Neuro-Oncology Advances

7(1), vdaf001, 2025 | https://doi.org/10.1093/noajnl/vdaf001 | Advance Access date 16 January 2025

Synthetic O-(2-^{18}F-fluoroethyl)-L-tyrosine-positron emission tomography generation and hotspot prediction via preoperative MRI fusion of gliomas lacking radiographic high-grade characteristics

Eric Suero Molina[†], Mehnaz Tabassum[†], Ghasem Azemi, Zeynep Özdemir, Wolfgang Roll, Philipp Backhaus, Philipp Schindler, Alex Valls Chavarria, Carlo Russo, Sidong Liu, Walter Stummer and , Antonio Di Ieva

All author affiliations are listed at the end of the article

Corresponding Author: Eric Suero Molina, MBA, FEBNS, Department of Neurosurgery, University Hospital Münster, Albert-Schweitzer-Campus 1, A1, D-48149 Münster, Germany (eric.suero@ukmuenster.de).

[†]These authors contributed equally.

Abstract

Background. Limited amino acid availability for positron emission tomography (PET) imaging hinders therapeutic decision-making for gliomas without typical high-grade imaging features. To address this gap, we evaluated a generative artificial intelligence (AI) approach for creating synthetic O-(2-^{18}F-fluoroethyl)-L-tyrosine ([^{18}F]FET)-PET and predicting high [^{18}F]FET uptake from magnetic resonance imaging (MRI).

Methods. We trained a deep learning (DL)-based model to segment tumors in MRI, extracted radiomic features using the Python PyRadiomics package, and utilized a Random Forest classifier to predict high [^{18}F]FET uptake. To generate [^{18}F]FET-PET images, we employed a generative adversarial network framework and utilized a split-input fusion module for processing different MRI sequences through feature extraction, concatenation, and self-attention.

Results. We included magnetic resonance imaging (MRI) and PET images from 215 studies for the hotspot classification and 211 studies for the synthetic PET generation task. The top-performing radiomic features achieved 80% accuracy for hotspot prediction. From the synthetic [^{18}F]FET-PET, 85% were classified as clinically useful by senior physicians. Peak signal-to-noise ratio analysis indicated high signal fidelity with a peak at 40 dB, while structural similarity index values showed structural congruence. Root mean square error analysis demonstrated lower values below 5.6. Most visual information fidelity scores ranged between 0.6 and 0.7. This indicates that synthetic PET images retain the essential information required for clinical assessment and diagnosis.

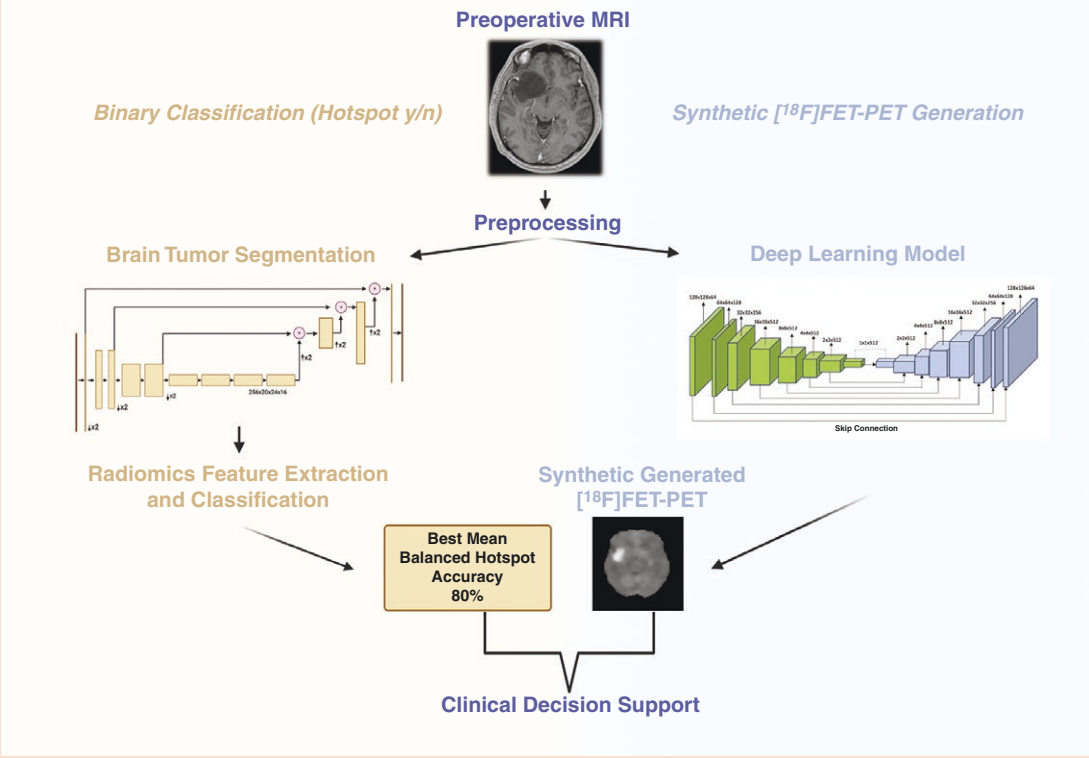
Conclusion. For the first time, we demonstrate that predicting high [^{18}F]FET uptake and generating synthetic PET images from preoperative MRI in lower-grade and high-grade glioma are feasible. Advanced MRI modalities and other generative AI models will be used to improve the algorithm further in future studies.

Key Points

- Evaluated generative artificial intelligence for synthetic O-(2-^{18}F-fluoroethyl)-L-tyrosine-positron emission tomography ([^{18}F]FET-PET) from magnetic resonance imaging for gliomas with lower-grade characteristics.
- Radiomics analysis with a Random Forest classifier achieved 80% accuracy in predicting high [^{18}F]FET uptake.

- Synthetic [18F]FET-PET images were clinically valuable in 85% of cases.
- Metrics showed high fidelity and structural congruence in synthetic images.
- This study paves the way for noninvasive imaging alternatives in regions lacking PET availability.

Graphical Abstract



Importance of the Study

This study addresses the limited availability of amino acid for positron emission tomography (PET) imaging, particularly O-(2-^{18}F-fluoroethyl)-L-tyrosine ([18F]FET)-PET, which is crucial for glioma diagnosis and treatment planning but is often unavailable in many regions. By developing a deep learning model capable of generating synthetic [18F]FET-PET images from preoperative magnetic resonance imaging, the research provides a significant step toward noninvasive and accessible imaging alternatives. This approach could enhance therapeutic decision-making, especially in areas where PET imaging is not available, improving overall management and outcomes of patients affected by glioma.

Gliomas are the most common malignant primary tumors in the adult population.[1,2] Lower-grade gliomas (LGG), a group of neuroepithelial neoplasms, arise from supporting glial cells of the central nervous system (CNS).[3] These tumors can be further classified by molecular features, such as isocitrate dehydrogenase (IDH) mutation.[4] Intratumoral heterogeneity is a challenge when diagnosing these tumors and failing to precisely biopsy regions of anaplastic malignancy can result in undergrading,[5,6] and, thus, suboptimal therapy strategies for the individual patient. Genetic variability within a tumor diminishes the efficacy of genomic analysis through invasive biopsies. However, it opens doors for noninvasive and repeatable medical imaging techniques that capture the complete tumor profile.

Current practice of preoperative LGG assessment involves magnetic resonance imaging (MRI) as the gold

standard. Magnetic resonance imaging offers good diagnostic performance for examining imaging features of LGG and their IDH mutations' status.[7] Advanced MRI research methods, such as 2-hydroxyglutarate magnetic resonance spectroscopy (MRS), can be used to detect the oncometabolite generated by the IDH mutation, that is, 2-hydroxyglutarate, with a sensitivity higher than the one related to the only use of MR imaging.[8–10] However, standard MRI protocols are limited in depicting small changes due to intratumoral heterogeneity, influencing decision-making in incidental findings.

Positron emission tomography (PET) with radiolabeled amino acids may unveil anaplastic foci and has, therefore, become standard practice in many parts of the world, not only in LGG but also in high-grade gliomas (HGG). Frequently used amino acids are O-(2-^{18}F-fluoroethyl)-L-tyrosine ([^{18}F]FET), [^{11}C]Methionine, and [^{18}F]Dopa, depending on clinical queries and regional availability. This noninvasive modality has provided clinicians with a powerful tool to identify tumor malignant subregions, monitor treatment response, and distinguish between tumor recurrence and pseudoprogression.[11–13] A significant advantage is that amino acids, such as [^{18}F]FET, are actively transported over the blood–brain barrier and are thus less susceptible to posttreatment-related effects.[14] Additionally, non-enhancing tumor regions with a higher risk of recurrence are difficult to unveil with MRI and have elevated uptake.[15] However, [^{18}F]FET-PET is not without limitations, primarily related to issues of accessibility and costs, but also low radiation exposure.

Moreover, [^{18}F]FET is still non-FDA approved and thus not available in the United States[16] or many high, middle- and low-income countries.[17] Nevertheless, this tracer has gained the endorsement of joint guidelines of the Response Assessment in Neuro-Oncology (RANO) working group and a joint European Association of Neuro-Oncology (EANO) and European Association of Nuclear Medicine (EANM) recommendations,[18,19] and it has become common practice in several centers worldwide, with physicians growing their clinical experience.[14,20]

Deep learning (DL) models have been paramount in advancing image-to-image translation[21] and image restoration.[22] Predominantly 2 networks, GAN[23] and the encoder-decoder network,[24] have expedited DL in these domains. Convolutional networks have proven useful in nuclear medicine for image denoising, attenuation correction, and image reconstruction.[25] Cross-modality image-to-image translation is a field where DL has proven effective. Synthetic computer tomography (CT) generation via MRI, either using deep convolutional neural networks,[26] or deep convolution adversarial networks[27] has been described. Several authors have effectively combined MRI data with ultra-low-dose radiotracer injections to synthesize diagnostic PET images for [^{18}F]Fluorodeoxyglucose ([^{18}F]FDG) studies in glioblastoma (GBM) patients and amyloid studies in individuals with cognitive issues.[28–31] More recently, authors have created synthetic PET imaging via MRI for [^{18}F]FDG-[25,32] and [^{11}C]Methionin-[33]PET in patients harboring brain neoplasms or amyloid for Alzheimer's diagnosis. This study is the first attempt to demonstrate the application of a DL model for synthetic [^{18}F]FET-PET generation trained on histologically confirmed gliomas lacking high-grade typical imaging characteristics. We thus analyze whether a 3D-U-Net architecture can succeed in generating PET images and recreating [^{18}F]FET uptake heatmaps out of features extracted from preoperative MRI.

Methods

Data Extraction and Patient Population

We aimed to review all patients with histologically diagnosed LGG (Grade 2 and 3) operated at the University Hospital Münster, Germany, between 2013 and 2023. Inclusion criteria for the synthetic generation task required the availability of both preoperative MRI and PET scans, not older than 3 months prior to surgery (Table 1). Patients were included if their preoperative MRI showed no evidence of necrosis and exhibited only patchy or focal contrast enhancement after gadolinium injection. For the hotspot prediction task, only preoperative MRI and binary classification of hotspot presence based on PET scan reports were required, allowing us to include all patients with MRI images and PET reports. In contrast, the synthetic PET generation task required both preoperative MRI and actual PET images, resulting in a slightly smaller cohort due to missing PET images for some patients (Table 1). Ethical board approval for this study was granted by the ethics committee of the University of Münster (2023-545-f-S). Due to the study's retrospective nature, obtaining informed consent from subjects was not sought.

We initially targeted individuals diagnosed with LGG based on clinical and radiological criteria. However, during our investigation, we encountered diverse tumor pathology, including glioblastomas classified as "molecular" GBMs and gliomas of various grades lacking typical high-grade features on MRI. The presence of molecular GBMs and other gliomas with atypical imaging features for their grade within our cohort highlights the challenges and limitations of relying solely on traditional imaging modalities for glioma classification and the potential utility of synthetic [^{18}F]FET-PET imaging in providing a more nuanced understanding of tumor biology.

Patient Demographics

We analyzed data from a cohort of 215 studies from 211 individuals. The gender distribution showed a predominance of males, with 119 men (55%) and 96 women (45%). Age distribution among the participants varied, with 89 individuals younger than 40 years (41%), the majority being between 40 and 69 years old (55%), and a small percentage aged 70 years or older (4%) (Table 1).

Astrocytoma accounted for 68% of the cases, with 99 identified at initial diagnosis and 48 at recurrence. Oligodendroglioma was detected in 28% of the patients, with 29 at initial diagnosis and 30 at recurrence. Glioblastoma and radiation necrosis were rare and present in a minority of the cases: 2% (4 primary diagnoses, 1 recurrence) and less than 1%, respectively. The demographics of the synthetic PET cohort, which are only slightly different, are shown in Table 1.

Table 1. Patient Demographics of the PET Synthetic Generation Task. MRI and PET Reports Were Used for the Hotspot Prediction Task, Whereas for Synthetic PET Generation, MRI and PET Images were Required, Resulting in a Slightly Smaller Cohort

	Synthetic [18F]FET-PET generation		[18F]FET-PET hotspot prediction	
Case count (number of patients included)	211 (204)		215 (187)	
Sex				
Male	116 (110)	55%	119 (99)	55%
Female	95 (94)	45%	96 (88)	45%
Age				
<40	92	44%	89	41%
40–69	112	53%	119	55%
≥70	7	3%	7	4%
Initial diagnosis	128	60%	130	60%
Recurrence	83	40%	85	40%
WHO grade				
1	1	<1%	1	1%
2	185	88%	189	88%
3	16	7%	16	7%
4	5	2%	5	2%
N/A	4	2%	4	2%
Tumor type				
Astrocytoma	143	68%	147	68%
Initial diagnosis	94	66%	99	67%
Recurrence	49	34%	48	33%
Oligodendroglioma	58	28%	59	28%
Initial diagnosis	30	52%	29	49%
Recurrence	28	48%	30	51%
Glioblastoma	5	2%	4	2%
Initial diagnosis	2	40%	1	25%
Recurrence	3	60%	3	75%
Necrosis	2	1%	2	<1%
Varia	3	1%	3	1%

Abbreviations: [18F]FET = O-(2-18F-fluoroethyl)-L-tyrosine; MRI = magnetic resonance imaging; PET = positron emission tomography; WHO = World Health Organization.

Acquisition of MRI Imaging

Magnetic resonance imaging images were extracted from the electronic medical records of patients treated at the University Hospital Münster, Germany. Preoperative MRI was often performed before admission and thus very often in different MRI machines. Twenty-five different MRIs were used to create the images analyzed in this study (Supplementary Table 1), contributing to the model's generalizability.

Acquisition of [18F]FET-PET Imaging

A 3T PET-MRI system (mMR; Siemens Healthcare GmbH) ($n = 50$) or a PET-CT (mCT, Siemens Healthcare GmbH, Erlangen, Germany) was used for [18F]FET-PET imaging. Following guidelines and institutional protocols, the injected activity was 2.5 MBq [18F]FET/kg body weight. [18F]FET was produced onsite in an in-house cyclotron unit. Positron emission tomography was acquired with a full dynamic acquisition mode starting with tracer injection for 40 min or with a static acquisition from 20 to 40 min postinjection. Twenty to forty minutes of static images were used for further analysis. Positron emission tomography images were reconstructed using the following parameters: PET-CT: time of flight, 400 × 400 matrix, 1.02 × 1.02 × 2.03 mm voxel size, 5 iterations, 21 subsets, 4 mm full-width-at-half-maximum Gaussian filter; PET-MRI: 344 × 344 matrix, 1.04 × 1.04 × 2.03 mm voxel size, 3 iterations, 21 subsets, 5 mm full-width-at-half-maximum Gaussian filter. Only attenuation-corrected sequences were used. A "hotspot" was defined as a region of significantly elevated focal tracer uptake above the background. The information on the presence of a hotspot was extracted from [18F]FET-PET reports and verified by an experienced and board-certified nuclear medicine physician.

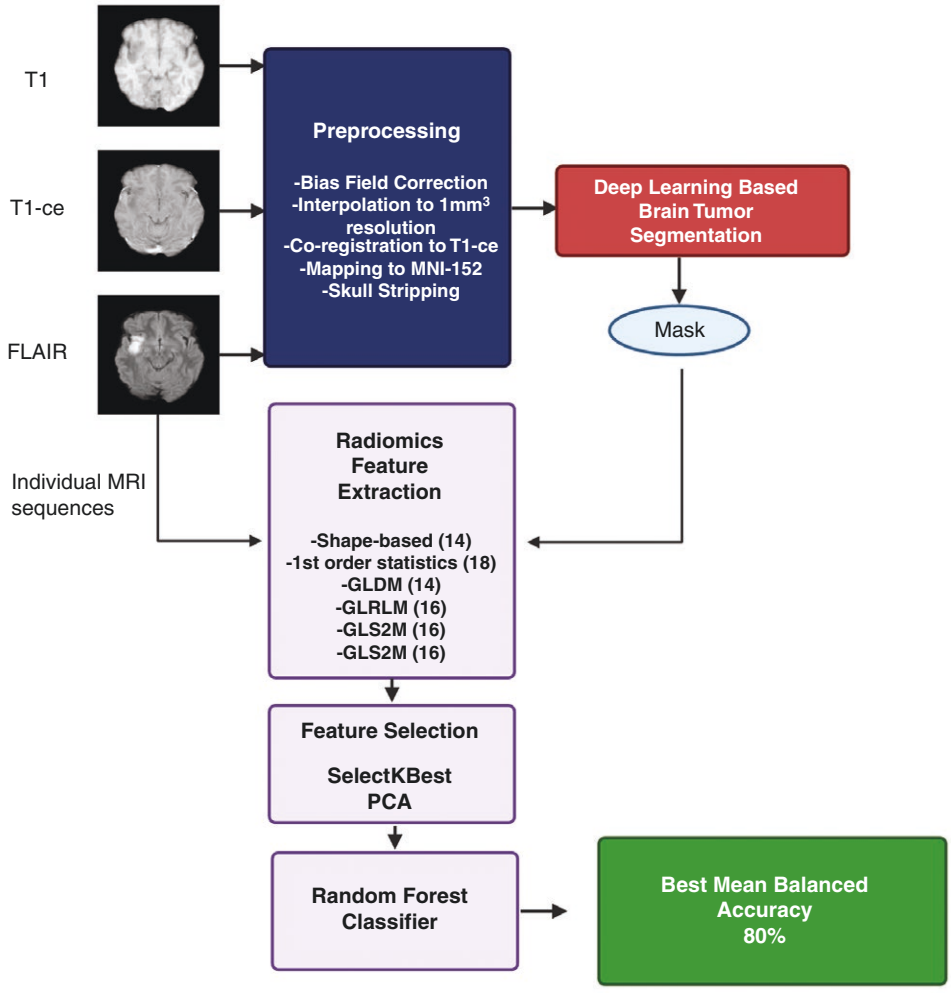


Figure 1. Predictive model for hotspot classification.

Preprocessing

A comprehensive preprocessing step was performed before introducing preoperative MRI sequences into our generative framework. Preprocessing was conducted on the images to remove artifacts and enhance signal quality without compromising information loss. The acquired MRI sequences and the [18F]FET-PET scans underwent the same preprocessing pipeline to ensure consistency and reliability in subsequent analyses. The preprocessing steps included bias field correction, interpolation to 1 mm^3 resolution, co-registration to contrast-enhanced weighted-T1 (T1-ce), mapping to the Montreal Neurological Institute template (MNI-152), and skull stripping. The pipeline was implemented using the advanced normalization tools (ANTs) software package[34] and the FMRIB Software Library (FSL), developed at the Oxford Center for Functional Magnetic Resonance Imaging of the Brain (FMRIB).[35] Advanced normalization tools provided robust capabilities for bias field correction, image resampling, co-registration, and skull stripping, while FSL was used to map the images to the MNI-152 template.

Hotspot Prediction Task

The primary objective of this task was to develop a predictive model capable of accurately identifying hotspots based on features extracted from preoperative MRI (Figure 1).

Tumor segmentation.—This study used the Variational Auto Encoder U-Net model introduced by Di Ieva et al.,[36] pretrained for segmenting tumors using MRI sequences. This model builds upon the Myronenko model,[37] recognized as the top-performing model in the Brain Tumor Segmentation (BraTS) 2018 challenge. The model used in this study presents a notable advantage in its versatility and provides various pretrained versions tailored for individual and MRI sequences (encompassing all possible combinations), a feature not present in the winning submissions to recent BraTS challenges. Importantly, this advantage made the model particularly suitable for this study, as we utilized only T1-weighted images, T1-ce, and fluid-attenuated inversion recovery (FLAIR) sequences.

Feature extraction and selection.—We employed the mask obtained from the tumor segmentation model described earlier to identify the tumor region, which was then applied to each MRI sequence. Subsequently, radiomics features were extracted exclusively within the identified tumor region. A total of 83 radiomics features were extracted from each MRI sequence using Python PyRadiomics package.[38] The extracted features include:
Shape-based features: these 14 features capture geometric properties such as volume, surface area, and compactness, providing insights into the overall tumor's morphology. First-order statistics: with 18 features, this category quantifies the distribution of voxel intensities within the tumor region, providing basic statistical descriptors such as mean, median, variance, skewness, and kurtosis. Gray level dependence matrix (GLDM): encompassing 14 features, GLDM features characterize the spatial dependencies between voxel intensities, offering insights into the texture and homogeneity of the tumor. Gray level run length matrix (GLRLM): comprising 16 features, GLRLM features quantify the length and distribution of consecutive voxel runs at different gray levels, providing information about the texture and coarseness of the tumor. Gray level size zone matrix (GLSZM): with 16 features, GLSZM features capture the size and spatial distribution of connected voxel clusters at different gray levels, reflecting the heterogeneity and complexity of the tumor. Neighboring gray tone difference matrix (NGTDM): this category includes 5 features that quantify the difference in voxel intensities between neighboring voxels, offering insights into the local variations and patterns within the tumor region.

Given the varying ranges of the extracted features, the MinMax scaling technique was used to normalize the features by linearly scaling them down into a fixed range. Also, considering the high dimensionality of the feature vectors compared to the number of samples, SelectKBest (SKB) and principal component analysis (PCA) techniques were employed to reduce potential overfitting.

Classification and performance evaluation.—A Random Forest classifier was employed to classify the radiomics features extracted from each MRI sequence. Specifically, 3 models were trained and tested, each corresponding to one of the MRI sequences utilized in the study. A rigorous 10-fold cross-validation procedure enhanced our classification approach's robustness. This procedure ensured that the performance metrics were reliably estimated across different subsets of the dataset. During each iteration of the cross-validation process, the optimal number of components for SKB and Principal PCA was determined to maximize the overall accuracy of the models.

Given the imbalance in the training set for each fold, a random undersampling technique was used to address this issue. This technique involved randomly removing instances from the majority class to balance the class distribution, thus preventing bias toward the majority class during training.

Performance evaluation metrics were recorded for each fold of the cross-validation process. The mean balanced accuracy, mean sensitivity, and mean specificity were considered the primary classification efficacy measures. These metrics provided insights into the overall performance of the classifiers in accurately distinguishing between different classes while accounting for potential class imbalances in the dataset.

Synthetic [^{18}F]FET-PET Generation Task

Generative framework.—Our study employs a generative adversarial network (GAN) framework, adapted from the models in Isola et al.,[39] for generating [^{18}F]FET-PET images from preoperative MRI. The code was written using Keras with a TensorFlow end, demonstrating the practical application and customization of the pix2pix model for various image translation tasks. This GAN framework is configured explicitly for conditional image generation, mapping an observed MRI image x and a random noise vector z to an output PET image y, formulated as G: → y.

The dataset was divided into a training set of 162 patients and an unseen test set of 49 patients to evaluate the model performance rigorously. All performance metrics reported in this study were computed on this test set, ensuring that the results reflect the model's generalization capability to unseen data and avoid any training data bias.

The inherent complexity of handling 3D Neuroimaging Informatics Technology Initiative (NifTi) MRI volume files and the necessity of pairing them with PET modalities present a significant challenge. The MRI and PET images differ in structural properties, resolution, and anatomical alignment. This disparity necessitates sophisticated preprocessing techniques to align and standardize these modalities effectively. We adopted a split-input fusion module to process different MRI modalities—T1-weighted imaging (T1), T1-ce, and FLAIR through feature extraction, concatenation, and self-attention as shown in Figure 2A.

Fusion of MRI modalities.—Each MRI sequence (T1, T1-ce, and FLAIR) is initially processed to highlight distinctive features within the images in the input fusion model for MRI data analysis. This is achieved through a convolutional process that filters the image data, effectively isolating essential structural elements within each MRI sequence. These processed sequences are then merged into a comprehensive image profile, which serves as a more detailed representation of the brain's anatomy. A self-attention mechanism is applied to refine this composite image further, which selectively enhances critical features for accurate PET image synthesis. This attention-enhanced image data is expected to improve the subsequent generation of PET images, ensuring they retain the essential details necessary for practical clinical evaluation.

Training procedure for generating [^{18}F]FET-PET images.—Pix2Pix GAN belongs to a category of conditional GAN[40] and is trained using joint optimization of 2 networks: the Generator (G) and Discriminator (D). The G network generates synthetic or fake data, while the D network determines whether the generated samples are authentic or fake. Both networks undergo alternating training to maximize the probability of correctly classifying authentic images and generated samples. The GAN achieves optimal training

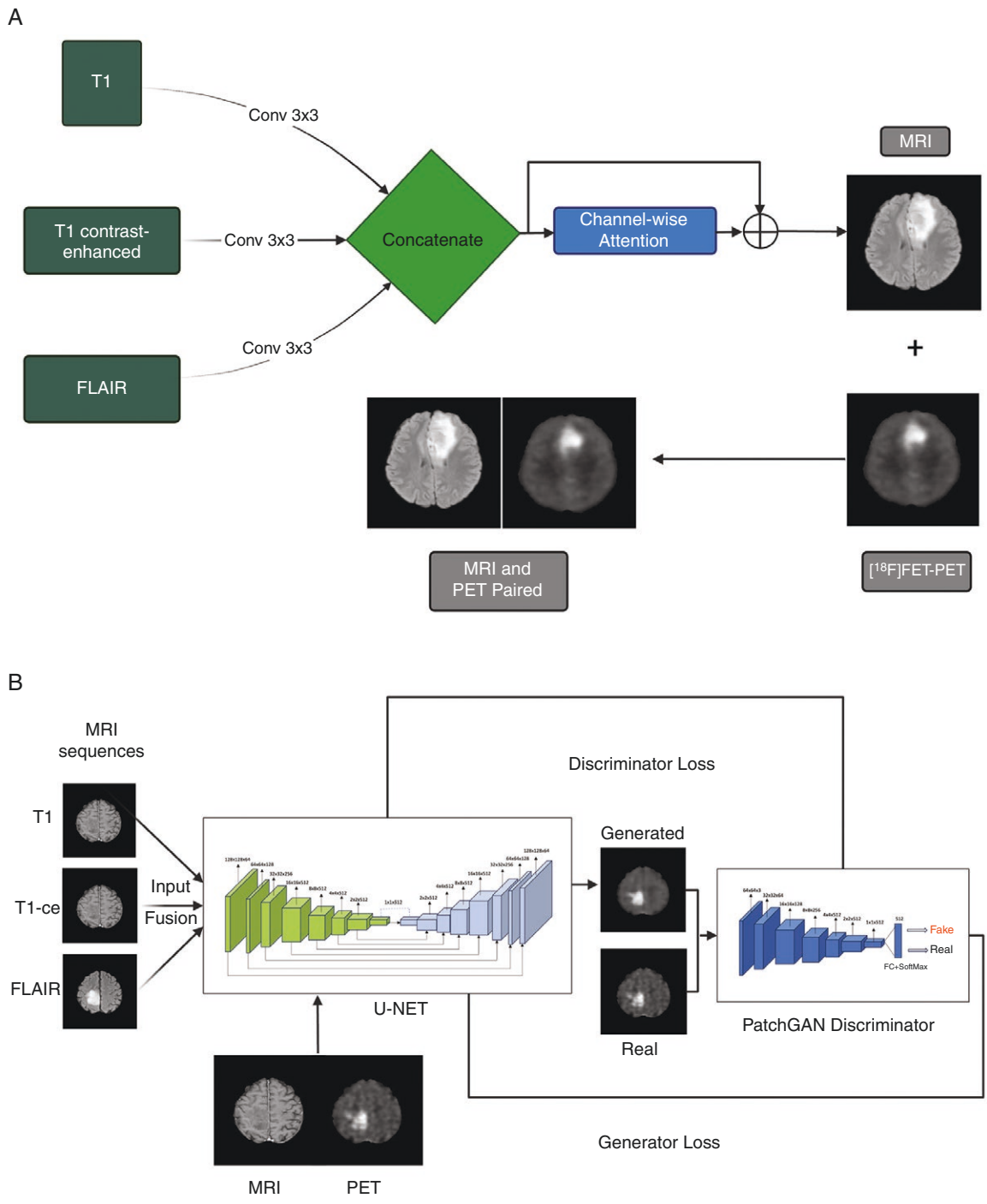


Figure 2. (A) MRI fusion module. (B) The architecture of the GAN Model. Abbreviations: MRI = magnetic resonance imaging; GAN = generative adversarial network.

when the generator can produce data samples as diverse as the original data distribution and deceive the discriminator into perceiving them as real.

Generator architecture.—The architecture adopted is a variant of the U-Net model, characterized by its encoder-decoder structure with intervening skip connections. The encoder compresses the input image into a dense, high-dimensional feature space. It comprises a series of convolutional layers that incrementally increase in channel capacity, as indicated by the sequence C64-C128-C256-C512-C512-C512-C512-C512-C512, where "C" denotes a convolutional layer, and the numeral represents the count of filters within that layer. This configuration allows for the

extraction of progressively complex features at multiple scales.

The decoder performs the inverse operation. It methodically upscales the abstract feature representation, reconceptualizing the spatial dimensions of the image through a symmetrical configuration of layers, denoted as CD512-CD512-CD512-C512-C256-C128-C64. "CD" here likely represents a deconvolutional or transposed convolutional layer that serves to upsample its input. The decreasing sequence of channel numbers reflects the layer-wise detail refinement process undertaken during image reconstruction.

Discriminator architecture.—The 70 × 70 PatchGAN discriminator follows a C64-C128-C256-C512 sequence, ending with a convolution to a 1-dimensional output and a Sigmoid function. Like the generator, the first C64 layer does not include BatchNorm, and all rectified linear units (ReLUs) are leaky. Variations of the discriminator, such as 1 × 1, 16 × 16, and 286 × 286 architectures, adjust the receptive field size, with the smallest discriminator using 1 × 1 spatial filters. This Markovian discriminator focuses on local image patches.

Figure 2B illustrates the workflow where MRI sequences (T1, T1-ce, and FLAIR) are fed into a U-Net generator. The generator processes these inputs to produce synthetic PET images, which, along with authentic PET images, are evaluated by the PatchGAN discriminator for authenticity at a patch level. The discriminator assesses if the generated PET patches are real or fake, contributing to the discriminator's loss. Simultaneously, the generator is trained to minimize the generator loss by producing images that closely resemble authentic PET images. This adversarial process iterates to enhance the fidelity of the generated PET images to actual PET scans, aiming to achieve precise and high-quality image translation for medical diagnostic purposes.

Loss functions.—Our conditional GAN model aims to fine-tune the generator and discriminator in a competitive setup. The generator strives to create images close to the target, while the discriminator seeks to differentiate between generated and authentic images. To achieve high-quality synthetic images, we utilize an L1 loss that calculates the average absolute difference between the synthetic and target images, ensuring pixel-level accuracy. The final objective harmonizes the GAN's adversarial goal with the precision mandated by the L1 loss, encapsulated in a single equation that represents the balance between the 2:

$$G^* = argmin_G \ argmax_D \left(L_{cGAN}(G, D) + \lambda L_{L1}(G)\right)$$

Here, λ is a weighting factor that balances the adversarial nature of the GAN with the pixel-wise accuracy enforced by the L1 loss.

Optimization strategy.—The generator G is trained to minimize this combined objective, while the discriminator D aims to maximize its part of the loss. We utilize minibatch Stochastic Gradient Descent and the Adam solver for optimization, with a learning rate of 0.0002 and momentum terms β1 = 0.5 and β2 = 0.999. The training process is conducted over 200 epochs with a batch size of 1.

This comprehensive approach, integrating a U-Net generator with skip connections and a Markovian PatchGAN discriminator, is optimized for high fidelity [18F]FET-PET image reproduction tasks. Including the L1 loss ensures that the generated images are realistic and maintain high structural fidelity to the target images, which is critical in medical imaging applications.

Evaluation Metrics of the Synthetic Generated [18F]FET-PET Images

Quantitative analysis.—The peak signal-to-noise ratio (PSNR) is a standard metric for image quality assessment, which compares the maximum potential power of a signal to the power of corrupting noise. The PSNR of our synthetic images was calculated using the equation:

$$PSNR = 10.log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

where MAXI represents the maximum possible pixel value of the image, and MSE denotes the mean squared error between the original and synthetic images. The structural similarity index (SSIM) measures the perceived quality of digital images and videos. Our computed SSIM values, defined by:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Where x and y are the original and synthetic images, respectively, indicate a substantial structural congruence between the 2. Root mean square error, a measure for the differences between predicted and observed values, was employed to compare pixel intensity errors directly.

$$RMSE(x, y) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}$$

Continuing with the evaluation of image quality metrics, the visual information fidelity (VIF) metric was employed to assess how closely the synthetic images approximate the information fidelity relative to the original, as perceived by the human visual system. Visual information fidelity was computed using the relation: $VIF(x, y) = \sum log\left(1 + \frac{\sigma_{xy}^2}{\sigma_x^2(\sigma_y^2+\sigma_n^2)}\right)$.

Where x denotes the original image, y the synthetic counterpart, σ_{xy}^2 the covariance of x and y, σ_x^2 and σ_y^2 the variances of x and y, respectively, and σ_n^2 represents the variance of the noise perceived by the human visual system.

Qualitative assessment.—For a comprehensive evaluation, *a group of senior physicians*, including specialists in nuclear medicine ($n = 1$), radiology ($n = 1$), and neurosurgery ($n = 2$), were presented with MRI sequences alongside real and synthetic PET images from a test cohort of 49 patients. Their assessment aimed to verify clinical accuracy

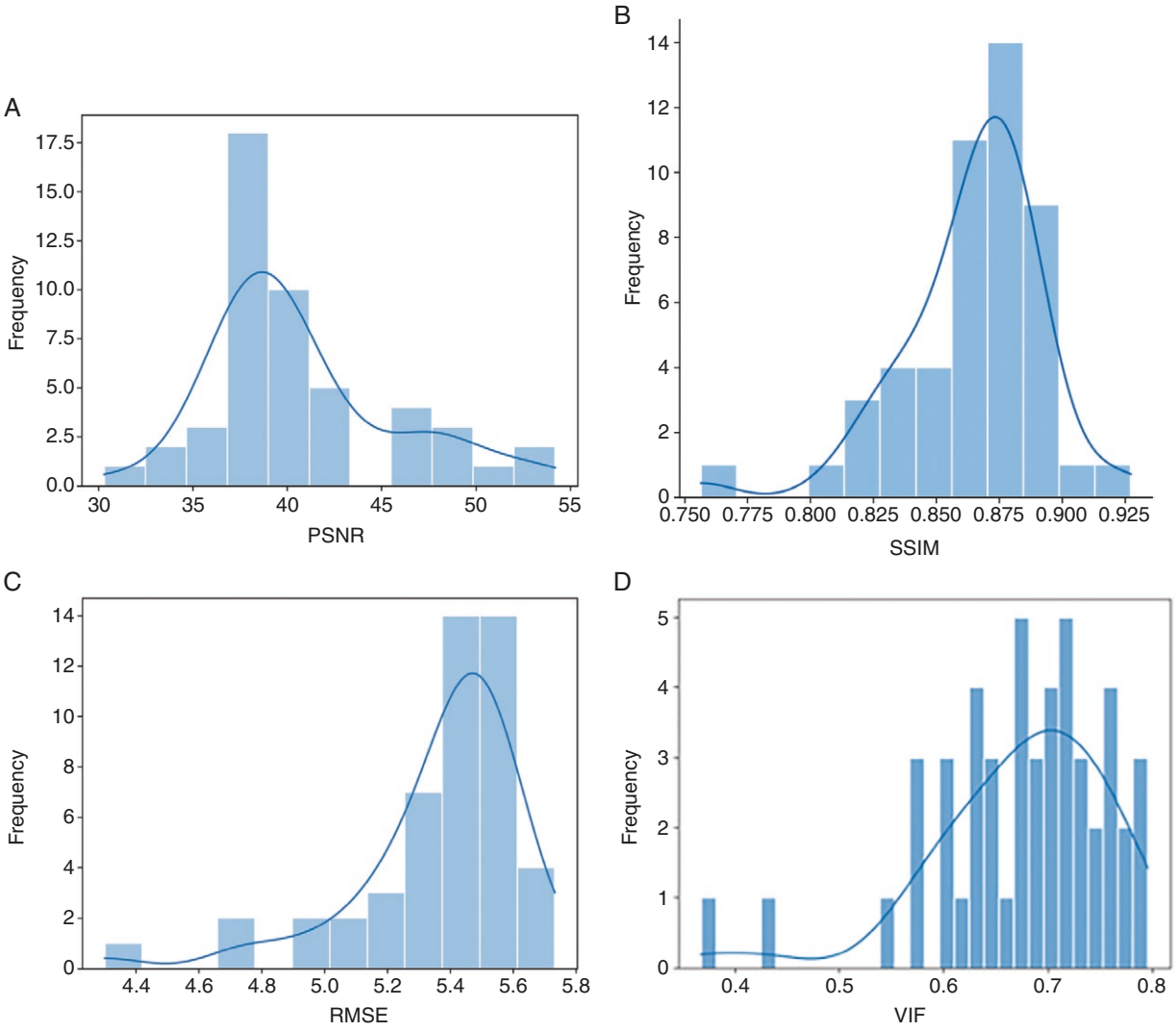


Figure 3. Comparative distribution of image quality metrics. This histogram represents the distribution of (A) peak signal-to-noise ratio (PSNR), (B) structural similarity index (SSIM), (C) root mean square error (RMSE), and (D) visual information fidelity (VIF) across a dataset of images. The overlaid curves indicate the kernel density estimation for each metric, providing insight into the central tendency and dispersion, which reflects image synthesis performance.

and alignment with radiological standards, focusing on the hotspot regions.

In the hotspot detection task, we ensured accuracy by employing a 2-step verification process: scans were initially reviewed by an expert nuclear medicine specialist and a radiologist, followed by validation from an experienced neurosurgeon. This multi-expert approach ensured that any identified hotspots in the scans were accurate and clinically relevant. For automatic segmentation in hotspot classification, we employed a state-of-the-art method to achieve consistent segmentation across all scans.

Results

After the initial search, 550 surgically treated patients with gliomas lacking high-grade radiographic characteristics were identified. However, according to the inclusion criteria, only 211 studies from 207 patients could be enrolled for the synthetic PET generation task due to the PET images' unavailability or missing MRI sequences for the model. The model for the binary classification task could be trained with a different cohort, that is, 215 studies from 187 patients (Table 1).

Synthetic [18F]FET-PET Generation Task

Similarity metrics.—In our analysis of the synthetic brain PET images generated by the proposed model, we employed 4 principal quality assessment metrics: PSNR, SSIM, and root mean square error (RMSE) (Figure 3).

Peak signal-to-noise ratio.—The histogram for PSNR displays a prominent peak around the value of 40 dB,

suggesting that synthetic images have high signal fidelity relative to the noise level.

Structural Similarity Index.—The histogram of SSIM values exhibits a skew toward the higher end, particularly in the range of 0.85–0.90, emphasizing the structural and textural similarity of the generated images to the real PET images.

Root mean square error.—The histogram for RMSE illustrates a concentration of lower values, with most errors clustering below 5.6, indicating that the model accurately predicted pixel intensity.

Regarding the qualitative assessment, the senior physicians unanimously agreed that the real and synthetic PETs were visually accurate by 90%. Only 5 of 49 cases (~10%) were noted where the model could not reproduce small regions of high [^{18}F]FET uptake, and in 3 cases (~5%), the FET positive area was underestimated. In one case, the model provided an accurate hotspot but underestimated the [^{18}F]FET-PET positive volume in visual analysis (Figure 4).

Visual information fidelity.—The histogram of VIF scores displays a distribution that suggests a concentration of data points in the middle range, with most values clustered around 0.6–0.7. This central aggregation indicates that most synthesized images maintain a moderate level of visual fidelity compared to the original images. The distribution's tails taper off toward both the lower and higher ends of the VIF scale, implying fewer images with very low or very high fidelity.

Hotspot accuracy.—The radiomics features extracted from 3 different MRI sequences—T1, T1-ce, and FLAIR—were analyzed using various feature selection methods (Supplementary Table 2). The feature selection for T1 was carried out using SKB, which resulted in an overall confusion matrix where 37 cases were correctly identified as positive, 10 as false positive, 48 as false negative, and 122 as true negative. This yielded a mean sensitivity of 72%, a mean specificity of 79%, and a mean balanced accuracy of 75% (Supplementary Table 2).

For the T1-ce sequence, the SKB method was also utilized for feature selection. The overall confusion matrix showed an improvement with 38 true positives, 9 false positives, 36 false negatives, and 134 true negatives. The mean sensitivity increased to 79%, the mean specificity to 81%, and the mean balanced accuracy was 80%, indicating a better performance than the T1 sequence.

Lastly, we used PCA in the FLAIR sequence for feature selection. The metrics for this sequence showed a mean sensitivity of 78%, a mean specificity of 79%, and a mean balanced accuracy of 78%.

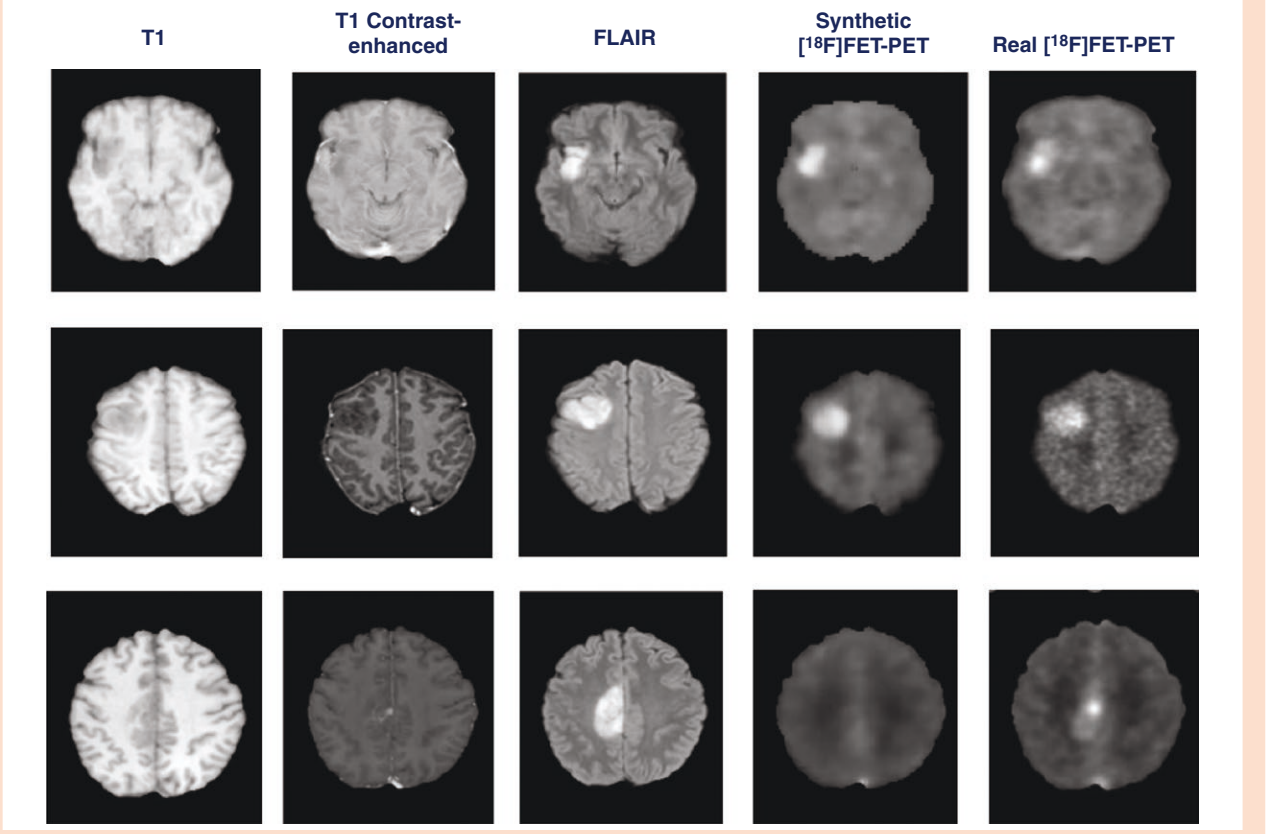


Figure 4. Representative cases of MRI-based synthetic [^{18}F]FET-PET: 2 accurate reconstructions (first 2 rows) and an example of underestimation of FET uptake (last row). Abbreviations: [^{18}F]FET = O-(2-^{18}F-fluoroethyl)-L-tyrosine; MRI = magnetic resonance imaging; PET = positron emission tomography.

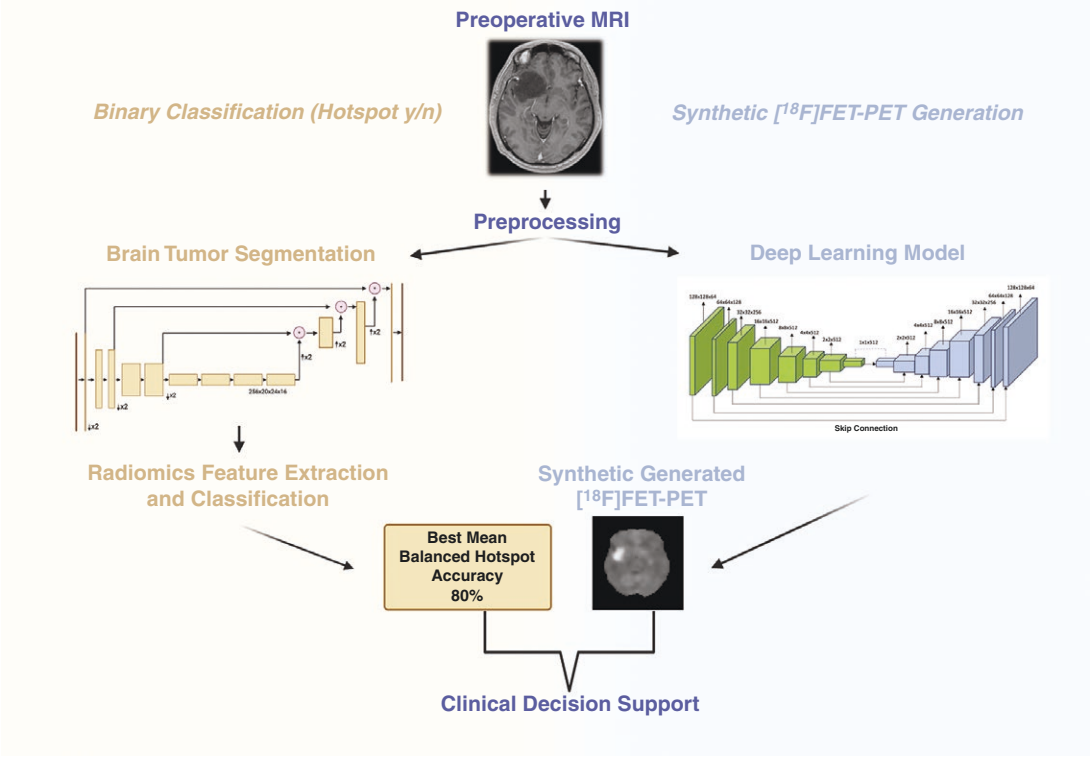


Figure 5. Workflow for MRI-based brain tumor analysis with deep learning enhanced clinical decision support. Abbreviation: MRI = magnetic resonance imaging.

Discussion

We present 2 different pipelines that use artificial intelligence (AI) modalities to create synthetic images of [18F]FET-PET and, after tumor segmentation, binary classify the preoperative MRI into the probability of having a significantly higher uptake of [18F]FET. Both models complement each other in terms of clinical value and decision-making. While both models were trained separately, they can be incorporated into a single pipeline to evaluate gliomas with lower-grade radiographic characteristics (Figure 5).

Clinical Value of MRI and [18F]FET-PET

Radiologically, IDH-mutant tumors present with different and often more subtle features than IDH wild-type gliomas. Specifically, these tumors favor the frontal lobe. They have well-defined tumor margins, no or fewer incidences of contrast enhancement, higher apparent diffusion coefficient, and lower relative cerebral blood volume than their IDH wild-type counterparts.[9] Thus, the classical definition of low-grade gliomas changed as "typical" MRI features do not necessarily indicate a low-grade tumor for the reasons mentioned above, and authors now prefer the term lower-grade glioma.[3]

[18F]FET is an artificial amino acid taken up by the L-Type amino acid transporter (LAT), which is overexpressed in gliomas,[41] resulting in a higher specificity compared to other tracers like [18F]FDG (correlated to high physiological brain uptake).[42] In LGG, [18F]FET-PET is frequently applied in initial work-up to define hotspots, revealing more aggressive tumor biology and eventually guiding biopsy and resection. The [18F]FET-PET positive volume, often referred to as biological tumor volume, is relevant for neurosurgical resection and radiation therapy planning.[43]

Differentiation between all tumor subtypes is challenging. Oligodendroglioma, for example, is part of the LGG group and presents with high [18F]FET uptake comparable to HGG. However, the uptake of oligodendroglioma is significantly higher compared to low-grade astrocytoma. Novel data is needed as most of the published studies do not implement the new World Health Organization CNS 2021 classification.[44]

Advances in Generative AI for Neuroimaging

Generative adversarial networks[45] have revolutionized medical imaging, bringing innovative solutions to image analysis, enhancement, and synthesis challenges. These advanced networks, through their unique architecture comprising a generator and a discriminator, excel in creating realistic medical images, aiding significantly in areas such as data augmentation, anomaly detection, and improved diagnostics. Generative adversarial networks have been particularly transformative in scenarios where

medical imaging data is scarce or unbalanced, helping to train more robust diagnostic models and contributing to enhanced patient care.

The introduction of Conditional GANs[46] marked a revolutionary shift in medical imaging, particularly in translating images across different modalities. This technique allows for the controlled generation of images, where the output is conditioned on certain input features, providing a more targeted approach to image synthesis. The pix2pix GAN[39] is a prominent example in this landscape, designed explicitly for image-to-image translation tasks. The pix2pix model utilizes paired training data to learn how to convert an image from one domain to another, making it highly effective for tasks like translating MRI scans to CT images or vice versa.

While the pix2pix GAN has seen extensive use in various medical imaging areas, its application in PET remains comparatively rare. Most research has centered around translating images between MRI and CT scans, with some exploring the conversion between standard radiological imaging (CT or MRI) and PET images.[47] Recently, a study by Choi et al.[48] explored generating Fluorodeoxyglucose (FDG)-PET images from Florbetaben (FBB)-PET images. To our knowledge, our study is the first to assess the feasibility of using DL to create images akin to [18F]-FET-PET with higher tracer uptake from conventional MRI images. Our project is at the cutting edge, aiming to use pix2pix GAN for synthesizing [18F]FET-PET images from various MRI inputs, potentially advancing neuroimaging understanding and contributing significantly to diagnosing and treating neurological disorders.

Computational Challenges in Synthetic Neuroimaging

Employing the pix2pix model in our task necessitates considerable computational effort. This arises primarily from the generator's need for an advanced and complex network architecture, crucial for accurately delineating the differences between MRI and PET imaging modalities. Handling this data-heavy task demands significant computational resources. Again, the discriminator plays a vital role in augmenting the computational requirements, as it rigorously ensures the realism and medical veracity of the synthetic images. This function requires extensive processing power for in-depth analysis. Also, the characteristic iterative training regimen of GANs, as seen in the pix2pix model, intensifies the computational demands. Each training cycle involves processing large volumes of data and performing sophisticated calculations, which must be repeated multiple times to train the model effectively. Despite converting 3D MRI data into 2D format following initial input fusion, this methodology remains resource-intensive due to the high level of fidelity required in medical imaging and the complexities associated with training DL models in this specialized domain.

Innovation in Synthetic [18F]FET-PET Generation

In our project, we adapted the architecture of the pix2pix GAN to suit our specific dataset better. The generator employs an encoder-decoder structure, with the encoder sequence configured as C64-C128-C256-C512-C512-C512-C512-C512. This arrangement progressively deepens to encode the input MRI data efficiently. The decoder follows a CD512-CD512-CD512-C512-C256-C128-C64 pattern to reconstruct the encoded data into detailed PET images. We have opted for a modified structure comprising layers C64-C128-C256-C512 for the discriminator. Following the last layer, a convolution maps the output to a 1-dimensional scale, concluding with a Sigmoid activation function. This specific inclusion of the Sigmoid function is strategic, enabling the model to capture continuous values crucial for distinguishing between pathological hotspot pixels and normal tissue in the PET images, an essential requirement given the nature of our dataset.

This combination of quantitative metrics, qualitative assessment, and a rigorous reference standard for hotspot identification underscores our commitment to producing high-quality and clinically valuable synthetic images.

DL Framework

The collective assessment of PSNR, SSIM, RMSE, and VIF metrics confirms the model's adeptness at synthesizing PET images with remarkable fidelity. The PSNR values point to excellent signal reconstruction capabilities, while the SSIM histogram indicates significant preservation of structural and textural congruence, especially noted in the upper range nearing 0.90. The RMSE values suggest precise pixel intensity accuracy, and the VIF scores, predominantly situated in the 0.6–0.7 range, further affirm the visual consistency of the synthetic images. This peak in the VIF distribution underscores the synthesized images' capacity to retain substantial informational content as perceived by the human visual system. Cumulatively, these metrics corroborate the robustness of the generative model in producing high-quality synthetic images, promising to enrich diagnostic procedures and clinical evaluations.

Our intention is not to replace the PET modality but rather to pave the way for a potential future alternative to physicians without access to nuclear medicine facilities, assisting in creating a better treatment strategy for lower-grade and eventually HGG.

We have demonstrated the feasibility of predicting [18F]FET-PET images from MRI using the T1-weighted, T1-post gadolinium weighted sequences and FLAIR. Our results indicate a reasonable accuracy with further potential for improvement. As part of our future work, we evaluate more advanced imaging modalities.

Limitations

The retrospective nature of this study is one of the limitations of our research. However, our model is reproducible due to the large patient cohort and heterogeneous MRI dataset. A large portion (85%) of the tumors demonstrated a higher [18F]FET uptake and were classified as having a hotspot. This is an expected imbalance, as the cohort was formed from surgically treated patients. This imbalance could lead to the model overestimating the incidence of FET-PET uptake.

Nevertheless, predominantly in patients without a higher uptake of [18F]FET, our model generated no false positive

synthetic PET image. Currently, the synthetic PET images do not allow for uptake quantification, including clinically used parameters such as tumor-to-background ratios and biological tumor volume. Therefore, we plan to synthesize absolute standard uptake value per pixel to be used in future studies, allowing for a reproducible and accurate hotspot assessment and tumor segmentation.

Conclusions

For the first time, we demonstrate that synthetic [18F]FET-PET generation is feasible. A DL-based framework created synthetic [18F]FET-PET images with reasonable accuracy from standard MRI. Even though our model can be further improved and should be prospectively validated in larger cohorts, it might already help in its current form in the assessment and treatment planning for patients harboring LGG.

Supplementary material

Supplementary material is available online at *Neuro-Oncology Advances* (https://academic.oup.com/noa).

Keywords

[18F]FET-PET | deep learning | GAN | glioma | radiomics

LAY SUMMARY

Gliomas are brain tumors that can grow quickly and can be difficult to treat. Positron emission tomography (PET) scans are special images that help doctors understand how active different parts of a tumor are, but many patients have trouble accessing them. In this study, the authors wanted to find out if computer programs could create images similar to PET scans using regular magnetic resonance imaging (MRI) scans, which are more widely available. To do this, they developed and tested computer models that used MRI scans to identify active tumor areas (hotspots) similar to what PET scans show. Their results showed that the computer-generated images provided useful information about the tumors and were similar to real PET scans.

Funding

Open *Access funding enabled and organized by Projekt DEAL.*

Acknowledgments

We thank Christian Thomas from the Institute of Neuropathology at the University of Münster for his help with data extraction. Figures 2 and 4 were created with Biorender.com.

Conflict of interest statement. None declared.

Authorship statement

Concept and design of this study were developed by E.S.M. and A.D.I. E.S.M., Z.Ö., and A.V.C. were responsible for data collection. E.S.M., M.T., S.L., C.R., and G.A. worked on the model development and interpretation of results. The manuscript was prepared by E.S.M., M.T., and G.A. and critically reviewed and edited by all authors.

Data availability

Patient data cannot be publicly shared due to data protection reasons, but the trained model can be provided to individuals upon request.

Affiliations

Department of Neurosurgery, University Hospital Münster, Münster, Germany (E.S.M., Z.O., A.V.C., W.S.); Computational NeuroSurgery (CNS) Lab, Macquarie Medical School, Macquarie University, Sydney, NSW, Australia (E.S.M., M.T., G.A., C.R., S.L., A.D.I.); Macquarie Neurosurgery & Spine, Macquarie University Hospital, Sydney, NSW, Australia (E.S.M., A.D.I.); Centre for Health Informatics, Australian Institute of Health Innovation, Macquarie University, Sydney, NSW, Australia (M.T., S.L.); Department of Nuclear Medicine, University Hospital Münster, Münster, Germany (W.R., P.B.); Clinic for Radiology, University Hospital Münster, Münster, Germany (P.S.); Department of Neurosurgery, Nepean Blue Mountains Local Health District, Kingswood, NSW, Australia (A.D.I.)

References

1. Alshiekh Nasany R, de la Fuente MI. Therapies for IDH-mutant gliomas. *Curr Neurol Neurosci Rep.* 2023;23(5):225–233.
2. Lapointe S, Perry A, Butowski NA. Primary brain tumours in adults. *Lancet.* 2018;392(10145):432–446.
3. Youssef G, Miller JJ. Lower grade gliomas. *Curr Neurol Neurosci Rep.* 2020;20(7):21.
4. Parsons DW, Jones S, Zhang X, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science.* 2008;321(5897):1807–1812.
5. Widhalm G, Kiesel B, Woehrer A, et al. 5-Aminolevulinic acid induced fluorescence is a powerful intraoperative marker for precise histopathological grading of gliomas with non-significant contrast-enhancement. *PLoS One.* 2013;8(10):e76988.
6. Jaber M, Wolfer J, Ewelt C, et al. The value of 5-Aminolevulinic acid in low-grade gliomas and high-grade gliomas lacking glioblastoma

imaging features: an analysis based on fluorescence, magnetic resonance imaging, 18F-Fluoroethyl tyrosine positron emission tomography, and tumor molecular factors. *Neurosurgery.* 2016;78(3):401–11.
7. Zhang B, Chang K, Ramkissoon S, et al. Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. *Neuro Oncol.* 2017;19(1):109–117.
8. Zhou H, Vallieres M, Bai HX, et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. *Neuro Oncol.* 2017;19(6):862–870.
9. Suh CH, Kim HS, Jung SC, Choi CG, Kim SJ. Imaging prediction of isocitrate dehydrogenase (IDH) mutation in patients with glioma: a systemic review and meta-analysis. *Eur Radiol.* 2019;29(2):745–758.
10. Di Ieva A, Magnussen JS, McIntosh J, et al. Magnetic resonance spectroscopic assessment of isocitrate dehydrogenase status in gliomas: the new frontiers of spectrobiopsy in neurodiagnostics. *World Neurosurg.* 2020;133:e421–e427.
11. Minniti G, Clarke E, Lanzetta G, et al. Stereotactic radiosurgery for brain metastases: analysis of outcome and risk of brain radionecrosis. *Radiat Oncol.* 2011;6:48.
12. Wen PY, Macdonald DR, Reardon DA, et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J Clin Oncol.* 2010;28(11):1963–1972.
13. Ouyang ZQ, Zheng GR, Duan XR, et al. Diagnostic accuracy of glioma pseudoprogression identification with positron emission tomography imaging: a systematic review and meta-analysis. *Quant Imaging Med Surg.* 2023;13(8):4943–4959.
14. Singnurkar A, Poon R, Detsky J. 18F-FET-PET imaging in high-grade gliomas and brain metastases: a systematic review and meta-analysis. *J Neurooncol.* 2023;161(1):1–12.
15. Pauleit D, Floeth F, Hamacher K, et al. *O*-(2-[18F]fluoroethyl)-L-tyrosine PET combined with MRI improves the diagnostic assessment of cerebral gliomas. *Brain.* 2005;128(Pt 3):678–687.
16. Wang L, Lieberman BP, Ploessl K, Kung HF. Synthesis and evaluation of 18F labeled FET prodrugs for tumor imaging. *Nucl Med Biol.* 2014;41(1):58–67.
17. Gallach M, Mikhail Lette M, Abdel-Wahab M, et al. Addressing global inequities in Positron Emission Tomography-Computed Tomography (PET-CT) for cancer management: a statistical model to guide strategic planning. *Med Sci Monit.* 2020;26:e926544.
18. Law I, Albert NL, Arbizu J, et al. Joint EANM/EANO/RANO practice guidelines/SNMMI procedure standards for imaging of gliomas using PET with radiolabelled amino acids and [18F]FDG: version 1.0. *Eur J Nucl Med Mol Imaging.* 2019;46(3):540–557.
19. Albert NL, Weller M, Suchorska B, et al. Response assessment in neuro-oncology working group and European Association for Neuro-Oncology recommendations for the clinical use of PET imaging in gliomas. *Neuro-Oncol.* 2016;18(9):1199–1208.
20. Albert NL, Galldiks N, Ellingson BM, et al. PET-based response assessment criteria for diffuse gliomas (PET RANO 1.0): a report of the RANO group. *Lancet Oncol.* 2024;25(1):e29–e41.
21. Kaji S, Kida S. Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiol Phys Technol.* 2019;12(3):235–248.
22. Mao X-J, Shen C, Yang Y-B. Image restoration using convolutional auto-encoders with symmetric skip connections. *arXiv preprint arXiv.* 2016.
23. Alotaibi A. Deep generative adversarial networks for image-to-image translation: a review. *Symmetry-Basel.* 2020;12(10):1705.
24. Li L, Fang Y, Wu J, Wang J, Ge Y. Encoder-decoder full residual deep networks for robust regression and spatiotemporal estimation. *IEEE Trans Neural Netw Learn Syst.* 2021;32(9):4217–4230.
25. Wang R, Liu H, Toyonaga T, et al. Generation of synthetic PET images of synaptic density and amyloid from 18F-FDG images using deep learning. *Med Phys.* 2021;48(9):5115–5129.
26. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys.* 2017;44(4):1408–1419.
27. Nie D, Trullo R, Lian J, et al. Medical image synthesis with deep convolutional adversarial networks. *IEEE Trans Biomed Eng.* 2018;65(12):2720–2730.
28. Xu J, Gong E, Ouyang J, Pauly J, Zaharchuk G. *Ultra-low-dose 18F-FDG Brain PET/MR Denoising using Deep Learning and Multi-contrast Information.* Vol 11313. SPIE; 2020.
29. Chen KT, Gong EH, Macruz FBD, et al. Ultra-low-dose 18F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. *Radiology.* 2019;290(3):649–656.
30. Ouyang J, Chen KT, Gong E, Pauly J, Zaharchuk G. Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. *Med Phys.* 2019;46(8):3555–3564.
31. Chen KT, Toueg TN, Koran MEI, et al. True ultra-low-dose amyloid PET/MRI enhanced with deep learning for clinical interpretation. *Eur J Nucl Med Mol Imaging.* 2021;48(8):2416–2425.
32. Ouyang J, Chen KT, Duarte Armindo R, et al. Predicting FDG-PET images from multi-contrast MRI using deep learning in patients with brain neoplasms. *J Magn Reson Imaging.* 2023;59(3):1010–1020.
33. Takita H, Matsumoto T, Tatekawa H, et al. AI-based virtual synthesis of methionine PET from contrast-enhanced MRI: development and external validation study. *Radiology.* 2023;308(2):e223016.
34. ANTs GitHub repository. https://github.com/ANTsX/ANTs. Date accessed March 22, 2024.
35. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. *Neuroimage.* 2012;62(2):782–790.
36. Di Ieva A, Russo C, Liu S, et al. Application of deep learning for automatic segmentation of brain tumors on magnetic resonance imaging: a heuristic approach in the clinical scenario. *Neuroradiology.* 2021;63(8):1253–1262.
37. Myronenko A. *3D MRI Brain Tumor Segmentation Using Autoencoder Regularization.* Cham; 2019.
38. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017;77(21):e104–e107.
39. Isola P, Zhu JY, Zhou TH, Efros AA. Image-to-image translation with conditional adversarial networks. 30th IEEE Conference on Computer Vision and Pattern Recognition (Cvpr 2017). 2017:5967–5976.
40. Smith KE, Smith AO. Conditional GAN for timeseries generation. *arXiv preprint arXiv:2006.16477.* 2020.
41. Hafliger P, Charles RP. The L-Type Amino Acid Transporter LAT1—An emerging target in cancer. *Int J Mol Sci .* 2019;20(10):2428.
42. Spaeth N, Wyss MT, Weber B, et al. Uptake of 18F-fluorocholine, 18F-fluoroethyl-L-tyrosine, and 18F-FDG in acute cerebral radiation injury in the rat: implications for separation of radiation necrosis from tumor recurrence. *J Nucl Med.* 2004;45(11):1931–1938.
43. Hayes AR, Jayamanne D, Hsiao E, et al. Utilizing 18F-fluoroethyltyrosine (FET) positron emission tomography (PET) to define suspected nonenhancing tumor for radiation therapy planning of glioblastoma. *Pract Radiat Oncol.* 2018;8(4):230–238.
44. Dunet V, Rossier C, Buck A, Stupp R, Prior JO. Performance of 18F-fluoro-ethyl-tyrosine (18F-FET) PET for the differential diagnosis of primary brain tumor: a systematic review and Metaanalysis. *J Nucl Med.* 2012;53(2):207–214.
45. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *arXiv preprint arXiv.* 2014.
46. Mirza M, Osindero S. Conditional generative adversarial nets. *arXiv preprint arXiv.* 2014.
47. Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: a review. *Med Image Anal.* 2019;58:101552.
48. Choi HJ, Seo M, Kim A, Park SH. Generation of conventional 18F-FDG PET images from 18F-Florbetaben PET images using generative adversarial network: a preliminary study using ADNI Dataset. *Medicina (Kaunas).* 2023;59(7):1281.