

Does Generative Adversarial Network Ensure Diversity in Data? Morphologic Evaluation of Synthetic Isocitrate Dehydrogenase-Mutant Glioblastomas in A Clinical Diagnostic Model

Ji Eun Park

Asan Medical Center

Dain Eun

University of Ulsan

Ho Sung Kim (✉ radhskim@gmail.com)

Asan Medical Center

Da Hyun Lee

Asan Medical Center

Ryoung Woo Jang

University of Ulsan

Namkug Kim

Asan Medical Center

Research Article

Keywords: isocitrate dehydrogenase, glioblastoma, generative adversarial network, prediction

Posted Date: December 28th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-126408/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Generative adversarial network (GAN) creates synthetic images to increase data quantity, but whether GAN ensures diversity is still unknown. We investigated whether GAN-based synthetic images provide sufficient morphologic variability to improve molecular-based prediction, as a rare disease of isocitrate dehydrogenase (IDH)-mutant glioblastomas. GAN was initially trained on 500 normal brains and 110 IDH-mutant high-grade astrocytomas, and paired contrast-enhanced T1-weighted and FLAIR MRI data were generated. Diagnostic models were developed from real IDH-wild type ($n = 80$) with real IDH-mutant glioblastomas ($n = 38$), or with synthetic IDH-mutant glioblastomas, or augmented by adding both real and synthetic IDH-mutant glioblastomas. Turing tests showed synthetic data showed reality (classification rate of 55%). Both the real and synthetic data showed that a more frontal or insular location (odds ratio [OR], 1.34 vs. 1.52; $P = .04$) and distinct non-enhancing tumor margins (OR, 2.68 vs. 3.88; $P < .001$), which become significant predictors of IDH-mutation. In an independent validation set, diagnostic accuracy was higher for the augmented model (90.9% [40/44] and 93.2% [41/44] for each reader, respectively) than for the real model (84.1% [37/44] and 86.4% [38/44] for each reader, respectively). The GAN-based synthetic images yield morphologically variable, realistic-seeming IDH-mutant glioblastomas. GAN will be useful to create a realistic training set in terms of diversity and quality, thereby improving diagnostic performance in a clinical model.

Introduction

Isocitrate dehydrogenase (IDH) mutation status of gliomas is a very important prognostic, diagnostic, and therapeutic biomarker¹. Although the frequency of IDH mutation in primary glioblastoma is low (~ 8%)^{1,2}, noninvasive imaging-based determination of IDH mutation status can predict response to anti-IDH treatment or vaccination³⁻⁶. In addition, radiologic suspicion of IDH-wild type may predict prognosis in patients with inoperable tumors⁵. Magnetic resonance imaging (MRI) has been shown to distinguish between tumors with wild-type and mutant IDH, but these studies have focused primarily on grade II/III gliomas⁷⁻¹⁰ or included a very limited number of IDH-mutant glioblastomas^{11,12} for visual analysis or deep learning. A multicenter cohort study of 496 patients with glioblastoma showed IDH mutation in 31 (6.3%)¹¹, limiting the ability of MRI to train a network to reliably predict IDH mutation status. In consequence, most studies seeking to improve the noninvasive identification of this subtype have lacked sufficient statistical power.

Although training data can be augmented by introducing random alterations, including rotations, translations, shearing, zooming, and flipping¹³, these manipulations do not generate "new" training data or address the overfitting problem¹⁴. A generative adversarial network (GAN), which can generate realistic imaging data that differ from the original data but are similar in distribution, may constitute an attractive solution¹³. GAN has been applied to MRI data to reduce noise¹⁵, but the ability of GAN-based synthetic data to model the morphologic characteristics of real tumors and their use in diagnostic models have not been tested to date. If GAN-generated imaging data reflect the morphologic characteristics of glioblastomas with mutant IDH, while varying in morphologic distribution, then these GAN-generated data can be used for training on future deep learning tasks.

To determine whether GAN-produced images reflect the morphologic characteristics of actual tumors, enabling their use as a future training set, a diagnostic model was created from the morphologic characteristics of actual and synthetic data. This model was used to determine whether the synthetic images affect performance and could be validated in an independent dataset. The purpose of this study was to investigate whether GAN-based generated IDH-mutant glioblastomas provide morphologic variability and improve molecular prediction of the IDH status of glioblastomas.

Materials And Methods

This study is reported in accordance with the Standards for Reporting of Diagnostic Accuracy Studies (STARD) 2015 guidelines¹⁶. The study protocol was approved by the institutional review board of Asan Medical Center, a tertiary referral hospital, which waived the requirement for informed consent because of the retrospective nature of the study.

Study Population

The study population consisted of a cohort of consecutive patients with histopathologically confirmed glioblastoma who underwent brain MRI from May 2017 to May 2020 (Fig. 1). Patients were included if they were histopathologically diagnosed with glioblastoma and their IDH mutation status was known, according to WHO 2016 criteria¹. A total of 214 patients met the inclusion criteria. Patients were excluded if (a) pre-operative contrast-enhanced T1-weighted imaging or fluid-attenuated inversion recovery imaging was not performed ($n = 14$), or (b) they had history of previous surgery ($n = 38$). The study population consisted of 162 patients, 65 men and 97 women, of mean \pm standard deviation age (SD) 56 ± 10.7 years, with 118 patients who underwent brain MRI from May 2017 to January 2019 assigned to the training set, and 44 patients who underwent brain MRI from February 2019 to May 2020 assigned to the validation set.

IDH Mutation Status

IDH mutation status was analyzed by members of the pathology division of our hospital who were blinded to the radiologic results. The reference standard consisted of immunohistochemical determination of IDH1 (R132H) protein expression¹⁷. Mutations in the *IDH1* and *IDH2* genes were determined by DNA pyrosequencing at diagnosis¹⁷.

All patients were tested for 1p/19q co-deletion status and found the 1p/19q co-deletion was negative, indicating astrocytomas.

Imaging Data Acquisition

All enrolled patients underwent MRI on a 3.0 T unit (Achieva or Ingenia, Philips Medical Systems) using a 16-channel or 32-channel head coil. The MRI protocols included T2-weighted imaging, fluid-attenuated inversion recovery (FLAIR) imaging, T1-weighted imaging, and contrast-enhanced T1-weighted imaging. The contrast-enhanced T1-weighted (CE-T1w) images were obtained at a high-resolution three-dimensional (3D) volume, using a gradient-echo T1-weighted sequence with the following parameters: repetition time (TR)/echo time (TE), 9.8/4.6 ms; flip angle, 10°; field of view (FOV), 256 mm; matrix, 512 × 512; and slice thickness, 1 mm with no gap. The parameters for FLAIR imaging included TR/TE, 9000/135 ms; flip angle, 90°; FOV, 240 mm; matrix, 512 × 512; and slice thickness, 4 mm with no gap.

Image Preprocessing

To prepare the training data, both CE-T1w and FLAIR images were subjected to skull stripping using HD-BET algorithms¹⁸. Each FLAIR image was co-registered to the corresponding CE-T1w image by within-subject registration using a rigid-body model, image reslicing, and SPM12 software¹⁹. The CE-T1w, FLAIR, and null images were combined into a three-channel image. A total of 19,595 three-channel images from 110 IDH-mutant patients were fed into the style-based GAN architecture (StyleGAN2) network to simultaneously generate synthetic IDH-mutant CE-T1w and FLAIR images.

Theory

GANs have been shown to generate realistic images from latent vectors. Although the latent vector sampled from a uniform distribution is traditionally provided to the GAN generator network^{20,21}, this approach leads to an unavoidable feature entanglement. Because feature disentangling is required for smooth image generation, StyleGAN first introduced the mapping network, $f: Z \rightarrow W$, which transforms latent $z \in Z$ from a uniform distribution to the intermediate latent vector $w \in W$. StyleGAN also successfully introduced adaptive instance normalization (AdaIN) to the generator network, enabling the computation of the invariant style y from the intermediate latent vector w .

Following the success of StyleGAN, StyleGAN2 further improved image-generation quality by redesigning the generator architecture, reducing the common artifacts observed in StyleGAN-generated images. The performance of the StyleGAN2 synthesis network g was improved by introducing several modifications (**Supplementary Fig. 1**). The applications of bias, noise, and normalization to the constant input at the beginning of the network architecture were removed. Then, bias and noise operations were added outside the styleblock. The AdaIN operation was divided into modulation and demodulation operations. The modulation operation scaled each input feature map of the convolution by its scaling value, which was determined by the incoming style. The demodulation operation normalized each output feature map to the L2 norm of each output channel. With these modifications, StyleGAN2 successfully removed common artifacts that were commonly observed in StyleGAN²².

Contrast-enhanced T1-Weighted and FLAIR Cogeneration and StyleGAN2 Implementation Details

Although the generation of multi-modality images is considered favorable, most medical image synthesis studies have focused only on the generation of single-modality images^{23–26}. By combining CE-T1w and FLAIR images into multichannel images, StyleGAN2 generated CE-T1w and FLAIR images simultaneously (**Supplementary Fig. 2**).

The sizes of the input latent vector z and the intermediate latent vector w were each set at 1×512 . The output image size was $3 \times 256 \times 256$; the first channel was the CE-T1w image, the second channel was the FLAIR image, and the last channel was the null image. The mapping network consisted of eight fully connected layers. Leaky ReLU activation with $\alpha = 0.2$ was used for activation function and bilinear filtering for all up and down sampling layers. The learning rate was set at 2×10^{-3} . An Adam optimizer was used with hyperparameters $\beta_1 = 0$, $\beta_2 = 0.99$, $\epsilon = 10^{-8}$ and minibatch size 32. The network was trained on a NVIDIA TITAN RTX 24 GB GPU. The training of 80,000 images took approximately 25 minutes, and the generation of 100 synthetic images took approximately 8 seconds. The network was iteratively trained for 4 million images. The code was modified from the original paper²², which is available at <https://github.com/NVlabs/stylegan2>. All experiments were implemented with the official tensorflow code of StyleGAN2 provided by the NVIDIA Corporation.

Sample Size and Rationale for the Training Network

StyleGAN2 was initially developed to train data using 500 datasets of normal appearing brain MRI, obtained from 393 men and 107 women of mean \pm SD age 49.4 ± 12.1 years. These datasets included contrast-enhanced T1-weighted and FLAIR images that were obtained for evaluation of brain metastases in patients with lung cancer, with all patients diagnosed as negative for metastases in brain parenchyma. The images created from StyleGAN2 were reviewed by two experts (J.E.P. and H.S.K., with 5 and 20 years of experience, respectively, in neuro-oncologic imaging). These evaluations confirmed that the generated imaging data yielded realistic images without artifacts.

The sample size was set at 100 for the training network to provide realistic data. Thus, synthetic data for IDH-mutant glioblastomas were generated from a dataset consisting of images of 110 patients, 57 men and 53 women, of mean \pm SD age 54 ± 12.3 years, with WHO grades III and IV IDH-mutant high-grade astrocytomas, including 49 IDH-mutant glioblastomas. The synthetic imaging data reflected the morphologic features of IDH-mutant type astrocytomas, as shown in **Supplementary Fig. 3**.

Imaging Analysis

Training was continued until the two expert radiologists found it difficult to distinguish between real and synthetic data.

Evaluation of Reality

Turing tests of each dataset were performed independently by the two observers 2 weeks before morphologic assessment. The evaluation was binary, with a score of 0 indicating that the data appeared fake and seemed to consist of GAN-generated synthetic data, whereas a score of 1 indicated that the data

appeared real²⁷. The correct classification rate and misclassification rates were calculated.

Morphologic Assessment

A radiologist (H.S.K., with 22 years of experience in neuroradiology) who did not participate in any other image review in this study selected single 2D FLAIR-weighted and contrast-enhanced T1-weighted images to be reviewed, with real and synthetic imaging data randomly shuffled. Two observers (J.E.P. and D.L., with 5 and 1 years of experience, respectively, as board-certified neuroradiologists) independently reviewed 200 MRI datasets, while being blinded to diagnosis and the evaluations of other observers. Feature categories were adapted from previous studies of IDH mutations in WHO grade II/III gliomas⁸⁻¹⁰. Tumor location was specified by epicenter, with locations grouped according to the frequency of IDH mutation, thereby reducing the number of variables for statistical analysis. The locations included the frontal or insular cortex, the thalamus or brainstem, and others. Patterns of contrast enhancement included rim enhancement surrounding central necrosis, nodular enhancement, and partial patchy enhancement. The areas surrounding regions of high signal intensity on non-enhancing FLAIR images were recorded as tumor dominant or edema dominant, and the margins surrounding these regions as clear or indistinct. Representative cases generated from synthetic data are shown in Fig. 2.

Statistical Analysis

Distribution of Morphologic Features

Continuous variables across datasets were compared by Student's *t*-tests, and categorical variables were compared by chi-square tests. All statistical analyses were performed using R software (version 3.6.1), with *P*-values < .05 regarded as statistically significant.

Significant Predictors for IDH Mutation

Inter-observer agreement on morphologic categories was evaluated by Cohen κ testing. Values of < 0 indicated no agreement, whereas values of 0–0.20, 0.21–0.40, 0.41–0.60, 0.61–0.80, and 0.81–1.0 indicated slight, fair, moderate, substantial, and almost perfect agreement, respectively. Morphologic categories with κ values \geq 0.5 were subject to univariable analysis. Discordant morphologic categories were subsequently resolved by consensus for variables in the model.

Univariate logistic regression analyses were performed to test whether morphologic criteria could predict IDH mutation status. Nagelkerke (Pseudo) R^2 was used as a summary statistic to determine the degree to which the overall model predicted the variation in IDH mutation positivity. Parameters significant in univariable analysis, defined as those with $P < 0.05$, were subsequently entered into the multivariable analysis. Multivariable binomial logistic regression was performed to predict IDH-mutant vs. IDH-wild type glioblastoma using a stepwise elimination process. Models were built separately for real IDH-wild type and IDH-mutant data ($n = 118$, model 1), real IDH-wild type and synthetic IDH-mutant data ($n = 118$, model 2), and real IDH-wild type, real IDH-mutant, and synthetic IDH-mutant data ($n = 156$, model 3).

Diagnostic Performance for IDH Mutation

Using the results from the multivariable regression analysis for each model, the probability of IDH mutation positive status was calculated for individual patients in the validation set. The diagnostic performance of the multivariable model was determined by calculating the area under the receiver operating characteristics (ROC) curve, with the diagnostic threshold determined using the Youden index. The three above models were compared.

Additionally, univariate logistic regression analysis was performed to determine whether age could predict IDH mutation status. The age-based prediction was subsequently combined with the image-based prediction using a logistic regression classifier in the training set with real data (model 1) and in the validation set.

Results

Patient Demographics

This study included 162 patients, consisting of 65 men and 97 women of mean \pm SD age 56 ± 10.7 years. Of these, 118 patients were included in the training set and 44 patients in the validation set. Patients with IDH-wild type were significantly older than patients with IDH-mutant type glioblastoma, both in the training set (57.4 vs. 46.3 years; $P < .001$) and in the validation set (56.4 vs. 41.6 years; $P < .001$). Table 1 shows the demographic characteristics of these patients, as well as the imaging characteristics of the real and synthetic datasets. A video (**Online Supplement**) shows continuous synthetic tumor on contrast-enhancing T1-weighted and FLAIR images.

Table 1
Clinical and Imaging characteristics of the study patients

Variables	Training Set			Validation Set					
	IDH-wild	IDH-mutant	IDH-mutant (GAN)	<i>P+</i>	<i>P*</i>	<i>P**</i>	IDH-wild	IDH-mutant	<i>P</i>
No. of patients	80	38	38				33	11	
Mean age (y)	57.4	46.3	-	-	< .001	< .001	56.4	41.6	< .001
Enhancement category				.01	< .001	.002			.01
Rim enhancing	47	19	9				24	3	
Thick nodular	28	8	5				6	3	
Patch enhancing	5	11	24				3	5	
Tumor location				.55	.01	.008			.36
Frontal or insula	29	25	25				16	8	
Other	42	10	12				13	2	
Thalamus or brainstem	9	3	1				4	1	
Necrosis				.35	.001	< .001			.007
Yes	72	23	19				30	6	
No	8	15	19				3	5	
Surrounding high signal intensity				.39	.001	< .001			< .001
Tumor dominant	35	29	32				8	11	
Edema dominant	45	9	6				25	0	
Margin of non-enhancing lesion				.10	< .001	< .001			< .001
Clear	4	20	27				0	6	
Indistinct	76	18	11				33	5	

Note: Data are expressed as the mean ± standard deviation. Abbreviation: IDH = isocitrate dehydrogenase. *P+* indicates differences between real and synthetic IDH-mutant data. *P** and *P*** indicate differences between real IDH-wild type and real IDH-mutant data and between real IDH-wild type and synthetic IDH-mutant data, respectively.

Evaluation of Reality

Turing tests of the imaging data showed that the correct classification rates by readers 1 and 2 were 55% and 62%, respectively, showing that it was difficult to distinguish between real and synthetic data. Reader 1 misclassified 22 real images as synthetic, while misclassifying 23 synthetic images as real. Reader 2 misclassified 20 real images as synthetic, while misclassifying 18 synthetic images as real. Examples of synthetic data correctly classified as synthetic are shown in Fig. 3.

Distribution of Morphologic Features

A comparison of imaging data of real and synthetic IDH-mutant glioblastomas showed no differences in tumor location (χ^2 test, $P = .55$), degree of necrosis ($P = .35$), and tissue ($P = .39$) and margins ($P = .10$) surrounding regions of high signal intensity. The patch enhancing pattern was observed more frequently in the synthetic than in the real imaging data ($P = .01$). Frontal or insular location was significantly more frequent in both patient ($P = .01$) and synthetic ($P = .008$) data in the training set, but not in the validation set.

Compared with imaging of IDH-wild type glioblastoma, imaging of IDH-mutant type glioblastoma showed that rim enhancing pattern was less frequent in both patients (highest $P = .01$) and in the synthetic dataset ($P = .002$). Similarly, internal necrosis was significantly less frequent in IDH-mutant than in IDH-wild type in both patients (highest $P = .001$) and in the synthetic dataset ($P < .001$). By contrast, distinct margins surrounding areas of high intensity were significantly more common in IDH-mutant than in IDH-wild type in the patients (highest $P = .001$) and in the synthetic dataset ($P < .002$).

Significant Predictors of IDH Mutation

The two readers showed moderate agreement regarding tumor location ($\kappa = 0.67$, $P < .001$), patterns of enhancement ($\kappa = 0.67$, $P < .001$), presence of necrosis ($\kappa = 0.65$, $P < .001$), and margins of non-enhancing lesions ($\kappa = 0.56$, $P < .001$).

Table 2 shows the results of univariable and multivariable logistic regression analyses. Multivariable analysis showed that, in both real and synthetic data, a more frontal or insular location ($\beta = 1.34$, $P = .02$ for real data; $\beta = 1.52$, $P = .04$ for synthetic data) and distinct margins of non-enhancing tumors ($\beta = 2.68$, $P < .001$ for real data; $\beta = 3.88$, $P < .001$ for synthetic data) were significant predictors of IDH mutation. Univariate analysis showed that absence of necrosis and presence of a patch enhancing pattern in both real and synthetic data were significant, whereas the multivariable model showed that the absence of necrosis was significant only for real data ($\beta = 1.91$, $P = .02$), and the presence of a patch enhancing pattern was significant only for synthetic data ($\beta = 3.46$, $P = .002$).

Table 2
Univariable and Multivariable Binomial Logistic Regression Analysis of Factors Predicting IDH mutation in the Training Dataset

Variables	With Real Data (Model 1) (n = 118)				With IDH-mutant Synthetic Data (Model 2) (n = 118)				Augmented with Real and Synthetic Data (Model 3) (n = 156)			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	Beta coefficient	P	Beta coefficient	P	Beta coefficient	P	Beta coefficient	P	Beta coefficient	P	Beta coefficient	P
Tumor location												
Other	Ref		Ref		Ref		Ref		Ref		Ref	
Frontal or insula	1.29	.004	1.34 (1.24, 1.45)	.02	1.10	.009	1.52 (1.04, 2.35)	.04	1.19	< .001	1.32 (1.24, 1.4)	.008
Thalamus or brainstem	0.95	.18	0.25 (0.09, 0.41)	.77	2.05	.06	3.33 (2.89, 3.77)	.77	1.35	.03	0.78 (0.64, 0.91)	.36
Absence of necrosis	1.77	< .001	1.91 (1.77, 2.05)	.017	2.19	< .001	-0.06 (-0.24, 0.12)	.95	1.98	< .001	1.13 (-0.52, 0.78)	.08
Enhancement category												
Rim enhancing	Ref		Ref		Ref				Ref			
Thick nodular	-0.35	.47	-1.50 (-1.64, -1.36)	0.05	-0.07	.90	0.25 (0.086, 0.414)	.78	-0.25	.54	-0.58 (0.66, -0.49)	.27
Patch enhancing	1.69	.005	0.035 (-0.135, 0.205)	0.97	3.22	< .001	3.46 (3.26, 3.66)	.002	2.46	< .001	1.97 (1.87, 2.07)	.002
Edema-dominant surrounding high signal intensity	-0.91	.04	-0.54 (-0.646, -0.434)	.35	-3.56	.001	-1.29 (-1.5, -1.08)	.26	-1.62	.001	-0.83 (-0.91, -0.75)	.11
Margin of non-enhancing lesions												
Indistinct	Ref		Ref		Ref		Ref		Ref		Ref	
Distinct	2.94	< .001	2.68 (2.56, 2.8)	< .001	3.84	< .001	3.88 (3.04, 4.72)	< .001	3.47	< .001	2.96 (2.86, 3.06)	< .001

Note.—Data in parentheses are 95% confidence intervals.

Diagnostic Performance for IDH Mutation

The results of diagnostic performance are shown in Table 3. The synthetic model (AUC, 0.96; 95% CI, 0.90–0.99) showed higher diagnostic performance than the real model (AUC, 0.86; 95% CI, 0.80–0.92) in the training set. In the validation set, the diagnostic performance was similar in both the real and synthetic model, with readers 1 and 2 showing AUCs of 0.71 (95% CI, 0.54–0.89) and 0.77 (95% CI, 0.56–0.98), respectively, for the real model, and AUCs of 0.75 (95% CI, 0.52–0.98) and 0.77 (95% CI, 0.56–0.98), respectively, for the synthetic model.

Table 3
Diagnostic performance of the Models for Prediction of IDH mutation

	Model from Real Data (Model 1)				Model from IDH-mutant Synthetic Data (Model 2)				Combined with Real and Synthetic Data (Model 3)			
	AUC	95% CI	Sensitivity	Specificity	AUC	95% CI	Sensitivity	Specificity	AUC	95% CI	Sensitivity	Specificity
Training set												
	0.864	0.789, 0.920	57.9%	95.0%	0.958	0.904, 0.986	84.2%	92.5%	0.899	0.841, 0.942	72.37%	91.25%
Validation set												
Reader 1	0.713	0.535, 0.892	36.4%	100%	0.747	0.517, 0.978	63.6%	100%	0.747	0.517, 0.978	63.6%	100%
Reader 2	0.773	0.565, 0.981	62.7%	63.6%	0.773	0.565, 0.981	62.7%	63.6%	0.821	0.653, 0.989	63.6%	93.9%
Combined with age information												
Reader 1	0.871	0.722, 1.00	63.6%	100%	0.826	0.662, 0.991	63.6%	100%	0.826	0.662, 0.991	63.6%	100%
Reader 2	0.855	0.710, 1.00	78.8%	81.8%	0.861	0.720, 1.00	69.7%	90.9%	0.861	0.720, 1.00	69.7%	90.9%
AUC = area under the receiver operating characteristics curve												

Effect of Data Augmentation

Use of an augmented model, in which synthetic data were added to real data, showed the same predictors of IDH-mutant as the synthetic model, with a multivariable analysis showing that a more frontal or insular location ($\beta = 1.32, P = .01$), the presence of a patch enhancing pattern ($\beta = 1.97, P = .002$), and distinct margins of non-enhancing tumors ($\beta = 2.96, P < .001$) were statistically significant. In the training set, the augmented model had a diagnostic performance (AUC, 0.90; 95% CI, 0.84–0.94) slightly higher than that of the real model (AUC, 0.86) and slightly lower than that of the synthetic model (AUC, 0.96). In the validation set, the augmented model showed slightly higher diagnostic performance (AUC, 0.75 for reader 1 and 0.82 for reader 2) than the synthetic or real model. The augmented model had greater diagnostic accuracy (90.9% [40/44] and 93.2% [41/44] for readers 1 and 2, respectively) than the real model (84.1% [37/44] and 86.4% [38/44] for readers 1 and 2, respectively).

Discussion

This study found that the morphologic characteristics exhibited by synthetic and real imaging data of IDH-mutant glioblastomas were generally similar, with the two datasets being similar in tumor location, margins, type of tissue surrounding areas of high signal intensity, and presence of necrosis, but not in contrast-enhancing patterns. Univariable analysis showed that the same morphologic characteristics, including tumor location, absence of necrosis, enhancement category, and margins and type of tissue surrounding non-enhanced regions, were predictive of IDH mutation in both the real and synthetic datasets. A multivariable diagnostic model derived from synthetic data showed higher predictive performance than a model derived from real data in the training set, with the two models having similar predictive performance in the independent validation set. Thus, the morphologic variability of GAN-based synthetic images of IDH-mutant glioblastomas was similar to that of actual images, suggesting that the former may serve as a realistic training set.

Models have shown the ability to distinguish between IDH-mutant and IDH-wild type gliomas with AUCs of 0.80–0.94^{8–10,28}. Based on the prevalence of IDH-mutant glioblastomas, the sample size required for sufficient training for deep learning is up to 1200 patients. This number, however, is difficult to achieve in practice and requires data augmentation. Previous studies using GAN^{23,29} showed that augmentation with synthetic data improved the diagnostic performance of the model, but those studies were more limited in that performance was measured in the training set. The performance of the synthetic and augmented models in the present study was similar to or higher than the performance of the real-data only model in both the training and validation sets. In addition, age was an important predictor of IDH mutation status, suggesting that the synthetic data generated by GAN may be useful for extracting image-based morphologic features and could be combined with age as an additional predictor.

GAN may have the ability to learn the complete distribution of data when given “sufficiently large” deep networks, sample size, and computation time³⁰. To utilize GAN to learn the characteristics of IDH-mutant glioblastoma, we first optimized the sample size for StyleGAN2, until GAN provided sufficiently realistic imaging data without artifacts. We then trained GAN with the images available for IDH-mutant high-grade astrocytomas to generate synthetic images and transfer them to IDH-mutant glioblastomas. This provided important evidence about training on a rare disease, generating certain types of images, such that style transfer could be useful for a pre-trained network to improve image quality (image reality). Subsequently, a specific outcome, such as a certain molecular subtype or diagnosis, would be appropriate in a latent space. The synthetic images created in the present study showed similar but not identical morphology to the training dataset, providing a smooth transition in the latent space³¹ with the GAN network.

Two-channel GAN was able to simultaneously generate contrast-enhanced T1-weighted (CE-T1w) and FLAIR images. This is important for GAN-based synthetic images because both images are necessary to characterize IDH mutations and may be useful for data augmentation in deep learning. Two-channel

GAN can fully determine the morphologic characteristics of conventional imaging data predictive of IDH mutation, including focal patch enhancement within areas of high signal intensity on FLAIR^{8,9}, and distinct margins of non-enhancing lesions^{8,10} determined by high signal intensity on FLAIR without contrast enhancement. Univariate analysis of all three models, the real, synthetic, and augmented models, yielded the same predictive factors, indicating that the distribution of morphologic variability was similar for real and synthetic data, and suggesting that the use of synthetic data for diagnostic training was feasible.

This study had several limitations. First, synthetic data were generated from IDH-mutant high-grade astrocytomas, not solely from glioblastomas, in which patchy enhancing patterns were more frequent. High-grade astrocytomas were included in GAN training because the IDH-mutant glioblastomas available for GAN training was small. Second, the GAN network was trained on both CE-T1w and FLAIR images; however, other diffusion-weighted or T2 images may be useful for predicting IDH mutations. For example, training of T2 images with T2-FLAIR mismatch sign^{32,33} will be helpful in predicting IDH mutations in low-grade gliomas. Third, sampling from GAN networks was random. The development of diagnostic models may depend on the sampling method. A more objective analysis requires the methodologic construction and testing of multiple diagnostic models, as well as their statistical improvement in the future.

In conclusion, the GAN-based synthetic images yielded morphologically variable, realistic but unseen IDH-mutant glioblastomas, and they were useful as realistic training sets to improve diagnostic performance. Our results provided evidence that synthetic IDH-mutant glioblastomas improved the visual diagnosis of tumors with IDH mutations and demonstrated the potential to improve noninvasive identification of IDH-mutant tumors, thus overcoming the small sample size inherent in imaging-based genomic and molecular prediction.

Declarations

Acknowledgement

This research was supported by National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (grant numbers: NRF-2020R1A2B5B01001707 and NRF-2020R1A2C4001748).

Author Contribution Statement

All authors reviewed the manuscript. J.E.P. contributed to conceptual design and writing the manuscript. D.E. contributed to generative adversarial network analysis, imaging processing, and writing the manuscript. H.S.K. contributed to editing the manuscript, conceptual design and project integrity. D.H.L. contributed to image analysis. R.W.J. contributed to database construction and data analysis. N.K. contributed to conceptual feedback, oversight, and software support.

Competing Interests

The authors declare no competing interests.

Corresponding author

Correspondence to Ho Sung Kim.

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* **131**, 803–820 (2016).
2. Ceccarelli, M. *et al.* Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* **164**, 550–563 (2016).
3. Wang, F. *et al.* Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* **340**, 622–626 (2013).
4. Shankar, G. M. *et al.* Genotype-targeted local therapy of glioma. *Proc Natl Acad Sci U S A* **115**, E8388–E8394 (2018).
5. Schumacher, T. *et al.* A vaccine targeting mutant IDH1 induces antitumour immunity. *Nature* **512**, 324–327 (2014).
6. Pusch, S. *et al.* Pan-mutant IDH1 inhibitor BAY 1436032 for effective treatment of IDH1 mutant astrocytoma in vivo. *Acta Neuropathol* **133**, 629–644 (2017).
7. Akkus, Z. *et al.* Predicting Deletion of Chromosomal Arms 1p/19q in Low-Grade Gliomas from MR Images Using Machine Intelligence. *J Digit Imaging* **30**, 469–476 (2017).
8. Kanazawa, T. *et al.* Imaging scoring systems for preoperative molecular diagnoses of lower-grade gliomas. *Neurosurg Rev* **42**, 433–441 (2019).
9. Maynard, J. *et al.* World Health Organization Grade II/III Glioma Molecular Status: Prediction by MRI Morphologic Features and Apparent Diffusion Coefficient. *Radiology* **296**, 111–121 (2020).
10. Zhou, H. *et al.* MRI features predict survival and molecular markers in diffuse lower-grade gliomas. *Neuro Oncol* **19**, 862–870 (2017).
11. Chang, K. *et al.* Residual Convolutional Neural Network for the Determination of IDH Status in Low- and High-Grade Gliomas from MR Imaging. *Clin Cancer Res* **24**, 1073–1081 (2018).

12. Korfiatis, P. & Erickson, B. Deep learning can see the unseeable: predicting molecular markers from MRI of brain gliomas. *Clinical radiology* **74**, 367–373 (2019).
13. Shorten, C. & Khoshgoftaar, T. M. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data* **6**, 60 (2019).
14. Engstrom, L. & Madry, A. A rotation and a translation suffice: Fooling cnns with simple transformations.
15. Chen, J., Chen, J., Chao, H. & Yang, M. Image blind denoising with generative adversarial network based noise modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3155–3164.
16. Cohen, J. F. *et al.* STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* **6**, e012799 (2016).
17. Weller, M. *et al.* European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. *Lancet Oncol* **18**, e315-e329 (2017).
18. Isensee, F. *et al.* Automated brain extraction of multisequence MRI using artificial neural networks. *Human brain mapping* **40**, 4952–4964 (2019).
19. Ashburner, J. & Friston, K. J. Unified segmentation. *Neuroimage* **26**, 839–851 (2005).
20. Radford, A., Metz, L. & Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434* (2015).
21. Karras, T., Aila, T., Laine, S. & Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* (2017).
22. Karras, T. *et al.* Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8110–8119.
23. Han, C. *et al.* Infinite brain MR images: PGGAN-based data augmentation for tumor detection In *Neural approaches to dynamics of signal exchanges*. 291–303 (Springer, 2020).
24. Yi, X., Walia, E. & Babyn, P. Generative adversarial network in medical imaging: A review. *Medical image analysis* **58**, 101552 (2019).
25. Beers, A. *et al.* High-resolution medical image synthesis using progressively grown generative adversarial networks. *arXiv preprint arXiv:1805.03144* (2018).
26. Han, C. *et al.* Learning more with less: Conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 119–127.
27. Geman, D., Geman, S., Hallonquist, N. & Younes, L. Visual Turing test for computer vision systems. *Proceedings of the National Academy of Sciences* **112**, 3618 (2015).
28. Kim, M. *et al.* Diffusion- and perfusion-weighted MRI radiomics model may predict isocitrate dehydrogenase (IDH) mutation and tumor aggressiveness in diffuse lower grade glioma. *Eur Radiol* **30**, 2142–2151 (2020).
29. Liu, S. *et al.* Isocitrate dehydrogenase (IDH) status prediction in histopathology images of gliomas using deep learning. *Sci Rep* **10**, 7733 (2020).
30. Goodfellow, I. *et al.* Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.
31. Wu, J., Zhang, C., Xue, T., Freeman, B. & Tenenbaum, J. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Advances in neural information processing systems*. 82–90.
32. Broen, M. P. G. *et al.* The T2-FLAIR mismatch sign as an imaging marker for non-enhancing IDH-mutant, 1p/19q-intact lower-grade glioma: a validation study. *Neuro Oncol* **20**, 1393–1399 (2018).
33. Patel, S. H. *et al.* T2-FLAIR Mismatch, an Imaging Biomarker for IDH and 1p/19q Status in Lower-grade Gliomas: A TCGA/TCIA Project. *Clin Cancer Res* **23**, 6078–6085 (2017).

Figures

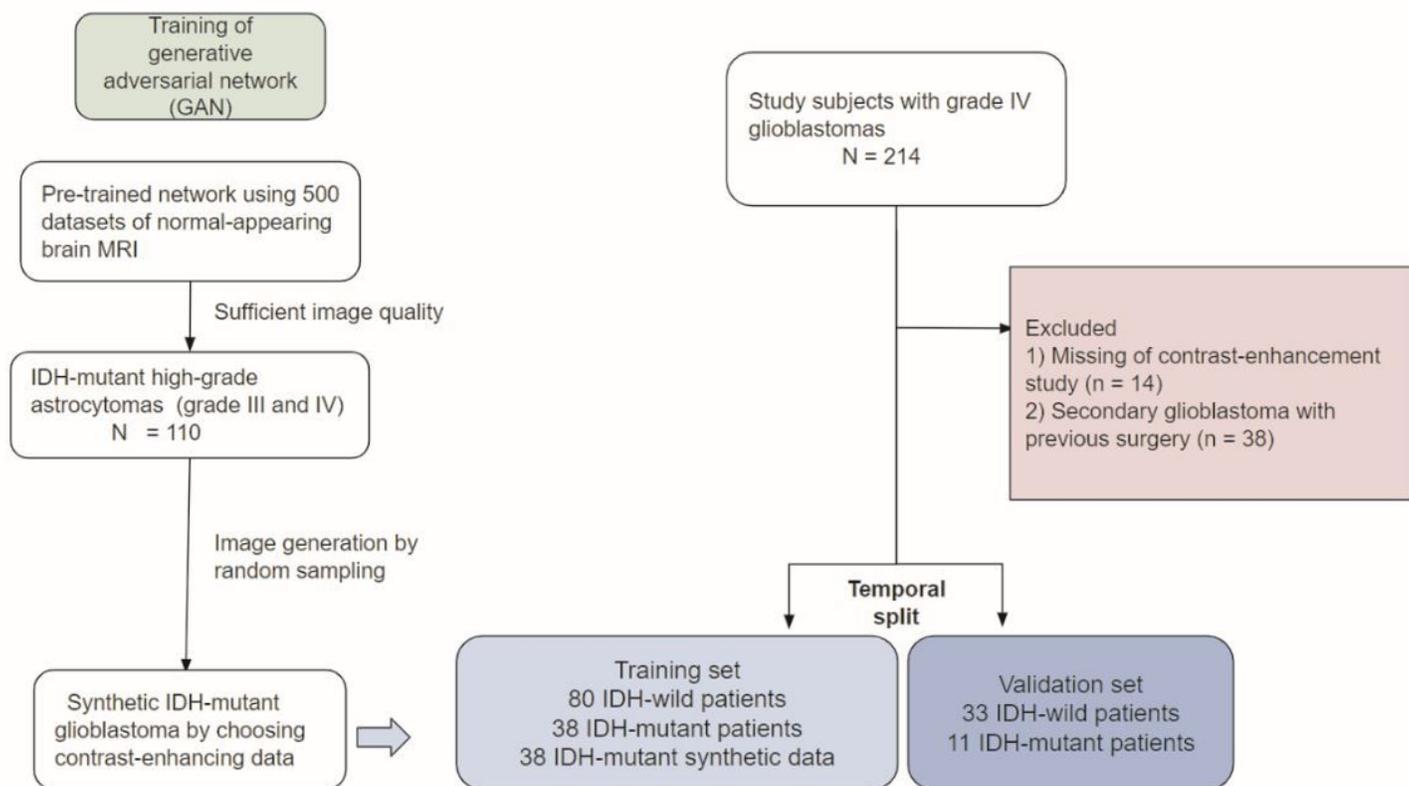


Figure 1

Process for inclusion of the study population and the training dataset for the generative adversarial network.

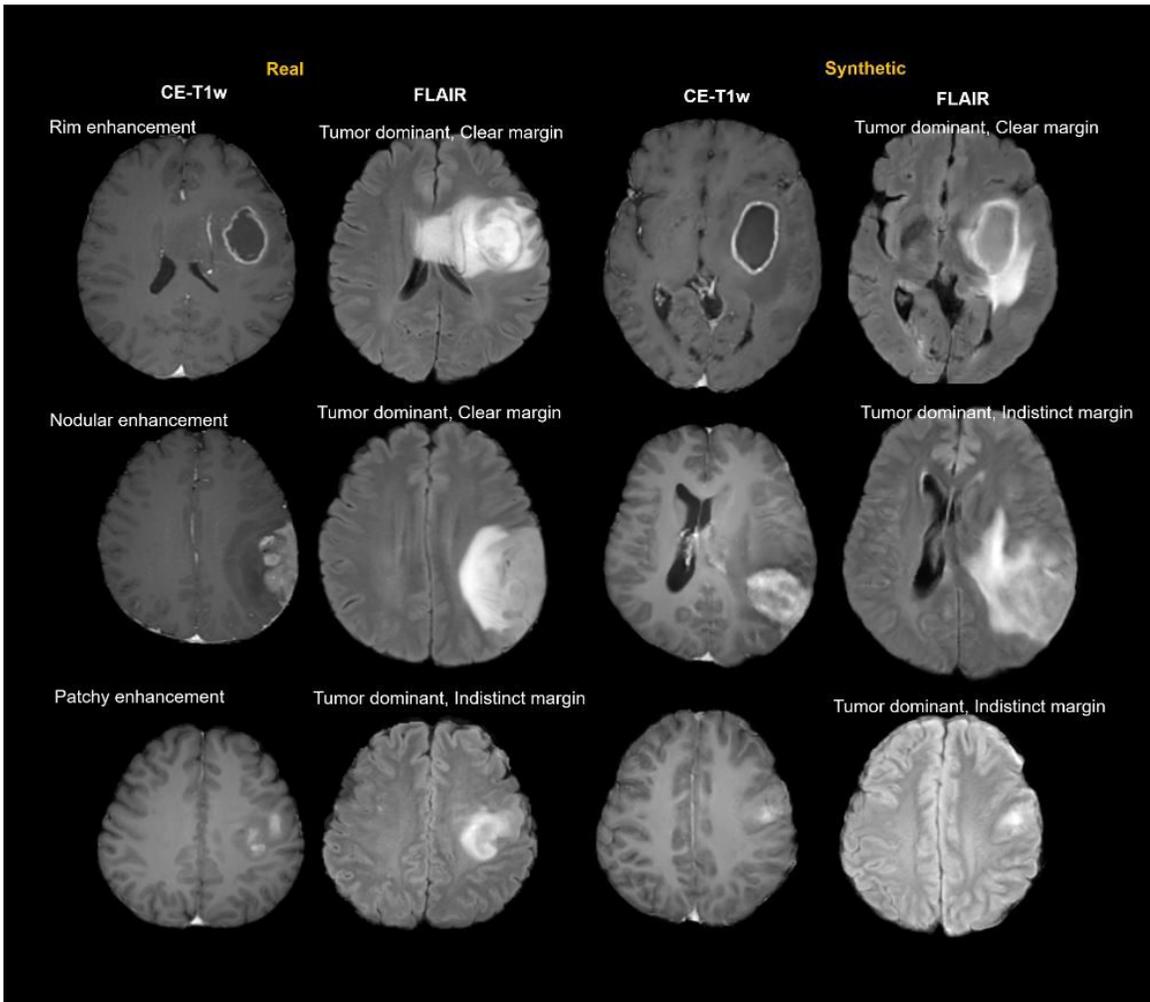


Figure 2
 Morphologic characteristics of real IDH-mutant glioblastomas (left) and synthetic IDH-mutant glioblastomas generated by a generative adversarial network (right) based on contrast-enhanced T1-weighted (CE-T1w) and paired FLAIR images. (A) CE-T1w images showing different contrast patterns of rim enhancement, thick nodular enhancement, and patch enhancement. (B) FLAIR images showing types of surrounding high signal intensity (tumor dominant and edema dominant) and margins of non-enhancing lesions (clear and indistinct). Although the appearances of synthetic images are similar to those of real images, there were no exact matches.

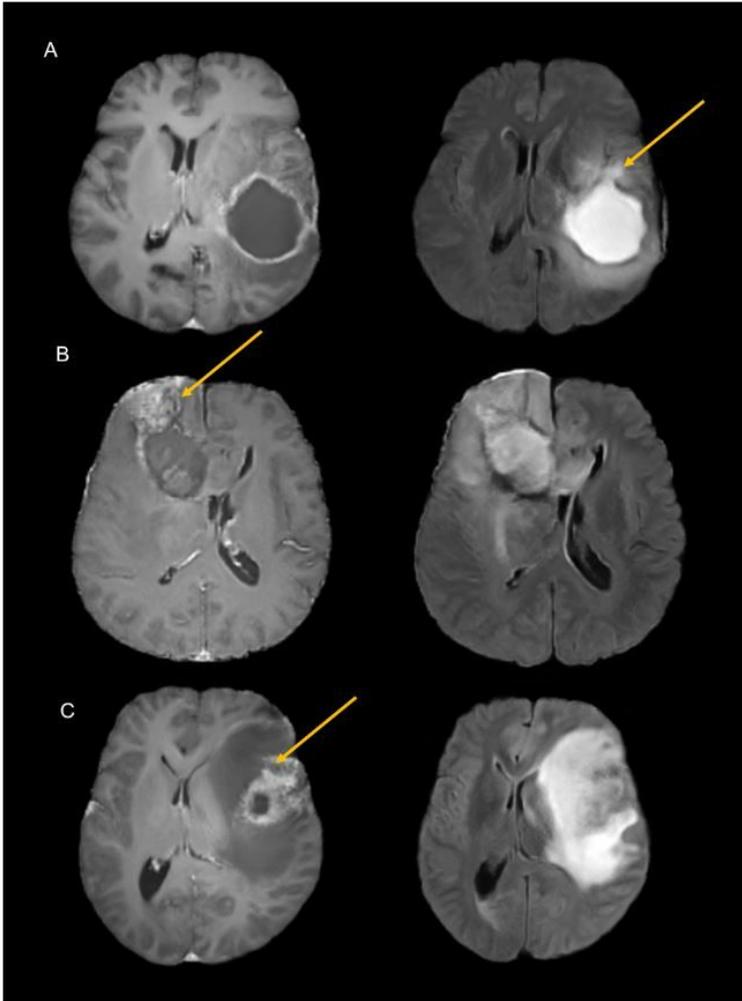


Figure 3

Representative synthetic images correctly determined to be synthetic by neuroradiologists. (A) Contrast-enhanced T1-weighted (CE-T1w) image similar to a real image, coupled with a FLAIR image showing an open rim of hypointensity, suggesting that the image was not real. (B) CE-T1w images showing nodular enhancement with a mesh-like artifact, suggesting that these images were not real. (C) CE-T1w images showing bizarre-shaped linear enhancement, suggesting that these images were not real.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SciReportsSupplementaryfinal.docx](#)