Neuro-Oncology Advances

4(1), 1–11, 2022 | https://doi.org/10.1093/noajnl/vdac060 | Advance Access date 22 April 2022

Combining radiomics and deep convolutional neural network features from preoperative MRI for predicting clinically relevant genetic biomarkers in glioblastoma

Evan Calabrese, Jeffrey D. Rudie, Andreas M. Rauschecker, Javier E. Villanueva-Meyer, Jennifer L. Clarke, David A. Solomon, and Soonmee Cha

Department of Radiology and Biomedical Imaging, University of California at San Francisco, San Francisco, California, USA (E.C., J.D.R., A.M.R., J.E.V.-M., S.C.); Center for Intelligent Imaging, University of California San Francisco, San Francisco, California, USA (E.C., J.E.V.-M.); Division of Neuro-Oncology, Departments of Neurology and Neurological Surgery, University of California San Francisco, San Francisco, California, USA (J.L.C.); Department of Pathology, University of California San Francisco, San Francisco, California, USA (D.A.S.); Clinical Cancer Genomics Laboratory, University of California San Francisco, San Francisco, California, USA (D.A.S.)

Corresponding Author: Evan Calabrese, MD, PhD, Department of Radiology and Biomedical Imaging, University of California San Francisco, 350 Parnassus Ave, Suite 307H, San Francisco, CA 94143-0628, USA (evan.calabrese@ucsf.edu).

Abstract

Background. Glioblastoma is the most common primary brain malignancy, yet treatment options are limited, and prognosis remains guarded. Individualized tumor genetic assessment has become important for accurate prognosis and for guiding emerging targeted therapies. However, challenges remain for widespread tumor genetic testing due to costs and the need for tissue sampling. The aim of this study is to evaluate a novel artificial intelligence method for predicting clinically relevant genetic biomarkers from preoperative brain MRI in patients with glioblastoma.

Methods. We retrospectively analyzed preoperative MRI data from 400 patients with glioblastoma, IDH-wildtype or WHO grade 4 astrocytoma, IDH mutant who underwent resection and genetic testing. Nine genetic biomarkers were assessed: hotspot mutations of *IDH1* or *TERT* promoter, pathogenic mutations of *TP53*, *PTEN*, *ATRX*, or *CDKN2A/B*, *MGMT* promoter methylation, *EGFR* amplification, and combined aneuploidy of chromosomes 7 and 10. Models were developed to predict biomarker status from MRI data using radiomics features, convolutional neural network (CNN) features, and a combination of both.

Results. Combined model performance was good for *IDH1* and *TERT* promoter hotspot mutations, pathogenic mutations of *ATRX* and *CDKN2A/B*, and combined aneuploidy of chromosomes 7 and 10, with receiver operating characteristic area under the curve (ROC AUC) >0.85 and was fair for all other tested biomarkers with ROC AUC >0.7. Combined model performance was statistically superior to individual radiomics and CNN feature models for prediction chromosome 7 and 10 aneuploidy, *MGMT* promoter methylation, and *PTEN* mutation.

Conclusions. Combining radiomics and CNN features from preoperative MRI yields improved noninvasive genetic biomarker prediction performance in patients with WHO grade 4 diffuse astrocytic gliomas.

Key Points

- We evaluated artificial intelligence models for predicting glioblastoma genetics from MRI.
- We assessed 9 different genetic biomarkers that are relevant for treatment or prognosis.
- Combining radiomics and convolutional image features yielded improved performance.

© The Author(s) 2022. Published by Oxford University Press, the Society for Neuro-Oncology and the European Association of Neuro-Oncology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Importance of the Study

Accurate radiogenomic assessment of preoperative brain MRI has the potential to improve care of patients with glioblastoma and could be particularly useful for helping to guide the use of emerging targeted therapies. This study demonstrates accurate preoperative radiogenomic assessment of several common glioblastoma genetic biomarkers including some that are relevant to investigational therapies. We further demonstrate that combining information from radiomics and deep learning feature extraction methods can lead to improved radiogenomic prediction accuracy compared to using either type of feature alone.

Glioblastoma is the most common primary brain malignancy but remains difficult to treat despite advances in surgical, radiation, and medical therapies.[1,2] The recent discovery that histologically defined glioblastoma encompasses several distinct tumor entities has accelerated progress in precision therapy targeting specific molecular biomarkers.[3,4] However, challenges remain for widespread tumor genetic testing due to costs and the need for adequate tissue sampling. In addition, in rare cases where biopsy is not possible or declined by the patient, glioblastoma is treated presumptively based on imaging alone, which precludes genetic testing. As our understanding of glioblastoma genetics improves, there is an increasing need for reliable, noninvasive methods for predicting genetic biomarkers relevant to prognosis and guiding targeted therapies.

Several prior studies have demonstrated that various clinically relevant glioblastoma genetic biomarkers can be inferred based on preoperative MRI using image analysis techniques. Most prior work has focused on image features extracted using radiomics or convolutional neural networks (CNNs) as the primary image-based predictors.[5–9] Radiomics features are a set of human-defined morphologic, grayscale, and pattern-based features that can be applied across multiple different imaging modalities and have proven useful for many different medical image-based inference tasks.[10] Radiomics features are relatively straightforward to define, conceptualize, and interpret, and are both standardized and reproducible.[11,12] In contrast, CNN-based features are abstract and often difficult to interpret multiscale features that are "learned" automatically using CNNs. CNN features are unique to each input dataset, which allows considerable versatility but also introduces susceptibility to overfitting and lack of reproducibility.[13] Relatively few prior studies have evaluated a combination of radiomics and CNN features for oncologic inference tasks[14–16] and studies specifically assessing genetic biomarker prediction in glioblastoma are lacking.

In this study, we present a hybrid approach for preoperative MRI-based glioblastoma genetic biomarker prediction, which combines both radiomics features and CNN features into a single unified artificial intelligence architecture. Our method leverages the strengths of both predefined radiomics features and learned CNN features to provide improved prediction performance. Predictive models were evaluated using a relatively large (n = 400) patient cohort with a diverse set of genetic biomarkers, many of which are useful either for determining prognosis or for guiding investigational targeted therapies.

Materials and Methods

Patient Population

All studies were performed in accordance with applicable guidelines and regulations and were approved by an institutional review board with a waiver for informed consent. The retrospective study cohort consisted of 400 adult patients with CNS WHO grade 4 diffuse astrocytic gliomas that had a final integrated diagnosis of either "Glioblastoma, IDH-wildtype" or "Astrocytoma, IDH-mutant" using the 2021 World Health Organization (WHO) Classification of Central Nervous System Tumors. All patients underwent preoperative MRI, initial tumor resection, and tumor genetic testing at the University of California San Francisco medical center between 2015 and 2021. The study cohort was identified by searching radiology and pathology databases (Nuance mPower) for all patients who received a preoperative planning MRI and had subsequent surgical pathology report within the following month with a final integrated diagnosis of glioblastoma, IDH-wildtype or astrocytoma, IDH-mutant, CNS WHO grade 4. Exclusion criteria included any prior history of brain tumor resection (n = 18) and incomplete or technically inadequate imaging studies (n = 28).

Genetic Biomarker Testing

Nine different pathogenic glioblastoma genetic biomarkers were evaluated: mutations or deletions of *IDH1*, *TP53*, *PTEN*, *ATRX*, *TERT*, and *CDKN2A/B*, *MGMT* promoter methylation, *EGFR* copy number amplification (including accompanying *EGFRVIII* rearrangement), and combined aneuploidy of chromosomes 7 and 10 (most commonly as the combination of trisomy 7 and monosomy 10).[17] Genetic biomarkers were chosen based on prevalence in the study cohort as well as their utility for determining tumor type (eg, *IDH1*, *ATRX*), informing chemotherapeutic management (eg, *MGMT*), determining prognosis (eg, *IDH*, *PTEN*), and/or guiding targeted therapies (eg, *CDKN2A/B*, *EGFR*). Gold standard molecular characterization was performed as part of routine clinical care at the time of diagnosis without knowledge of the imaging predictors developed during this study. Methods varied during the study period depending on the test availability and insurance reimbursement. Three hundred and five cases in the cohort were characterized with a capture-based targeted next-generation

DNA sequencing panel as previously described.[18] For the remaining cases, molecular characterization was accomplished by fluorescence in situ hybridization (PTEN, EGFR) and immunohistochemical staining (IDH1, ATRX, TP53).[19,20] All *IDH* mutations identified by immunohistochemistry were confirmed by genetic sequencing. *MGMT* promoter methylation status was determined using a quantitative methylation sensitive PCR assay.[21]

Image Acquisition

A standardized preoperative brain tumor MRI protocol was performed at 3.0 Tesla (Discovery 750, GE Healthcare) with a dedicated 8-channel head coil (Invivo). The imaging protocol included 6 3D sequences: T2-weighted, T2-weighted FLAIR, susceptibility-weighted (SWI), pre- and postcontrast T1-weighted images, and arterial spin labeling (ASL) perfusion images, as well as 2 2D sequences: isotropic diffusion-weighted images (DWIs) and 55-direction high angular resolution diffusion imaging (HARDI). Complete parameters for the standardized MRI protocol are provided elsewhere.[22] Some image data from this study are publicly available via the 2021 BraTS challenge, and we plan to share the remainder as a separate collection.[23,24]

Image Preprocessing

All image data were deidentified and converted to NIfTI format using dcm2niix v1.0 (https://github.com/rordenlab/dcm2niix) using default parameters. HARDI data were eddy current corrected and processed using the Eddy and DTIFIT modules from the Functional MRI of the Brain Software Library (FSL) version 6.0.2 yielding mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), and fractional anisotropy (FA) images.[25] All images were registered and resampled to the 3D space defined by the T2/FLAIR image (1-mm isotropic resolution) using automated nonlinear registration (Advanced Normalization Tools 2.3.5).[26,27] Resampled coregistered data were then skull stripped (including face removal) using a previously described and publicly available method (https://github.com/ecalabr/brain_mask).[22] Finally, all images underwent N4 coil bias correction and intensity normalization.[28] Processing time varied from approximately 18 min to approximately 35 min per study using a standard desktop computer with a 16-core processor, 128 GB of RAM, and an Nvidia RTX Titan graphics processing unit.

Deep Learning-Based Automated Tumor Subcompartment Segmentation

A previously validated deep learning algorithm was used to generate automated 3D segmentations of 3 key components of glioblastoma that are seen on MRI: enhancing and central nonenhancing/necrotic tumor (together comprising the tumor core) and the surrounding T2/FLAIR abnormality typically referred to as "edema," which may also contain infiltrative nonenhancing tumor outside of the tumor core.[29,30] All segmentations were manually corrected by a group of trained radiologists with varying experience (including authors EC and JR) using ITK-SNAP v3.8.0[31] and subsequently approved by 2 attending neuroradiologists with >15 years of experience each as part of the 2021 BraTS challenge.[24] Manual annotators were not provided any molecular characterization data.

Radiomics Feature Extraction

Radiomics features were extracted using PyRadiomics 2.2.0 command line tools with default extraction parameters.[32] Input images included 11 image contrasts: T1 pre, T1 post, T2, T2/FLAIR, SWI, DWI, ASL, MD, AD, RD, and FA. Input segmentations included 5 different tumor components: whole tumor, tumor core, and each of the 3 individual tumor compartments. The default set of shape features (n = 26), first-order grayscale features (n = 19), and higher-order grayscale features (n = 75) were extracted yielding 5300 radiomics features per patient.

Radiomics Feature Selection

An automated radiomics feature selection process was employed to select the 32 best features for inference using scikit-learn 0.24 with a 5-fold cross-validation approach to reduce selection bias. First, univariate feature selection with mutual information was used to select the 1024 best correlated features. Next, the best features for each cross-validation fold were ranked using recursive feature elimination with a random forest classifier, and the 32 features with the best average rank across folds were selected as the final feature set. Parameters for recursive feature elimination included 1000 estimators for the random forest classifier and a step size of 16 features for each feature elimination step.

Network Architecture

A CNN classifier (CNN limb) was constructed using TensorFlow 2.4. The CNN limb consisted of a 3D multiscale deep convolutional autoencoder with 1, 1, 2, and 2 bottleneck residual blocks per level, respectively, and a max-pool down-sampling layer (pool size = [2, 2, 2]) between each level.[33] Bottleneck blocks included leaky ReLU activation, batch normalization, and 40% feature dropout. The number of features per layer was set at 32 for the top level and doubled after each pooling step. A parallel random forest classifier limb was constructed for radiomics feature inference (radiomics limb) using scikit-learn 0.24. The random forest classifier parameters included 32 input features and 1000 estimators. Each limb yielded a single output logit, which was converted to a probability using the sigmoid logistic function. The 2 output probabilities (1 from the CNN limb and 1 from the radiomics limb) were averaged to create a final combined model probability. A probability threshold of $P \geq .5$ was used to determine a positive result.

Network Inputs and Training

Network training was implemented with 5-fold cross-validation and a train/test split of 80%/20%. Separate

networks were trained for each genetic biomarker. CNN limb inputs consisted of 96 × 96 × 96 voxel cubes encompassing the whole tumor (including surrounding FLAIR abnormality) after scaling to fit with a 5-voxel margin in each dimension. Data augmentation included random dimension swaps and rotations of ±0–90 degrees in each of the cardinal axes. Training data were sampled such that each batch consisted of an equal number of positive and negative cases. Model weights were initialized using the Glorot initializer.[34] Training comprised 40 epochs with a batch size of 10 cases on an Nvidia DGX-2 system. Neither training validation nor early stopping were used. Binary cross entropy loss between predictions and genetic biomarker results was minimized using the Adam optimizer.[35] Radiomics limb inputs consisted of the 32 best radiomics features identified during feature selection. Radiomics data were balanced using SMOTE feature synthesis.[36] Training was accomplished using the built in "fit" method from the sci-kit learn RandomForestClassifier class.

Network Performance Evaluation

Genetic biomarker prediction was evaluated using receiver operating characteristic (ROC) analysis as well as several different scalar metrics for the ROC best point: sensitivity, specificity, accuracy, precision, the *F*1 statistic, the Matthews correlation coefficient (MCC), and Youden's index. Statistically significant differences between ROC curves were assessed using DeLong's method with a threshold of $P < .05$.[37,38]

Results

Genetic Biomarker Testing

Genetic biomarker testing was not uniform across the patient cohort. Gene test frequency, included variants, and prevalence in the study cohort are presented in Table 1.[17,39,40]

In addition to the listed biomarkers, all cases were also tested for *IDH2* p.R172 mutations, however, none were identified in the study cohort. Of note, 29 of 400 patients in the cohort had tumors with *IDH1* p.R132 hotspot mutations corresponding with an integrated diagnosis of "Astrocytoma, IDH-mutant, CNS WHO grade 4". While these tumors are now classified independently of IDH-wildtype glioblastomas in the 2021 WHO Classification of CNS Tumors, they were included in the study cohort to allow prediction of *IDH* mutations.

Image Preprocessing, Segmentation, and Radiomics Feature Extraction

Figure 1 shows a schematic flow chart of the relevant image preprocessing, tumor segmentation, and radiomics feature extraction steps including all instances where user input was required. All study data were successfully processed through all steps of the pipeline. Data processing time varied depending on the amount of manual input required, but automated steps were accomplished in approximately 30 min per case.

Deep Neural Network Architecture

A graphical representation of the deep CNN architecture used for genetic biomarker inference is provided in Supplementary Figure 1. The total number of trainable parameters was 3 564 279 for each CNN instance. Identical models were trained for each of the evaluated genetic biomarkers. Code for the CNN limb is available at https://www.github.com/ecalabr/gliomarad.

Model Inputs and Outputs

Example image inputs and genetic biomarker prediction outputs for a single patient are shown in Supplementary Figure 2. Image inputs included 3D image volumes of the

Table 1. Genetic biomarker testing details for the study cohort

Genetic biomarker	Positive/ tested	Male/ female	Age ± Std.	Variants included
IDH1	29/400	241/159	60 ± 13	Any p.R132 hotspot mutation (most often p.R132H)
MGMT	209/381	233/148	60 ± 13	Methylation of ≥2/17 promoter CpG sites
TP53	174/398	239/159	60 ± 13	Any pathogenic mutation or deletion
PTEN	210/399	240/159	60 ± 13	Any pathogenic mutation or deletion
ATRX	34/396	238/158	60 ± 13	Any pathogenic mutation or deletion
TERT	239/305	178/127	60 ± 14	Either c.-146C>T or c.-124C>T promoter hotspot mutation
CDKN2A/B	200/305	178/127	60 ± 14	Homozygous/biallelic deletion of *CDKN2A* and *CDKN2B*
EGFR	101/399	240/159	60 ± 13	Focal gene amplification (including accompanying *EGFRVIII* structural rearrangement)
Chromosomes 7 and 10	221/305	178/127	60 ± 14	Combined polysomy of 7 and monosomy of 10

For each biomarker, the number of positive cases and total number of tested cases are listed along with patient sex and average age. Specific genetic variants that were considered are listed in the far-right column.

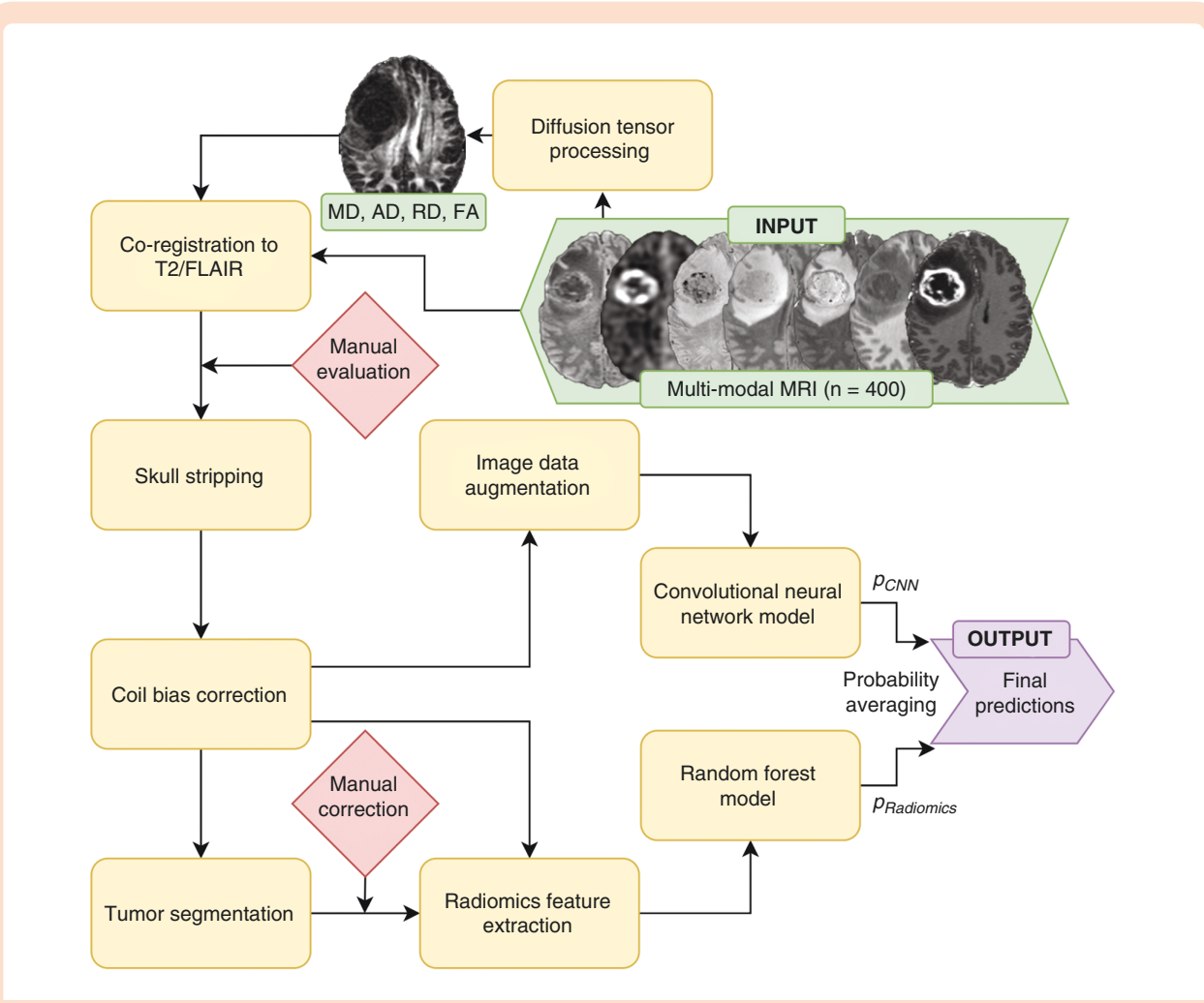

Figure 1. Flow chart of image processing steps. MR images (INPUT) were passed through several sequential automated image processing steps (rounded boxes) with minimal human input (diamond boxes). Each limb of the prediction model (convolutional neural network and radiomics) yields a single output logit, which is transformed into a probability using the sigmoid function. Final predictions (OUTPUT) were generated by averaging probabilities from the 2 limbs. Diffusion tensor derived contrasts include: mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), and fractional anisotropy (FA).

entire tumor region for each of the 11 MRI contrasts included in the study. Outputs consist of probabilities of each individual genetic biomarker being present from the radiomics and CNN models, respectively, as well as a combined probability.

Genetic Biomarker Prediction Performance

ROC curves for each genetic biomarker are shown in Figures 2–4, and a summary of evaluation metrics is provided in Table 2. Both radiomics and CNN methods were able to predict each genetic biomarker with performance that could not be explained by chance alone. For some genetic biomarkers, such as *CDKN2A/B* homozygous deletion, the CNN feature-based model showed better performance by all metrics, while for others, such as *TERT* promoter mutation, the radiomics feature-based model showed better performance. Combined model receiver operating characteristic area under the curve (ROC AUC) was highest for *ATRX* (0.97) and *IDH1* (0.96) mutations; however, these biomarkers had the lowest prevalence in the study cohort and therefore may be better assessed using metrics that account for class imbalance such as precision (0.42 and 0.65, respectively) or MCC (0.60 and 0.64, respectively). Combined model ROC AUC was lowest for *MGMT* promoter methylation (0.77) and *PTEN* mutation/deletion (0.77).

Comparison With Isolated Radiomics or CNN Feature-Based Models

The combined radiomics and CNN feature model showed improved ROC AUC compared to either the radiomics or CNN models individually for all evaluated genetic biomarkers. Combined model ROC AUC was statistically superior (DeLong's $P < .05$) to at least one of the individual models (radiomics or CNN) for all genetic biomarkers except for *IDH1* mutations and was superior to both individual models (radiomics and CNN) for combined chromosome 7 and 10 aneuploidy, *MGMT* promoter methylation, and *PTEN* mutation/deletion. Combined model performance by other metrics

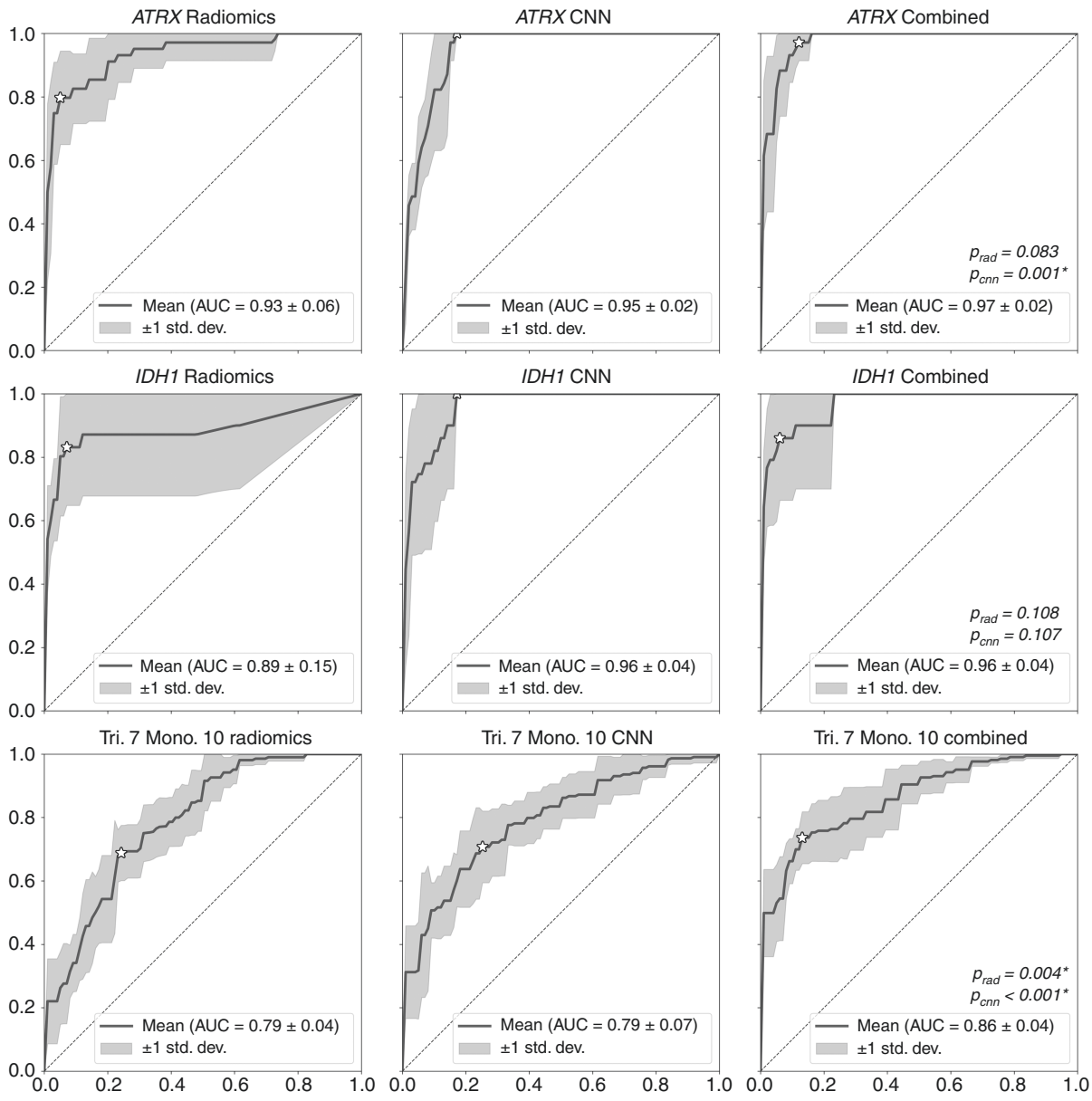

Figure 2. Fivefold cross-validation receiver operating characteristic curves for radiomics, convolutional neural network, and combine models for predicting pathogenic alterations in *ATRX*, *IDH1*, and chromosomes 7 and 10. Best points are indicated with an outlined star. Average area under the curve is included in the legend of each subfigure. *P* values for DeLong's test for ROC curve difference between the combined model and the radiomics (p_{rad}) and CNN (p_{cnn}) models, respectively, are reported with "*" denoting statistical significance. CNN, convolutional neural network; ROC, receiver operating characteristic.

was variable but generally better than or comparable to the better of the 2 individual models. In some cases, select combined model metrics were slightly worse than the better of the 2 individual models (eg, the precision metric for *ATRX*).

Discussion

In this study, we evaluated an artificial intelligence architecture that combines both radiomics and CNN image features into a single model for predicting genetic biomarkers in glioblastoma. We found that combining these 2 types of image features resulted in significantly improved genetic biomarker prediction performance in most cases and synergistic improvement in some cases. In our dataset, prediction performance of the individual radiomics and CNN models varied significantly between genetic biomarkers with the radiomics model performing better in some cases and the CNN model performing better in others. This highlights the fact that radiomics and CNN feature relevance may vary significantly depending on the specific inference task. One potential advantage of the proposed

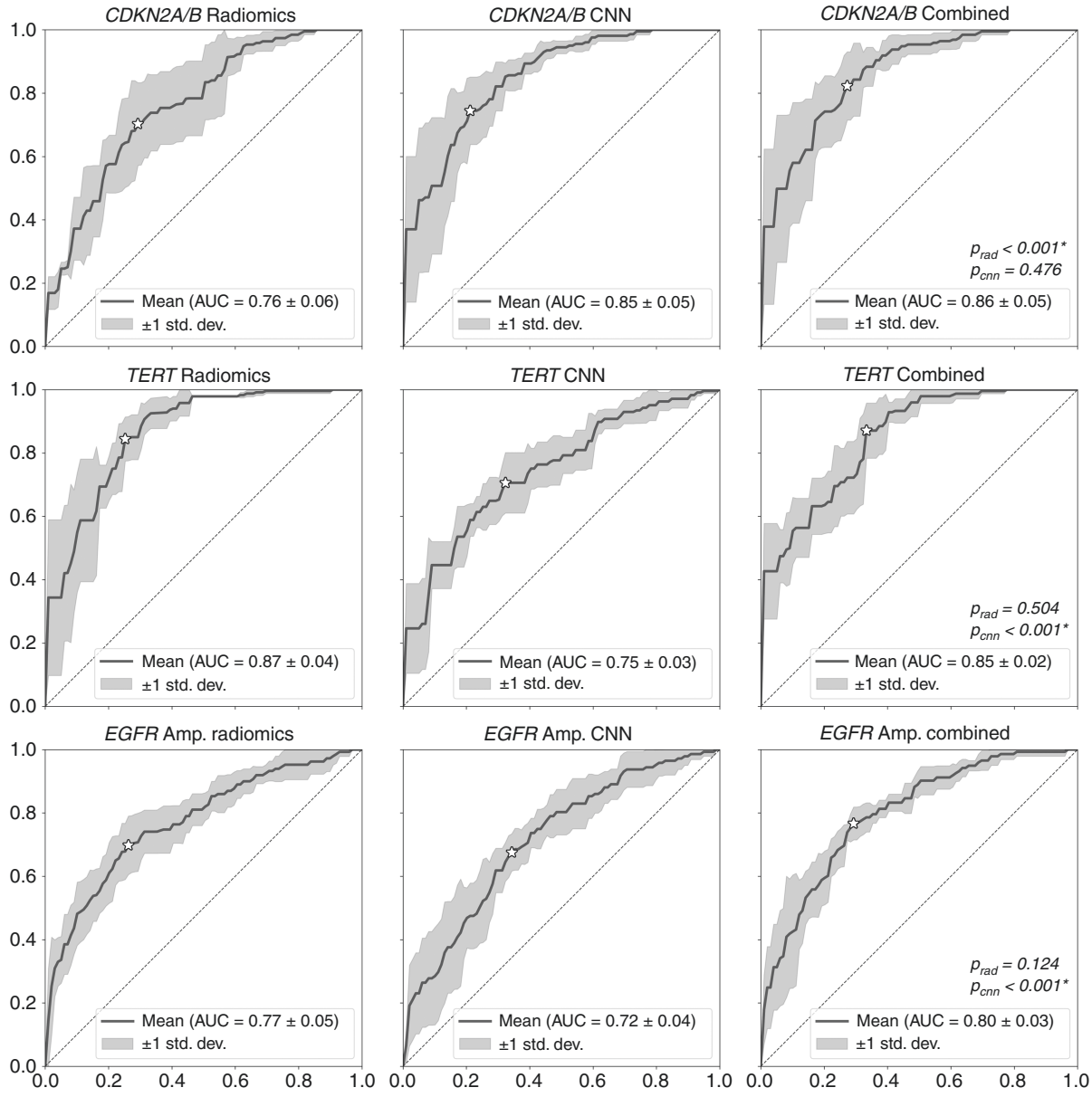

Figure 3. Fivefold cross-validation receiver operating characteristic curves for radiomics, convolutional neural network, and combine models for predicting pathogenic alterations in *CDKN2A/B*, *TERT* promoter, and *EGFR*. Best points are indicated with an outlined star. Average area under the curve is included in the legend of each subfigure. *P* values for DeLong's test for ROC curve difference between the combined model and the radiomics (p_{rad}) and CNN (p_{cnn}) models, respectively, are reported with "*" denoting statistical significance. CNN, convolutional neural network; ROC, receiver operating characteristic.

combined approach is that it can balance the contribution of radiomics and CNN features and can perform equally well for tasks that are better suited for either type of image feature.

In addition to presenting a new method for genetic biomarker prediction in glioblastoma, this study provides new insight into the feasibility of image-based genetic prediction for several genetic biomarkers that have not previously been well evaluated. For example, our model was able to predict *CDKN2A/B* homozygous deletion with a ROC AUC of 0.86 and accuracy of 79%. This genetic biomarker is present in approximately 60% of IDH-wildtype glioblastomas, and can be targeted using small molecule inhibitors of CDK4/6, which have shown promise for treating other tumor types and are currently being evaluated in glioblastoma.[41,42] These results suggest that image-based genetic inference may be useful for a variety of different glioblastoma genetic biomarkers beyond *IDH1* and *MGMT*, which have largely dominated prior related work. However, image-based genetic inference can only be accurate if the target biomarker produces an imaging phenotype. It is conceivable, that some genetic alterations may not yield

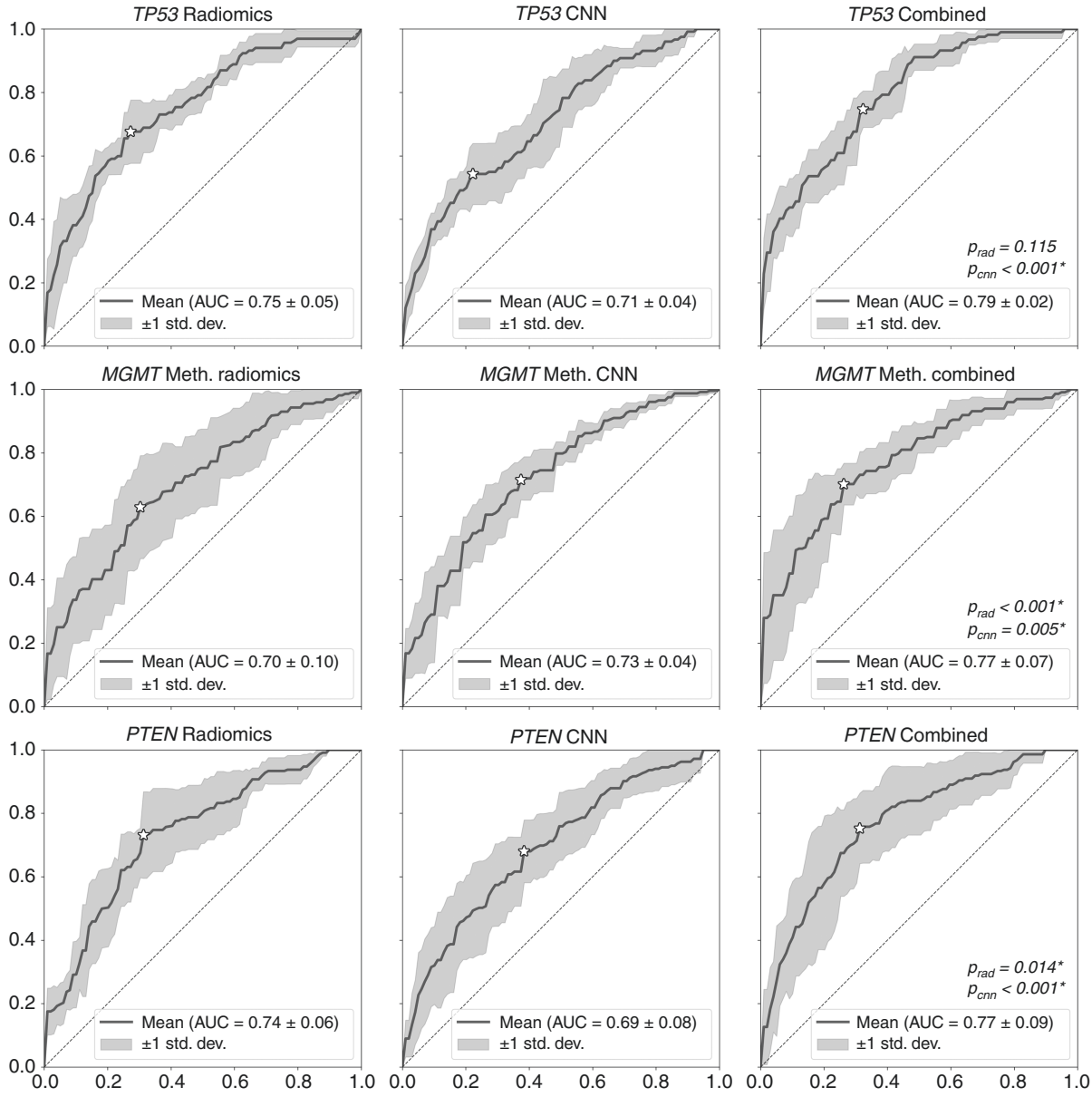

Figure 4. Fivefold cross-validation receiver operating characteristic curves for radiomics, convolutional neural network, and combine models for predicting pathogenic alterations in *TP53*, *MGMT*, and *PTEN*. Best points are indicated with an outlined star. Average area under the curve is included in the legend of each subfigure. *P* values for DeLong's test for ROC curve difference between the combined model and the radiomics (p_{rad}) and CNN (p_{cnn}) models, respectively, are reported with "*" denoting statistical significance. CNN, convolutional neural network; ROC, receiver operating characteristic.

discernable imaging phenotypes and therefore may not be amenable to image-based inference. By similar logic, genetic alterations that frequently occur together may be difficult to distinguish. For example, in our study cohort, only 8 patients had either *ATRX* or *IDH1* mutation, but not both, and the combined model was only able to identify 4/8 (50%) of these discordant mutations. The low prevalence of these genetic biomarkers also highlights the issue of class imbalance, which can lead to overly optimistic ROC AUC values. For example, the best ROC AUC values in our study were achieved for biomarkers with the lowest prevalence (ie, *ATRX*, *IDH1*). Model performance on biomarkers with low prevalence may therefore be better evaluated using metrics that account for class imbalance such as precision or MCC.

One major limitation of this study is the lack of external validation. Unfortunately, glioblastoma genetic testing is still not widely adopted, particularly for newer less studied genetic biomarkers. In addition, there is wide variation in preoperative brain tumor MR imaging protocols in terms of sequences, field strength, and acquisition dimensionality. As result, many publicly available glioblastoma MRI

Table 2. Performance metrics for the, radiomics, convolutional neural network (CNN), and combined models for each genetic biomarker

Biomarker	Model	AUC	$p_{combined}$	Sens.	Spec.	Acc.	Prec.	*F*1	MCC	Youden
ATRX	Radiomics	0.93	.083	0.80	0.95	0.94	0.59	0.68	0.65	0.75
	CNN	0.95	.001*	1.00	0.83	0.84	0.34	0.51	0.53	0.83
	Combined	0.97		0.97	0.88	0.89	0.42	0.59	0.60	0.85
IDH1	Radiomics	0.89	.108	0.83	0.93	0.92	0.47	0.60	0.59	0.76
	CNN	0.96	.107	1.00	0.83	0.84	0.30	0.47	0.50	0.83
	Combined	0.96		0.86	0.94	0.93	0.52	0.65	0.64	0.80
Trisomy 7 Monosomy 10	Radiomics	0.79	.004*	0.69	0.76	0.71	0.88	0.77	0.40	0.45
	CNN	0.79	<.001*	0.71	0.75	0.72	0.88	0.78	0.41	0.45
	Combined	0.86		0.74	0.87	0.77	0.94	0.82	0.55	0.60
CDKN2A/B	Radiomics	0.76	<.001*	0.70	0.71	0.70	0.82	0.76	0.39	0.41
	CNN	0.85	.476	0.74	0.79	0.76	0.87	0.80	0.51	0.53
	Combined	0.86		0.82	0.73	0.79	0.85	0.84	0.54	0.55
TERT	Radiomics	0.87	.504	0.84	0.75	0.82	0.92	0.88	0.54	0.59
	CNN	0.75	<.001*	0.71	0.68	0.70	0.89	0.79	0.33	0.38
	Combined	0.85		0.87	0.67	0.83	0.90	0.89	0.52	0.54
EGFR amplification	Radiomics	0.77	.124	0.70	0.74	0.72	0.59	0.64	0.42	0.44
	CNN	0.72	<.001*	0.68	0.66	0.66	0.51	0.58	0.32	0.33
	Combined	0.80		0.77	0.71	0.73	0.58	0.66	0.45	0.47
TP53	Radiomics	0.75	.115	0.68	0.73	0.71	0.50	0.58	0.37	0.40
	CNN	0.71	<.001*	0.54	0.78	0.71	0.50	0.52	0.31	0.32
	Combined	0.79		0.75	0.68	0.70	0.48	0.59	0.39	0.42
MGMT promoter methylation	Radiomics	0.70	<.001*	0.63	0.70	0.66	0.77	0.69	0.32	0.33
	CNN	0.73	.005*	0.72	0.63	0.68	0.75	0.73	0.34	0.34
	Combined	0.77		0.70	0.74	0.72	0.81	0.75	0.43	0.44
TEN	Radiomics	0.74	.014*	0.73	0.69	0.71	0.72	0.73	0.42	0.42
	CNN	0.69	<.001*	0.68	0.62	0.65	0.66	0.67	0.30	0.30
	Combined	0.77		0.75	0.69	0.72	0.72	0.74	0.44	0.44

Acc., accuracy; *F*1, *F*1 statistic; MCC, Matthew's correlation coefficient; Prec., precision; ROC AUC, receiver operating characteristic area under the curve; Sens., sensitivity; Spec., specificity; $p_{combined}$: *P* value for DeLong's test for ROC curve difference compared to the combined model. "*" denotes $P < .05$.

datasets do not provide genetic data and/or include only a limited number of MRI sequences (typically T1, T2, FLAIR, and postcontrast imaging). For these reasons, there are no publicly available datasets that could reasonably be used to externally validate the models proposed in this study. It should be noted that recent efforts to create a generalizable model for *MGMT* promoter methylation status prediction in a multi-institution glioblastoma MRI dataset yielded relatively poor performance, which is presumed to be related to data heterogeneity and/or the limited variety of available MRI sequences.[24] Given these findings, presented results should be interpreted as "best case scenario" as performance on external data would almost certainly be worse. However, as the clinical importance of glioblastoma genetics continues to grow and preoperative MRI protocols continue to improve, it seems likely that data relevant to this study will become more widely available in the future. To this end, the codebase for this study has been made publicly available at https://www.github.com/ecalabr/gliomarad and image data are publicly available via the 2021 BraTS challenge.[23,24]

As the clinical potential of glioblastoma radiogenomics continues to mature, there is increasing need for new predictive modeling methods that can better utilize available image data. Our results suggest that there may be a role for combining radiomics and CNN features for improved genetic biomarker prediction accuracy in glioblastoma. This approach has shown promise for other radiologic/oncologic inference tasks and may ultimately prove to be an important step toward integration of image-based genetic inference into clinical workflows. Future work will be necessary to evaluate the optimal architectures for combining radiomics and CNN features and for assessing generalizability and portability to other datasets.

Supplementary Material

Supplementary material is available at *Neuro-Oncology Advances* online.

Keywords

artificial intelligence | deep learning | glioblastoma | radiogenomics | radiomics.

Funding

Research was supported by the National Institutes of Health (NIH) Ruth L. Kirschstein Institutional National Research Service Award under award number T32EB001631 and by the Radiological Society of North America Research & Education (RSNA R&E) Foundation under grant number RR2011 to E.C. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or RSNA R&E Foundation. Genomic profiling was supported by the UCSF Glioblastoma Precision Medicine Program sponsored by the Sandler Foundation.

Acknowledgments

The authors acknowledge the University of California San Francisco resident research fund, which provided financial support for graphics processing hardware.

Conflict of interest statement. None declared.

Authorship Statement. Experimental design, analysis, and interpretation: E.C., J.D.R., A.M.R., J.E.V.-M., J.L.C., D.A.S., and S.C. Manuscript writing and revision: E.C., J.D.R., A.M.R., J.E.V.-M., J.L.C., D.A.S., and S.C. Final approval: E.C., J.D.R., A.M.R., J.E.V.-M., J.L.C., D.A.S., and S.C. Computational and statistical analysis: E.C., J.D.R., and A.M.R.

References

1. Krex D, Klink B, Hartmann C, et al. Long-term survival with glioblastoma multiforme. *Brain.* 2007;130(10):2596–2606.
2. Stupp R, Hegi ME, Mason WP. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol.* 2009;10(5):459-466. doi:10.1016/S1470-2045(09)70025-7. Epub 2009 Mar 9. PMID: 19269895
3. Touat M, Idbaih A, Sanson M, Ligon KL. Glioblastoma targeted therapy: updated approaches from recent biological insights. *Ann Oncol.* 2017;28(7):1457–1472.
4. Weller M, Stupp R, Hegi M, Wick W. Individualized targeted therapy for glioblastoma: fact or fiction? *Cancer J.* 2012;18(1):40–44.
5. Li ZC, Bai H, Sun Q, et al. Multiregional radiomics features from multiparametric MRI for prediction of MGMT methylation status in glioblastoma multiforme: a multicentre study. *Eur Radiol.* 2018;28(9):3640–3650.
6. Zhang J, Barboriak DP, Hobbs H, Mazurowski MA. A fully automatic extraction of magnetic resonance image features in glioblastoma patients: automatic extraction of MRI features in glioblastoma patients. *Med Phys.* 2014;41(4):042301.
7. Chang P, Grinband J, Weinberg BD, et al. Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas. *Am J Neuroradiol.* 2018;39(7):1201–1207.
8. Han L, Kamdar MR. MRI to MGMT: predicting methylation status in glioblastoma patients using convolutional recurrent neural networks. In: Biocomputing 2018. World Scientific; 2017:331–342. PMID: 29218894; PMCID: PMC5728677. PSB 2018; January 3-7; 2018; Fairmont Orchid, Big Island of Hawaii.
9. Chang K, Bai HX, Zhou H, et al. Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging. *Clin Cancer Res.* 2018;24(5):1073–1081.
10. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* 2014;5(1):1–9.4006.
11. Zhao B, Tan Y, Tsai WY, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep.* 2016;6:23428. doi:10.1038/srep23428. PMID: 27009765; PMCID: PMC4806325.
12. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology.* 2020;295(2):328–338.
13. Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput.* 2017;29(9):2352–2449.
14. Lao J, Chen Y, Li ZC, et al. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Sci Rep.* 2017;7(1):1–8.
15. Li Z, Wang Y, Yu J, Guo Y, Cao W. Deep learning based radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci Rep.* 2017;7(1):1–11.
16. Xiao T, Hua W, Li C, Wang S. Glioma grading prediction by exploring radiomics and deep learning features. In: Proceedings of the Third International Symposium on Image Computing and Digital Medicine. ISICDM 2019. Association for Computing Machinery; Aug 24-26; 2019; Sofitel Xian on Renmin Square Hotel, Xian China.
17. Goodenberger ML, Jenkins RB. Genetics of adult glioma. *Cancer Genet.* 2012;205(12):613–621.
18. Kline CN, Joseph NM, Grenert JP, et al. Targeted next-generation sequencing of pediatric neuro-oncology patients improves diagnosis, identifies pathogenic germline mutations, and directs targeted therapy. *Neuro Oncol.* 2017;19(5):699–709.
19. Chaurasia A, Park SH, Seo JW, Park CK. Immunohistochemical analysis of ATRX, IDH1 and p53 in glioblastoma and their correlations with patient survival. *J Korean Med Sci.* 2016;31(8):1208–1214.
20. Takano S, Ishikawa E, Sakamoto N, et al. Immunohistochemistry on IDH 1/2, ATRX, p53 and Ki-67 substitute molecular genetic testing and predict patient prognosis in grade III adult diffuse gliomas. *Brain Tumor Pathol.* 2016;33(2):107–116.
21. Kitange GJ, Carlson BL, Mladek AC, et al. Evaluation of MGMT promoter methylation status and correlation with temozolomide response in orthotopic glioblastoma xenograft model. *J Neurooncol.* 2009;92(1):23–31.
22. Calabrese E, Rudie JD, Rauschecker AM, Villanueva-Meyer JE, Cha S. Feasibility of simulated postcontrast MRI of glioblastomas and lower-grade gliomas by using three-dimensional fully convolutional neural networks. *Radio: Art Intell.* (2021);3(5):e200276.

23. Calabrese E, Villanueva-Meyer J, Rudie J, et al. The University of California San Francisco Preoperative Diffuse Glioma (UCSF-PDGM) MRI dataset [published online ahead of print August 30, 2021], arXiv, arXiv:210900356 [cs, eess], preprint: not peer reviewed. https://doi.org/10.48550/arXiv.2109.00356. Accessed January 4, 2022.
24. Baid U, Ghodasara S, Bilello M, et al. The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification dataset [published online ahead of print July 5, 2021], arXiv, arXiv:210702314 [cs], preprint: not peer reviewed. https://doi.org/10.48550/arXiv.2107.02314. Accessed July 9, 2021.
25. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. *Neuroimage.* 2012;62(2):782–790.
26. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage.* 2011;54(3):2033–2044.
27. Avants B, Epstein C, Grossman M, Gee J. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal.* 2008;12(1):26–41.
28. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging.* 2010;29(6):1310–1320.
29. Bakas S, Reyes M, Jakab A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge [published online ahead of print November 5, 2018], arXiv, arXiv:181102629 [cs, stat], preprint: not peer reviewed. https://doi.org/10.48550/arXiv.1811.02629. Accessed February 1, 2019.
30. Wang G, Li W, Ourselin S, Vercauteren T. Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In: Crimi A, Bakas S, Kuijf H, Menze B, Reyes M, eds. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Lecture Notes in Computer Science*. Springer International Publishing; Cham, Switzerland. 2018:178–190.
31. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage.* 2006;31(3):1116–1128.
32. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017;77(21):e104–e107.
33. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition [published online ahead of print December 10, 2015], arXiv, arXiv:151203385 [cs], preprint: not peer reviewed. https://doi.org/10.48550/arXiv.1512.03385. Accessed December 8, 2019.
34. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics; May 13-15; 2010; Chia Laguna Resort, Sardinia, Italy. https://proceedings.mlr.press/v9/glorot10a.html. Accessed March 2, 2022.
35. Kingma DP, Ba J. Adam: a method for stochastic optimization [published online ahead of print January 29, 2017], arXiv, arXiv:14126980 [cs]. https://doi.org/10.48550/arXiv.1412.6980. Accessed October 28, 2020.
36. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* 2002;16(1):321–357.
37. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837–845.
38. Sun X, Xu W. Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Process Lett.* 2014;21(11):1389–1393.
39. Crespo I, Vital AL, Nieto AB, et al. Detailed characterization of alterations of chromosomes 7, 9, and 10 in glioblastomas as assessed by single-nucleotide polymorphism arrays. *J Mol Diagn.* 2011;13(6):634–647.
40. Inda MM, Fan X, Muñoz J, et al. Chromosomal abnormalities in human glioblastomas: gain in chromosome 7p correlating with loss in chromosome 10q. *Mol Carcinog.* 2003;36(1):6–14.
41. Cen L, Carlson BL, Schroeder MA, et al. p16-Cdk4-Rb axis controls sensitivity to a cyclin-dependent kinase inhibitor PD0332991 in glioblastoma xenograft cells. *Neuro Oncol.* 2012;14(7):870–881.
42. Taylor JW, Parikh M, Phillips JJ, et al. Phase-2 trial of palbociclib in adult patients with recurrent RB1-positive glioblastoma. *J Neurooncol.* 2018;140(2):477–483.