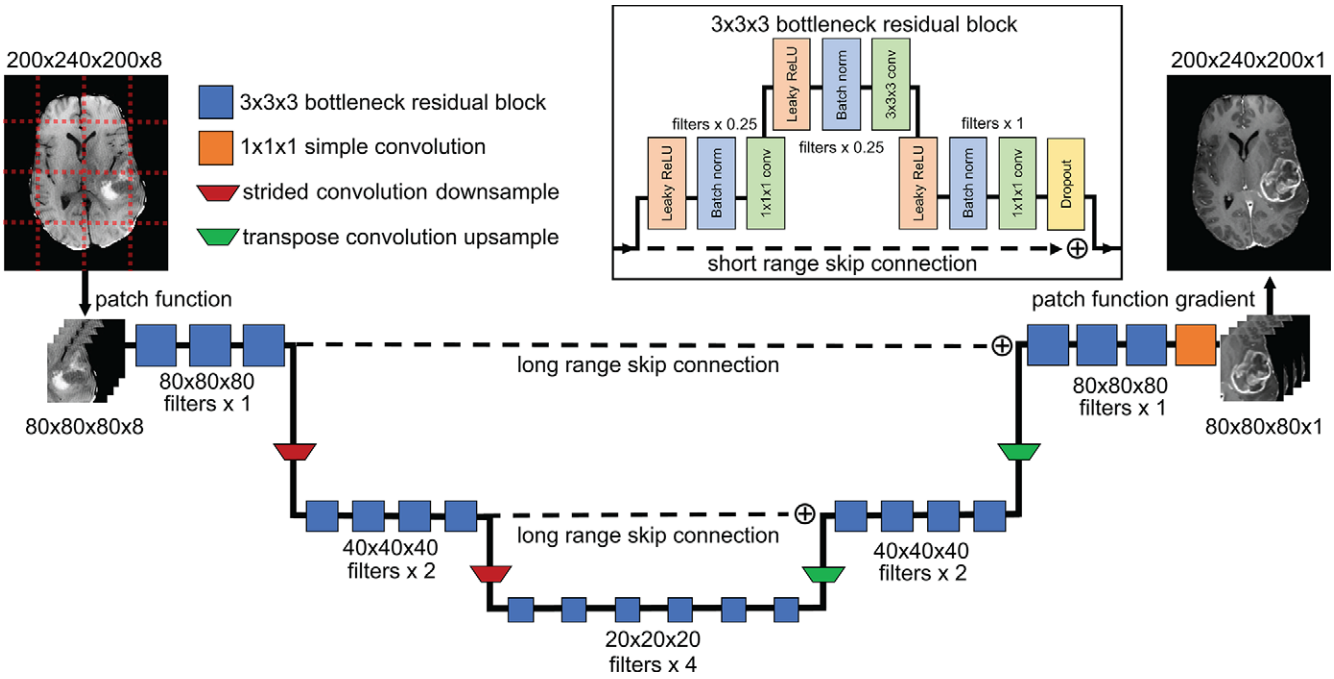

Figure 1: Schematic of the deep convolutional neural network architecture. Preprocessed precontrast images (left) of a 200 × 240 × 200–voxel size acquired by using eight MRI series are used as network inputs. Input images are divided into 80 × 80 × 80–voxel training patches and then fed into a fully convolutional deep neural network with three-dimensional convolution bottleneck residual blocks (blue squares), strided convolution downsampling (red trapezoids), transpose convolution upsampling (green trapezoids), and long-range skip connections with feature concatenation (dashed lines). A 1 × 1 × 1 convolutional layer (orange square) is used to map features to the final output image patches (right). An inset schematic of the residual block architecture is included (upper right). Conv = convolution, ReLU = rectified linear unit.

blinded set of either 80 real or 80 simulated, 3D, T1-weighted postcontrast image volumes of either high-grade (grade IV) or lower-grade (grade II–III) diffuse gliomas. Readers were asked to assess whether or not they believed the images were real or simulated, whether or not abnormal enhancement was present within the tumor region, and what histologic tumor grade was most likely given the image appearance. Additional details are provided in Appendix E1 (supplement).

External Validation

Model performance was further evaluated by using the 2019 Multimodal Brain Tumor Segmentation Challenge training dataset (25–27) consisting of 286 patients with glioblastoma and lower-grade gliomas with T1 pre- and postcontrast, T2, and FLAIR images. Only the training portion of the dataset was used because it included tumor segmentations, which were required for quantitative analysis. A "reduced model" was trained on internal study data by using only the corresponding three precontrast series (T1, T2, and FLAIR) and was evaluated by using the entire external dataset. The reduced model was then refined by training for five additional epochs by using 50% of the external data and was evaluated by using the remaining 50% of the external dataset ("refined reduced model"). Table 1 provides characteristics from the patients within the external dataset.

Statistical Analysis

Statistical significance was determined by using a two-tailed *t* test implemented in SciPy version 1.5.3 *(https://www.scipy.org/)*, with a *P* value less than .05 considered as indicating statistical significance. For comparing proportion data such as accuracy, a two-tailed *z* test was used, with a *P* value less than .05 considered as indicating statistical significance. When applicable, Bonferroni correction was used to control for multiple comparisons.

Results

dCNN Performance

A graphic schematic of the dCNN architecture used for this study is presented as Figure 1. Each 25-epoch training session lasted approximately 14 hours, distributed across four GPUs with up to four concurrent training folds running at the same time (total of 16 GPUs). The mean squared error loss between real and simulated images decreased throughout training and converged to oscillation in each case.

Quantitative Analysis of Simulated Postcontrast Images

The proposed fully convolutional dCNN network yielded simulated postcontrast T1-weighted images that were qualitatively similar to real postcontrast images, regardless of the tumor grade or image plane (Fig 2). Quantitative analysis showed strong similarity (and low voxelwise error) between real and simulated postcontrast images both across the whole brain and within the tumor (Fig 3). All quantitative metrics were slightly higher across the whole brain than within the tumor. Results were similar for the full model and the reduced model (trained using only internal T1, T2, and FLAIR images). Importantly, real postcontrast images were more closely related to simulated images than to real precontrast T1-weighted images by all metrics and in all evaluated regions, which indicated that the simulated images were

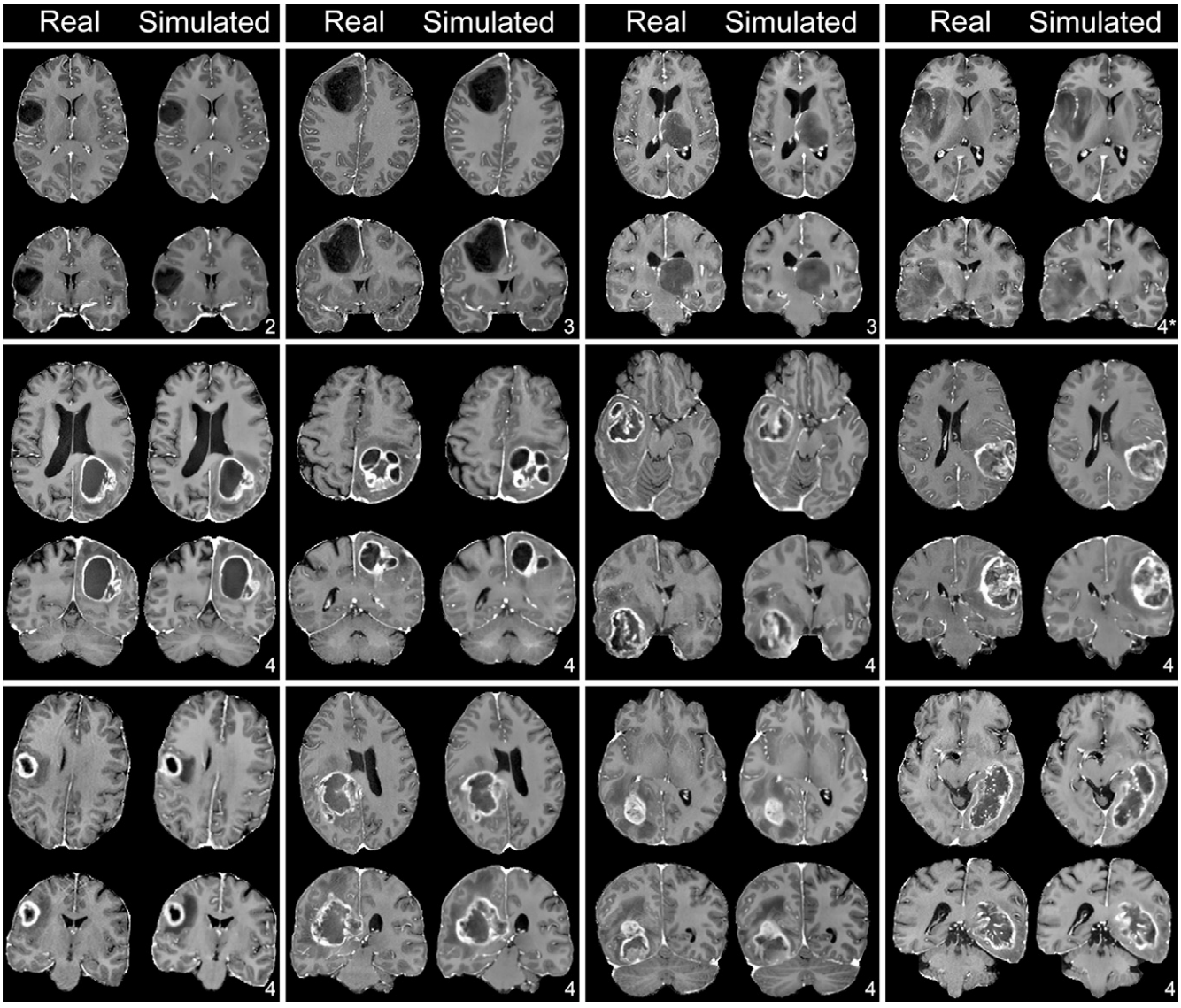

Figure 2: Simulated versus real axial and coronal postcontrast T1-weighted brain images of patients with glioblastoma and lower-grade gliomas. Each panel represents a separate patient, with real axial and coronal images shown on the left side of the panel and corresponding simulated images shown on the right side of the panel. The glioma grade (2–4) is indicated at the bottom right of each panel. The use of "4*" denotes a grade 4 glioma with the R132H gain-of-function mutation in isocitrate dehydrogenase 1. Of note, none of the patients presented in this figure had any history of prior treatment or biopsy.

closer to the truth than the precontrast inputs. All metrics were lower within the tumor core than across the whole brain after Bonferroni correction. Subgroup analysis based on the tumor grade showed lower performance in the tumor region of lower-grade gliomas than in glioblastomas (Fig E2 [supplement]).

Enhancing Tumor Component Overlap

Average overlap between the simulated and real contrast-enhancing tumor component as measured by the Dice coefficient was 0.65 ± 0.25 (95% CI: 0.63, 0.67) across the study group (Fig 4). Stratified by tumor grade, the average Dice coefficients were 0.65 ± 0.20 (95% CI: 0.63, 0.67) for glioblastomas and 0.58 ± 0.4 (95% CI: 0.49, 0.68) for lower-grade gliomas. After eliminating very small regions of enhancement (< 0.1 mL), the average Dice coefficients were 0.64 ± 0.20 (95% CI: 0.62, 0.66) for glioblastomas and 0.73 ± 0.37 (95% CI: 0.64, 0.82) for lower-grade gliomas. False-positive enhancement (ie, in simulated images but not in real images) greater than 0.1 mL in total volume was identified in 7.35% (five of 68) of lower-grade gliomas evaluated in this study. All five cases of false-positive enhancement are shown in Figure 5. False-negative enhancement greater than 0.1 mL was not identified.

Relative Contribution of Input Series

We also assessed the relative contribution of different image types in the training process. In general, models trained with all combinations of T1-weighted series plus one other input series showed lower performance by all metrics compared with the full model (Fig 6); however, not all of these differences reached statistical significance. Along with T1-weighted images, the most important input series (ie, smallest increase in error and smallest decrease in similarity) across the whole brain was the

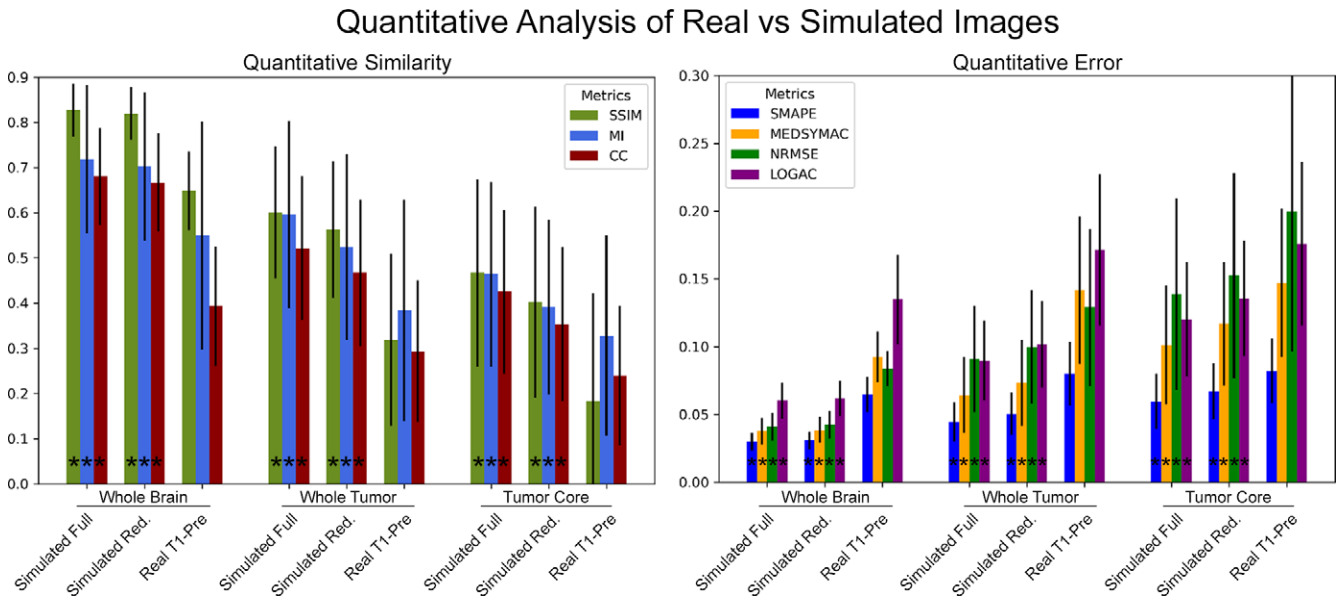

Figure 3: Quantitative comparisons between real and simulated postcontrast images using both image similarity (left) and error (right) metrics. Real T1-weighted postcontrast images were compared with simulated images from the full model (Simulated Full), simulated images from the reduced input model (Simulated Red.), and real T1-weighted precontrast (Real T1-Pre) images. Comparisons between real precontrast and real postcontrast images are provided for context. Results are shown across the whole brain, within the tumor region, and within the tumor core. * denotes a statistically significant difference compared with Real T1-Pre images. CC = cross-correlation, LOGAC = log accuracy ratio, MEDSYMAC = median symmetric accuracy, MI = histogram mutual information, NRMSE = normalized root mean square error, SMAPE = symmetric mean absolute percent error, SSIM = structural similarity index.

T2-weighted series, followed by the FLAIR- and susceptibility-weighted series. The T2-weighted series was also the most important additional series within the tumor core, followed by the susceptibility- and FLAIR-weighted series. Models trained on all other combinations of the T1-weighted series plus one other series had lower performance than the full model by all metrics.

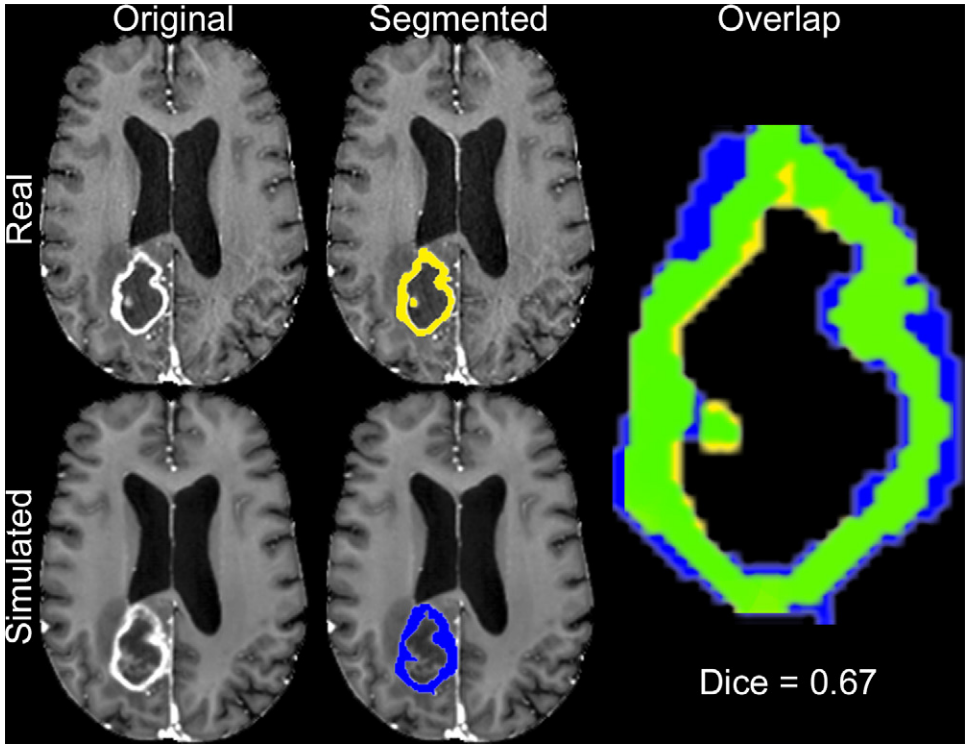

Figure 4: Illustration of Dice overlap analysis. The enhancing tumor component is segmented from real and simulated postcontrast images by using a previously published automated method. Dice overlap of the enhancing component is then calculated for each case. The real image segmentation is shown in yellow, the simulated image segmentation is shown in blue, and their intersection is shown in green. In this example, the Dice coefficient is 0.67.

Radiologist Assessment of Simulated Images

Results of the neuroradiologist reader study are presented in Table 2. Neuroradiologist readers performed similarly when assessing simulated and real images for both assessing the tumor grade (accuracy, 90.6% vs 87.7%; P = .87) and detecting the presence of abnormal tumor enhancement (accuracy, 86.9% vs 74.2%; P = .36). Interrater agreement for the tumor grade was substantial (κ, 0.70) and was almost perfect for tumor enhancement (κ, 0.90). However, it should be noted that the "truth" for the presence of abnormal tumor enhancement was based on an automated tumor segmentation algorithm and might have therefore produced results that were slightly different from those of the human readers. In contrast, reader performance for detect-

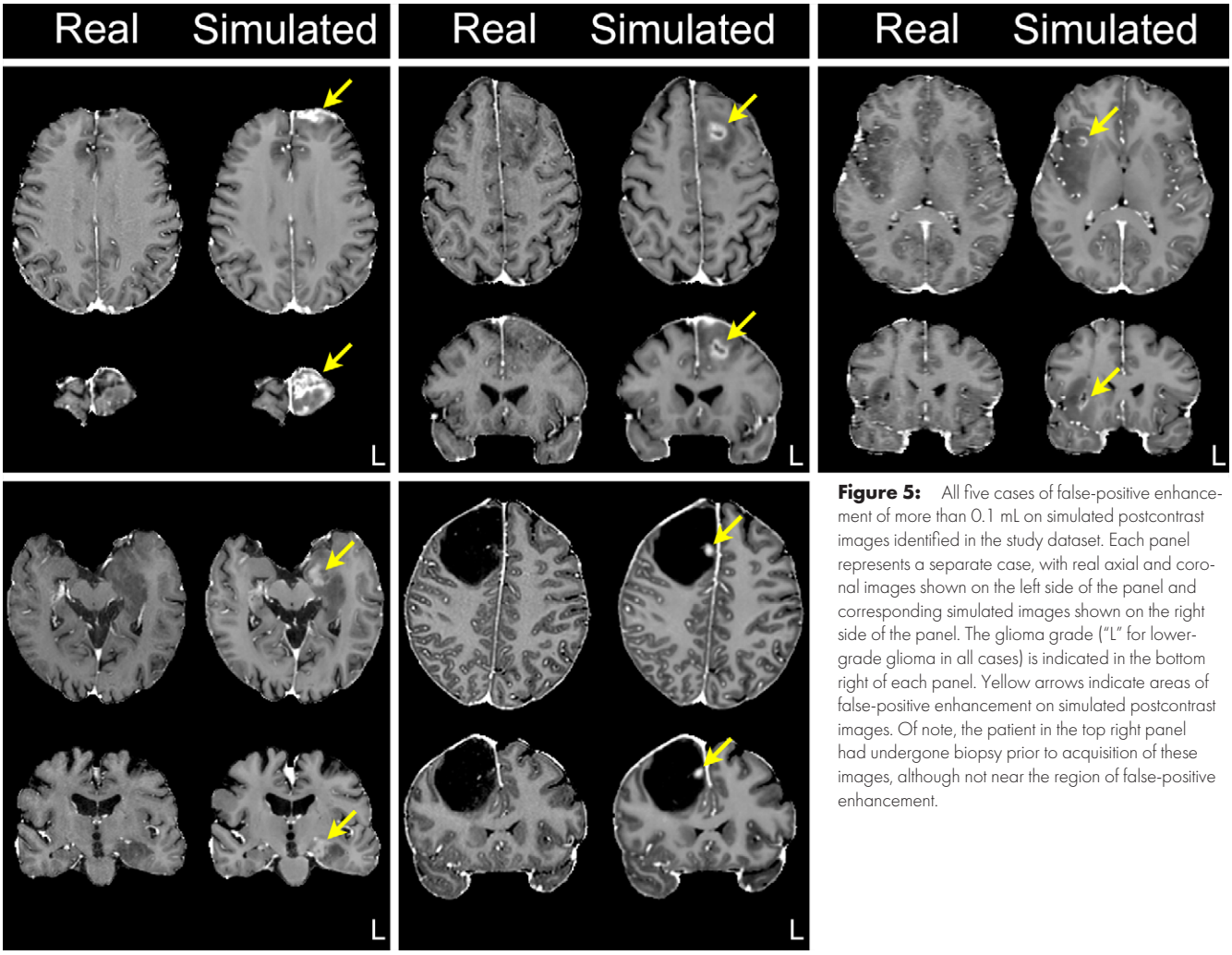

Figure 5: All five cases of false-positive enhancement of more than 0.1 mL on simulated postcontrast images identified in the study dataset. Each panel represents a separate case, with real axial and coronal images shown on the left side of the panel and corresponding simulated images shown on the right side of the panel. The glioma grade ("L" for lower-grade glioma in all cases) is indicated in the bottom right of each panel. Yellow arrows indicate areas of false-positive enhancement on simulated postcontrast images. Of note, the patient in the top right panel had undergone biopsy prior to acquisition of these images, although not near the region of false-positive enhancement.

ing real versus simulated images was poor (accuracy, 59.1%), with poor interrater agreement (κ, −0.23) being shown. Although it was not the purpose of this study to assess non–glioma-related abnormal enhancement, radiologists' review of images demonstrated several incidental examples of this type of abnormal enhancement, such as meningiomas or developmental venous anomalies. Representative examples showing both real and simulated images of these incidental findings are included as Figure E4 (supplement).

External Dataset Validation

The reduced model had lower performance on the external validation data; however, performance was nearly completely recovered after refining the model by training for an additional five epochs with 50% of the external data (the refined reduced model), with some evaluation metrics showing no statistical difference across the whole brain or within the tumor core (Fig 7). Enhancing tumor segmentation overlap analysis on the external dataset yielded an average Dice coefficient of 0.62 ± 0.27 (95% CI: 0.59, 0.65), which was not significantly different from internal dataset results. Example simulated images from the external dataset generated by using the refined reduced model are shown in Figure E3 (supplement).

Discussion

This study presents a deep learning method for generating simulated postcontrast brain MR images from precontrast images in patients with glioma. The fundamental assumption of our model was that the underlying factors contributing to glioma enhancement, such as perfusion, vascularity, and permeability, had some effect on the signal of precontrast imaging (28). Postcontrast image simulation can therefore be understood as a multiple regression of precontrast input signals to predict areas of enhancement. For this study, the "regression model" was a dCNN; however, this choice simply reflects the current state of the art rather than the theoretical optimal model for this task. In this context, simulated images can be interpreted as a probabilistic representation of real postcontrast images. However, regardless of context, simulated images must closely recapitulate real images on both a voxelwise and structural level to be diagnostically viable.

The primary method for evaluating simulated postcontrast images in this study was with quantitative image similarity and error metrics. We found that real postcontrast images were more closely related to simulated postcontrast images than to precontrast images, both across the whole brain and within the tumor, regardless of the tumor grade. Model performance was slightly lower in the lower-grade glioma

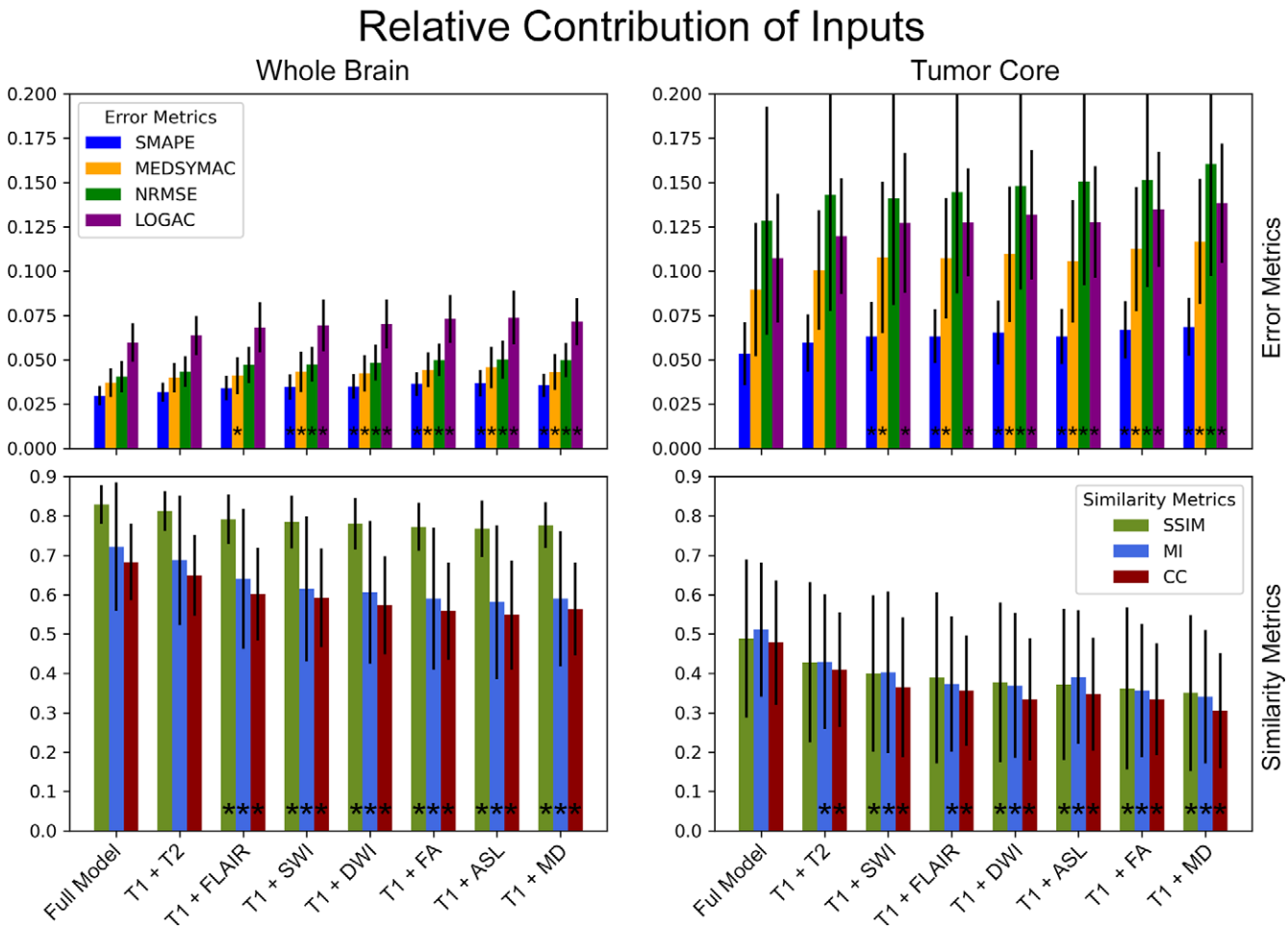

Figure 6: Quantitative image similarity and error analysis of real versus simulated postcontrast images derived from the models trained on the full set of inputs (Full Model) and each combination of a T1-weighted series plus one other series. Error metrics (top) and similarity metrics (bottom) are calculated across the whole brain (left) and within the tumor core (right). * denotes a statistically significant difference compared with the Full Model. ASL = arterial spin labeling, CC = cross-correlation, DWI = diffusion-weighted imaging, FA = fractional anisotropy, FLAIR = fluid-attenuated inversion recovery, LOGAC = log accuracy ratio, MD = mean diffusivity, MEDSYMAC = median symmetric accuracy, MI = histogram mutual information, NRMSE = normalized root mean square error, SMAPE = symmetric mean absolute percent error, SSIM = structural similarity index, SWI = susceptibility-weighted imaging.

group, likely because of their relative underrepresentation in the study dataset. These results show strong numerical correspondence between real and simulated postcontrast images but provide relatively little understanding of practical similarity for diagnostic tasks.

Enhancing tumor Dice overlap analysis can be interpreted as a more practically relevant metric for assessing the presence and/or extent of abnormal enhancement. Our model achieved an enhancing tumor Dice overlap of 0.65 ± 0.25 (95% CI: 0.63, 0.67), which we considered moderate performance, given the irregular enhancement often seen in gliomas and the high interrater variability of enhancing tumor segmentations reported in prior studies (29). However, our analysis did identify false-positive enhancement within a subset of lower-grade gliomas, which could be problematic for future clinical deployment of this method if not addressed. For

Table 2: Results of the Neuroradiologist Reader Study

Metric	Tumor Grade	Tumor Enhancement	Simulated vs Real
Accuracy	90.6%	83.4%	59.1%
Precision	88.2%	93.3%	60.4%
Recall	93.8%	83.3%	52.5%
F1	0.90	0.88	0.56
MCC	0.81	0.63	0.18
κ	0.70	0.90	−0.23

Note.—Each column represents a task that readers were asked to assess for each case: "Tumor Grade" represents whether the tumor was most likely to be high-grade (WHO grade IV) or lower-grade (WHO grade II–III); "Tumor Enhancement" represents whether abnormal contrast enhancement was present in the tumor; and "Simulated vs Real" represents whether the images were felt to be simulated or real. F1 = F1 score, κ = Cohen κ for interrater agreement, MCC = Matthews correlation coefficient, WHO = World Health Organization.

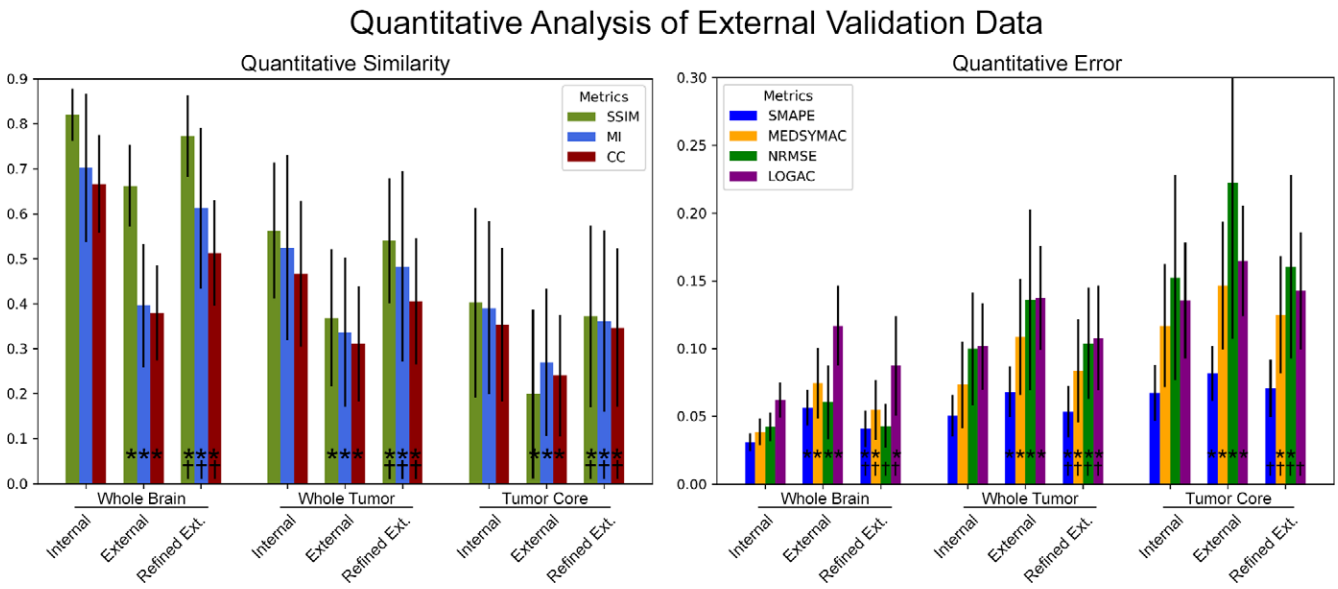

Figure 7: Quantitative analysis of model performance on external validation data. Quantitative comparisons between real and simulated postcontrast images using both image similarity (left) and error (right) metrics are shown. Real T1-weighted postcontrast images are compared with simulated images derived from the reduced input model using internal data (Internal), external Multimodal Brain Tumor Segmentation Challenge data using the reduced model (External), and external data using the refined reduced model (Refined Ext.). Results are shown across the whole brain, within the tumor region, and within the tumor core. * denotes a statistically significant difference compared with the internal dataset results. † denotes a statistically significant difference between external and refined external results. CC = cross-correlation, LOGAC = log accuracy ratio, MEDSYMAC = median symmetric accuracy, MI = histogram mutual information, NRMSE = normalized root mean square error, SMAPE = symmetric mean absolute percent error, SSIM = structural similarity index.

example, this limitation could be addressed by including more lower-grade glioma cases or by oversampling nonenhancing tumors. In addition, prior nonsurgical treatment such as chemotherapy or radiation could affect the algorithm output and could contribute to false-positive enhancement; however, none of the false-positive enhancements identified in this study had received prior treatment.

Interestingly, the contribution of the various precontrast image series was not the same in normal brain tissue and within the tumor. T2-weighted and FLAIR images were the most significant contributors (other than T1-weighted images) across the whole brain, whereas susceptibility-weighted images were also a significant contributor within the tumor. The contribution of T2-weighted and FLAIR images likely relates to their thin sections and parenchymal contrast; however, the contribution of susceptibility-weighted images is less apparent. One possibility is that increased perfusion and/or neovascularization alter the magnetic susceptibility of enhancing tissue (30). Surprisingly, arterial spin labeling perfusion images were not a major contributor despite the close relationship between perfusion and glioma enhancement (31), possibly owing to the high physiologic and technical variability of arterial spin labeling imaging (32).

Compared with prior similar work, this study used a larger patient sample and included validation with an external dataset (8). External validation results were somewhat limited by differences in underlying data. Most notably, a substantial portion of Multimodal Brain Tumor Segmentation Challenge precontrast data is two-dimensional, whereas the majority of the internal dataset was 3D. Despite these differences, model performance on the external dataset was nearly completely recovered by training the model for five additional epochs with half of the external data, and enhancing tumor overlap results were not significantly different. These results suggest that the proposed model is readily applicable to other datasets through a relatively rapid transfer learning approach. Applicability to other types of enhancing intracranial pathologic conditions was not directly assessed; however, we observed some incidental enhancing lesions in our sample that were represented in simulated images with varying accuracy.

One limitation of our results is the apparent blurring of simulated postcontrast images compared with real images. This effect, which is likely related to minor misregistration of precontrast inputs and/or feature resampling in the dCNN, could obscure subtle areas of enhancement. Such contrast blurring in simulated images could be particularly problematic for determining whether abnormal enhancement represents true progression or pseudoprogression in treated tumors. Although this blurring is likely overall detrimental to image quality and interpretability, it was also observed that simulated images are less affected by patient motion artifacts (Fig E3 [supplement], bottom right panel). In addition, our neuroradiologist reader study results suggest that simulated images were not easily distinguishable from real images when presented individually and did not significantly affect the ability to detect tumor enhancement or determine the tumor grade.

There are several possible approaches to building on this work. For example, including a larger number of cases, increasing the quality and quantity of precontrast images, and training on larger patches or whole-brain images could improve results. The use of image discriminators to guide training, as is done in generative adversarial networks, may also yield improvements. In addition, the methods presented here could be adapted to other similar tasks, including simulation of other MRI series (eg, simulated perfusion imaging), simulation of postcontrast images

in other modalities (eg, simulated postcontrast CT), or intermodality simulation (eg, simulated MRI from CT).

The results from this study suggest that there may be a role for deep learning–based simulated postcontrast MR images to help reduce the need for gadolinium-based contrast agents in some cases. Although we focused on preoperative gliomas, it would be straightforward to apply this method to postoperative gliomas, other tumor types, or other enhancing brain diseases. Future analysis of model performance in multiple disease states will be important, particularly because a specific diagnosis is not always known a priori. In addition, future work is necessary to assess the diagnostic accuracy of simulated postcontrast images, which was not addressed in this study.

Acknowledgments: The authors acknowledge the University of California San Francisco Center for Intelligent Imaging and Dr Sharmila Majumdar for support, including support related to technical expertise and deep learning processing hardware.

Author contributions: Guarantors of integrity of entire study, E.C., S.C.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, E.C., J.D.R., J.E.V.M., S.C.; clinical studies, E.C., J.E.V.M., S.C.; statistical analysis, E.C., A.M.R.; and manuscript editing, all authors

Disclosures of Conflicts of Interest: **E.C.** disclosed Radiological Society of North America (RSNA) resident research grant (No. RR2011). no relevant relationships. **J.D.R.** disclosed American Society of Neuroradiology (ASNR) research grant in artificial intelligence; member of *Radiology: Artificial Intelligence* 2019-2021 trainee editorial board. **A.M.R.** disclosed RSNA scholar grant and ASNR trainee grant, both paid to author's institution; former member of *Radiology: Artificial Intelligence* trainee editorial board. **J.E.V.M.** disclosed grant support to author's institution from GE Healthcare for unrelated work. **S.C.** disclosed no relevant relationships.

References

1. Prince MR, Zhang H, Zou Z, Staron RB, Brill PW. Incidence of immediate gadolinium contrast media reactions. AJR Am J Roentgenol 2011;196(2):W138–W143.
2. Gulani V, Calamante F, Shellock FG, Kanal E, Reeder SB; International Society for Magnetic Resonance in Medicine. Gadolinium deposition in the brain: summary of evidence and recommendations. Lancet Neurol 2017;16(7):564–570.
3. McDonald RJ, McDonald JS, Kallmes DF, et al. Intracranial gadolinium deposition after contrast-enhanced MR imaging. Radiology 2015;275(3):772–782.
4. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. Med Phys 2017;44(4):1408–1419.
5. Lei Y, Harms J, Wang T, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. Med Phys 2019;46(8):3565–3581.
6. Tanenbaum LN, Tsiouris AJ, Johnson AN, et al. Synthetic MRI for clinical neuroimaging: results of the Magnetic Resonance Image Compilation (MAGiC) prospective, multicenter, multireader trial. AJNR Am J Neuroradiol 2017;38(6):1103–1110.
7. Guo J, Gong E, Fan AP, Goubran M, Khalighi MM, Zaharchuk G. Predicting 15O-water PET cerebral blood flow maps from multi-contrast MRI using a deep convolutional neural network with evaluation of training cohort bias. J Cereb Blood Flow Metab 2020;40(11):2240–2253.
8. Kleesiek J, Morshuis JN, Isensee F, et al. Can virtual contrast enhancement in brain MRI replace gadolinium?: a feasibility study. Invest Radiol 2019;54(10):653–660.
9. Narayana PA, Coronado I, Sujit SJ, Wolinsky JS, Lublin FD, Gabr RE. Deep learning for predicting enhancing lesions in multiple sclerosis from noncontrast MRI. Radiology 2020;294(2):398–404.
10. Sun H, Liu X, Feng X, et al. Substituting gadolinium in brain MRI using DeepContrast. In: Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). Piscataway, NJ: Institute of Electrical and Electronics Engineers, 2020; 908–912.
11. Calabrese E, Villanueva-Meyer JE, Cha S. A fully automated artificial intelligence method for non-invasive, imaging-based identification of genetic alterations in glioblastomas. Sci Rep 2020;10(1):11852.
12. Gorgolewski K, Burns CD, Madison C, et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. Front Neuroinform 2011;5:13.
13. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. Neuroimage 2012;62(2):782–790.
14. Andersson JLR, Graham MS, Zsoldos E, Sotiropoulos SN. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. Neuroimage 2016;141:556–572.
15. Andersson JLR, Sotiropoulos SN. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. Neuroimage 2016;125:1063–1078.
16. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 2011;54(3):2033–2044.
17. Fonov VS, Evans AC, McKinstry RC, Almli CR, Collins DL. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. Neuroimage 2009;47(Suppl 1):S102.
18. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 2010;29(6):1310–1320.
19. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. ArXiv1505.04597 [preprint] https://arxiv.org/abs/1505.04597. Posted May 18, 2015. Accessed November 11, 2020.
20. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. ArXiv 1512.03385 [preprint] https://arxiv.org/abs/1512.03385. Posted December 10, 2015. Accessed November 11, 2020.
21. Kingma DP, Ba J. Adam: a method for stochastic optimization. ArXiv 1412.6980 [preprint] https://arxiv.org/abs/1412.6980. Posted December 22, 2014. Accessed November 11, 2020.
22. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 2004;13(4):600–612.
23. Morley SK, Brito TV, Welling DT. Measures of model performance based on the log accuracy ratio. Space Weather 2018;16(1):69–88.
24. Wang G, Li W, Ourselin S, Vercauteren T. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: Crimi A, Bakas S, Kuijf H, Menze B, Reyes M, eds. Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries. Vol 10670, Lecture Notes in Computer Science. Cham, Switzerland: Springer, 2018; 178–190.
25. Bakas S, Reyes M, Jakab A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. ArXiv 1811.02629 [preprint] https://arxiv.org/abs/1811.02629. Posted November 5, 2018. Accessed November 11, 2020.
26. Bakas S, Akbari H, Sotiras A, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data 2017;4(1):170117.
27. Menze BH, Jakab A, Bauer S, et al. The Multimodal Brain Tumor Image Segmentation benchmark (BRATS). IEEE Trans Med Imaging 2015;34(10):1993–2024.
28. Earnest F 4th, Kelly PJ, Scheithauer BW, et al. Cerebral astrocytomas: histopathologic correlation of MR and CT contrast enhancement with stereotactic biopsy. Radiology 1988;166(3):823–827.
29. Kleesiek J, Petersen J, Döring M, et al. Virtual raters for reproducible and objective assessments in radiology. Sci Rep 2016;6(1):25007.
30. Bagley LJ, Grossman RI, Judy KD, et al. Gliomas: correlation of magnetic susceptibility artifact with histologic grade. Radiology 1997;202(2):511–516.
31. Noguchi T, Yoshiura T, Hiwatashi A, et al. Perfusion imaging of brain tumors using arterial spin-labeling: correlation with histopathologic vascular density. AJNR Am J Neuroradiol 2008;29(4):688–693.
32. Clement P, Mutsaerts H-J, Václavů L, et al. Variability of physiological brain perfusion in healthy subjects: a systematic review of modifiers—considerations for multi-center ASL studies. J Cereb Blood Flow Metab 2018;38(9):1418–1437.