Journal of
Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Classification of brain tumor isocitrate dehydrogenase status using MRI and deep learning

Sahil Nalawade
Gowtham K. Murugesan
Maryam Vejdani-Jahromi
Ryan A. Fisicaro
Chandan G. Bangalore Yogananda
Ben Wagner
Bruce Mickey
Elizabeth Maher
Marco C. Pinho
Baowei Fei
Ananth J. Madhuranthakam
Joseph A. Maldjian

Sahil Nalawade, Gowtham K. Murugesan, Maryam Vejdani-Jahromi, Ryan A. Fisicaro, Chandan G. Bangalore Yogananda, Ben Wagner, Bruce Mickey, Elizabeth Maher, Marco C. Pinho, Baowei Fei, Ananth J. Madhuranthakam, Joseph A. Maldjian, "Classification of brain tumor isocitrate dehydrogenase status using MRI and deep learning," *J. Med. Imag.* **6**(4), 046003 (2019), doi: 10.1117/1.JMI.6.4.046003.

Journal of Medical Imaging 6(4), 046003 (Oct–Dec 2019)

Classification of brain tumor isocitrate dehydrogenase status using MRI and deep learning

Sahil Nalawade,[a,†] Gowtham K. Murugesan,[a,†] Maryam Vejdani-Jahromi,[a] Ryan A. Fisicaro,[a] Chandan G. Bangalore Yogananda,[a] Ben Wagner,[a] Bruce Mickey,[b] Elizabeth Maher,[c] Marco C. Pinho,[a] Baowei Fei,[a,d] Ananth J. Madhuranthakam,[a] and Joseph A. Maldjian[a,*]
[a]UT Southwestern Medical Center, Department of Radiology, Dallas, Texas, United States
[b]UT Southwestern Medical Center, Department of Neurological Surgery, Dallas, Texas, United States
[c]UT Southwestern Medical Center, Department of Neurology and Neurotherapeutics, Dallas, Texas, United States
[d]UT Dallas, Department of Bioengineering, Richardson, Texas, United States

Abstract. Isocitrate dehydrogenase (IDH) mutation status is an important marker in glioma diagnosis and therapy. We propose an automated pipeline for noninvasively predicting IDH status using deep learning and T2-weighted (T2w) magnetic resonance (MR) images with minimal preprocessing (N4 bias correction and normalization to zero mean and unit variance). T2w MR images and genomic data were obtained from The Cancer Imaging Archive dataset for 260 subjects (120 high-grade and 140 low-grade gliomas). A fully automated two-dimensional densely connected model was trained to classify IDH mutation status on 208 subjects and tested on another held-out set of 52 subjects using fivefold cross validation. Data leakage was avoided by ensuring subject separation during the slice-wise randomization. Mean classification accuracy of 90.5% was achieved for each axial slice in predicting the three classes of no tumor, IDH mutated, and IDH wild type. Test accuracy of 83.8% was achieved in predicting IDH mutation status for individual subjects on the test dataset of 52 subjects. We demonstrate a deep learning method to predict IDH mutation status using T2w MRI alone. Radiologic imaging studies using deep learning methods must address data leakage (subject duplication) in the randomization process to avoid upward bias in the reported classification accuracy. [DOI: 10.1117/1.JMI.6.4.046003]

Keywords: isocitrate dehydrogenase; magnetic resonance imaging; convolutional neural network; deep learning; tumor classification; segmentation.

Paper 19149R received Jun. 28, 2019; accepted for publication Nov. 18, 2019; published online Dec. 10, 2019.

1 Introduction

In 2008, it was reported that some glioblastoma multiformes (GBM) harbor a mutation in a gene coding for the citric acid cycle enzyme isocitrate dehydrogenase (IDH).[1] Subsequent studies revealed that the majority of low-grade gliomas (LGGs) possess a mutant form of IDH, and that the mutant enzyme catalyzes the production of the oncometabolite 2-hydroxyglutarate (2-HG).[2] Although this product of the mutant form of IDH is believed to play a role in the initiation of the neoplastic process, it has been observed that gliomas that contain the mutant enzyme have a better prognosis than tumors of the same grade that contain only the wild type IDH. This observation implies that IDH mutated and IDH wild type gliomas are biologically different tumors[2] and led the World Health Organization to designate them as such in the latest revision of their classification of gliomas.[3] Although a presumptive diagnosis of an IDH mutated glioma may be made on the basis of magnetic resonance (MR) spectroscopy for 2-HG,[4–7] at the present time, the only way to definitively identify an IDH mutated glioma is to perform immunohistochemistry or gene sequencing on a tissue specimen acquired through biopsy or surgery. Because the differences between IDH mutated and IDH wild type gliomas may have implications for their treatment, especially if inhibitors of the mutant IDH enzyme currently in development prove to halt their growth, there is interest in attempting to distinguish between these two tumor types prior to surgery. As noted above, one avenue of research involves using MR spectroscopy to measure levels of 2-HG in the tumor.[5,8–10] More recent studies have attempted to utilize machine learning techniques to analyze diagnostic MR images and predict IDH mutation status in gliomas using anatomic differences between the two tumor types.

Delfanti et al.[11] demonstrated that genomic information with fluid-attenuated inversion recovery (FLAIR) MR imaging could be used for the classification of patient images into IDH wild type, and IDH mutation with and without 1p/19q codeletion. The main determinants for classification were tumor border and location, with IDH mutant tumors having well-defined or slightly ill-defined borders and predominantly a frontal localization and IDH wild type tumors demonstrating undefined borders and location in nonfrontal areas. Chang et al.[12] developed a deep learning residual network model for predicting IDH mutation with preprocessing steps, including resampling, coregistration of multiple MR sequences, bias correction, normalization, and tumor segmentation. Using a combination of imaging and age, the model demonstrated a testing accuracy of 89.1% and an area under the curve (AUC) value of 0.95 for IDH mutations for all image sequences combined. Zhang et al.[13] used 103 LGG subjects for training a support vector machine for classifying IDH mutation status, achieving an AUC of 0.83 on testing data. In another approach, Chang et al.[14] similarly demonstrated that IDH mutation status can be determined using

[*] Address all correspondence to Joseph A. Maldjian, E-mail: Joseph.Maldjian@UTSouthwestern.edu

[†] These authors contributed equally to this work

2329-4302/2019/$28.00 © 2019 SPIE

T2-weighted (T2w), T2w-FLAIR, and T1-weighted (T1w) pre- and postcontrast images. Preprocessing steps in their work included co-registration of all sequences, intensity normalization using zero mean and unit variance, application of a three-dimensional (3-D) convolutional neural network (CNN)-based whole tumor segmentation tool for segmenting the lesion margins, cropping the output tumor mask on all input imaging sequences, and resizing individual image slices to 32 × 32 with four input sequence channels. The mean accuracy result from the model was 94% with a fivefold cross-validation accuracy ranging from 90% to 96%.[14] Common to all of these previous methods is the involvement of preprocessing steps, typically including some form of brain tumor presegmentation or region of interest (ROI) extraction, and utilizing multiparametric or 3-D near-isotropic MRI data that are often not part of the standard clinical imaging protocol.[12,14]

In this work, we propose a fully automated deep learning-based pipeline using a densely connected network model, which involves minimal preprocessing and requires only standard T2w images. A similar approach has been previously used for the identification of the O^6-methylguanine-deoxyribose nucleic acid (DNA) methyltransferase methylation status and prediction of 1p/19q chromosomal arm deletion.[15] Clinical T2w images are acquired in a short time frame (typically around 2 min) and are robust to motion with current acquisition methods. Almost universally, high-quality T2w images are acquired during clinical brain tumor workups. The preprocessing steps preserve the original image information without the need for any resampling, skull stripping, ROI, or tumor presegmentation procedures. The advantage of a dense network model is that it passes the weights from all the previous blocks to the subsequent blocks, preserving the information from the initial layer and aiding in the classification.

The ability to quickly and accurately classify IDH status noninvasively can help with better planning, counseling, and treatment for brain tumor patients, especially in cases where biopsy is not feasible due to unfavorable tumor locations. A methodological contribution that we specifically make to the radiologic deep learning literature is on further clarifying the approach to data randomization for two-dimensional (2-D) models. In slice-wise deep learning models, simple randomization of all the slices across training, validation, and testing sets can lead to subject duplication where different slices from the same subject are seen by the algorithm across groups. Adjacent slices of the same tumor can carry significantly similar information, biasing the measured slice-wise performance. Slice randomization should be done on a subject-wise basis to avoid this pitfall. The deep learning approach presented is fully automated and can be easily implemented in the clinical workflow using only T2w MR images. We also compared IDH image-based classification performance across several widely used deep learning models.

2 Materials and Methods

2.1 *Subjects*

Two hundred and sixty subjects from The Cancer Imaging Archive (TCIA)[16] dataset were selected, including 120 high-grade gliomas (HGGs)[17] and 140 LGGs,[18] and based on their preoperative status from a pool of 461 subjects. The genomic information was provided through the U.S. National Cancer Institute's Genomic Data Commons data portal.[19] The genomic data were available in the following three classes: IDH mutated, IDH wild type, and not available (N/A). The genomic data of the N/A type were excluded from the pool of 461 subjects. MRI data were filtered for any visible artifacts in the images. The final dataset consisted of 260 subjects based on the available genomic information, MRI data, preoperative status, and lack of image artifacts on the T2w images. Out of the 461 subjects in the TCIA, 292 were preoperative. Out of these, 22 subjects did not have T2w images. Of the remaining 270 subjects, 10 had obvious motion artifacts, leaving 260 subjects in the final dataset.

A standard 80:20 data split was employed with 80% training and 20% testing (held-out). The 80% training was further split into a standard 80:20 split of 80% training and 20% validation. The final dataset of 260 subjects was thus randomly divided into a training set (208 subjects, including ∼96 HGGs and 112 LGGs) and a test set (52 subjects, including ∼24 HGGs and 28 LGGs). This process was repeated separately for each fold during the fivefold cross validation.

For each fold of the cross validation, 208 subjects with, on average, 9728 axial slices of T2w images were selected for training and validation (7177 slices: no tumor, 1110 slices: IDH mutated, and 1441 slices: IDH wild type). The start and end slices of the tumor (edge slices) were manually labeled for each T2 dataset and verified by a neuroradiologist. These edge slices were excluded from training to provide more robust ground truth data representative of the tumor, rather than partial volume data from edge slices. All slices were included for the testing set. Each T2w slice was manually assigned only one label (no tumor, IDH mutated, or IDH wild type). To address any class imbalance due to the higher number of no tumor slices, class weights were assigned based on the labels in the training dataset. Although this was a slice-wise training model, slices of subjects in the testing set were not mixed into the training set. This is a critical step related to the data leakage problem in 2-D networks, especially for radiologic deep learning studies.[20,21] This was necessary to avoid bias during testing and an overinflation of the measured accuracies. Fifty-two subjects with 2522 axial slices (1839 slices: no tumor, 299 slices: IDH mutated, and 384 slices: IDH wild type) were not included in the training or validation and were used for testing for each fold. Classification was done on a slice-wise basis (2-D) followed by majority voting across all slices to provide a patient-level classification. Note that we use the term slice-wise to refer to classification of each 2-D axial image for IDH status. Similarly, the term subject-wise is used for classification of IDH status for each subject. We used a straightforward majority voting scheme to determine subject-wise classification based on the majority IDH classification of the individual 2-D slices. Subjects classified with an equal number of IDH mutated and IDH wild type tumor slices were assigned to the IDH wild type group.

2.2 *Image Processing*

Minimal standard preprocessing of the T2w images from the TCIA dataset was performed prior to training (Fig. 1). The images were converted from DICOM to NifTI format using dcm2nii,[22] bias corrected to remove radio frequency inhomogeneity using the N4 bias correction algorithm, zero-mean intensity normalized to between −1 and 1, and resampled to 128 × 128 image dimensions to improve the computational efficiency during training. The Inception V4 model, however, required an input image size of 299 × 299 as a design constraint

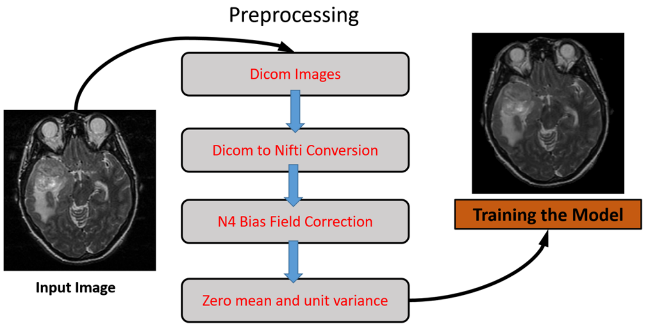

Fig. 1 Flowchart of preprocessing steps prior to training the deep learning model.

of this model when originally constructed.[23,24] The total preprocessing time for each subject was <1 min.

2.3 *Model Training*

The following models were used for classification of the T2w images into IDH mutated and IDH wild type classes: residual network (ResNet-50), densely connected network (DenseNet-161), and Inception-v4. Our choice of network architecture was based on the best performers from the ImageNet challenge for 2015 (ResNET) and 2017 (DenseNet and Inception V4). The DenseNet model, designed by Huang et al.,[25] received the best paper award at the Conference on Computer Vision and Pattern Recognition 2017. The models were trained with the Pycharm and Python IDEs using the Keras python package with TensorFlow backend engines. Fine tuning of the three classes was performed on all models. The three-class labels for each slice were no tumor, IDH mutated, and IDH wild type. The models were originally trained on ImageNet data with three channels (RGB). For our implementation, the three-channel input was provided as a central slice with the two immediate surrounding slices. If the central slice was the first or last slice, the surrounding slices were assigned as no value.

2.4 *ResNet-50 Model*

The residual network was implemented as proposed by He et al.[26] Each residual connection adds the input of the block to the output, helping to preserve information from the previous block. A deep residual network framework was added to the model while maintaining parameter numbers to address issues with convergence in the originally proposed model. The residual net used the kernel initializer as "He normal" for weight initialization. On top of the residual network model, a flattened output was added and sent to the dense layer with the rectified linear unit (relu) activation and a dropout of 0.5. The final layer of the model was the classification layer with a softmax activation and the number of classes as the output. The residual network model used for training was ResNet-50 (Fig. 2).

2.5 *Inception-v4 Model*

The Inception model architecture was designed by the Google research team.[23,24] The Inception-v4 model is a deep architecture with 41 million parameters and the model is designed with inception blocks and reduction blocks. The inception blocks are used in a sequential manner with reduction blocks except for the last inception block, which has an average pooling layer and a dropout layer before the classification layer.

2.6 *DenseNet-161 Model*

The DenseNet model was based on the design by Huang et al.[25] This model was inspired by the residual network model, which allows the residual connections to pass information from the

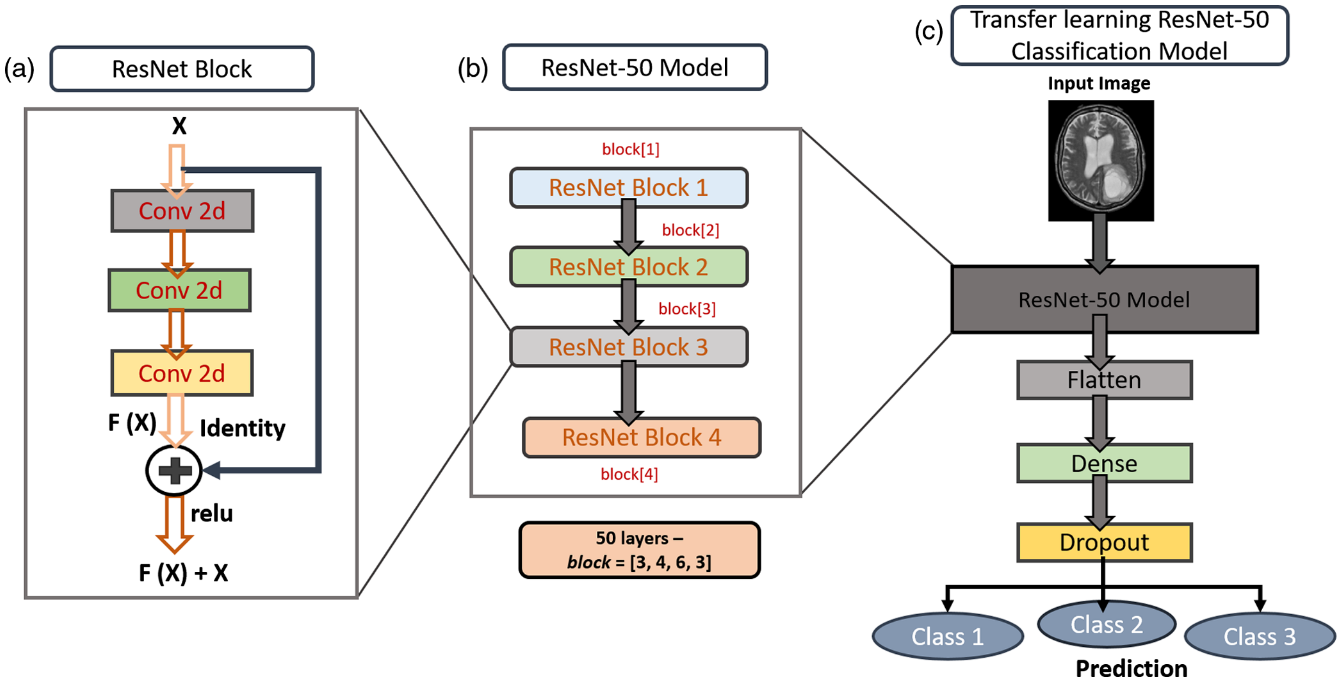

Fig. 2 Architecture of ResNet-50 (50 layers) model used for IDH classification.

previous layer to the subsequent layer. Dense networks have advantages over other networks by alleviating the vanishing gradient problem with feature propagation through the dense connection to the subsequent layers.

The features passed to the subsequent layers in the DenseNet model are not added by summation but are combined using concatenation. Each block has connections from the previous block such that $L =$ number of blocks and the number of connections for each block is $L \times (L + 1)/2$, creating a dense connectivity pattern or DenseNet. The DenseNet-161 model architecture is shown in Fig. 3, which illustrates a five-block approach where the first block is the input layer and each of the subsequent four blocks are characterized by 2-D convolution layers with filter sizes of (1×1) and (3×3), respectively. The pretrained model was used to transfer learning and was used for classification based on the trained information. A 161-layer DenseNet model was used for model training.

2.7 *Training, Testing, and Statistical Analysis*

Model training was performed on a Nvidia Tesla P100, P40, K40/K80 GPU with 384 GB RAM and the model accuracy was assessed for 200 epochs. The optimizer used for training was the stochastic gradient descent[27] as described by Zhang[28] and the learning rate was set to 10^{-7}, with a decay of 10^{-7} and momentum of 0.8. Data augmentation was performed on the training dataset, which included vertical and horizontal flip, random rotation, translation, shear, zoom shifts, and elastic transformation to minimize overfitting the data. The results were analyzed by assessing accuracy, precision, sensitivity, specificity, and F1 score values. Slice-wise model testing was performed based on the output from the 2-D model. Subject-wise classification was performed based on majority voting across IDH mutated and IDH wild type tumor slices. This classification accuracy was computed on the independent test dataset that was separate from the testing and validation datasets. Comparison between the models across fivefold cross-validation accuracies was performed using the Mann–Whitney rank sum test.

2.8 *Model Training Times*

The DensetNet-161 model took ∼110 h for training, while the ResNet-50 model and the Inception V4 model took ∼56 and ∼32 h, respectively. Testing time for individual subject classification was <40 s for all models.

3 Results

3.1 *Training, Validation, and Testing Accuracy*

Table 1 shows the accuracy comparison between the ResNet-50, DenseNet-161, and Inception-v4 models. The DenseNet-161 model outperformed the Inception-v4 model and performed slightly better than the ResNet-50 model. Averaged across the fivefolds, the slice-wise accuracy of the DenseNet-161 model was $90.5 \pm 1.0\%$ with an AUC of 0.95 on the held-out test dataset of 52 subjects. The mean slice-wise accuracies of the Resnet-50 model and the Inception-v4 model were $89.7 \pm 1.1\%$ with an AUC of 0.95 on the held-out test dataset and $76.1 \pm 3.7\%$ with an AUC of 0.86, respectively. Testing times across slices for each subject using the DenseNet-161, ResNet-50, and Inception V4 models were ∼40, ∼15, and ∼30 s, respectively. Subject-wise determination of IDH mutation status by majority voting across classified slices took <1 s.

3.2 *Accuracy, Precision, Recall/Sensitivity, Specificity, F1 Score, and AUC Comparison*

Average metrics were computed across folds and classes. The classification accuracy, precision, recall/sensitivity, specificity, F1 score, and AUC for slice-wise IDH classification with the DenseNet-161 model were $90.5 \pm 1.0\%$, $79.9 \pm 3.4\%$, $83.1 \pm 3.2\%$, $94.8 \pm 0.5\%$, $81.3 \pm 3.2\%$, and 0.95, respectively. The classification accuracy, precision, recall/sensitivity, specificity, F1 score, and AUC for slice-wise IDH classification with the ResNet-50 model were $89.7 \pm 1.1\%$, $79.3 \pm 3.3\%$, $81.7 \pm 3.2\%$, $94.1 \pm 0.8\%$, $80.2 \pm 3.1\%$, and 0.95 and for the Inception-v4 model were $76.1 \pm 3.7\%$, $59.4 \pm 2.7\%$, $59.2 \pm 2.6\%$, $84.5 \pm 3.1\%$, $58.2 \pm 2.1\%$, and 0.86, respectively.

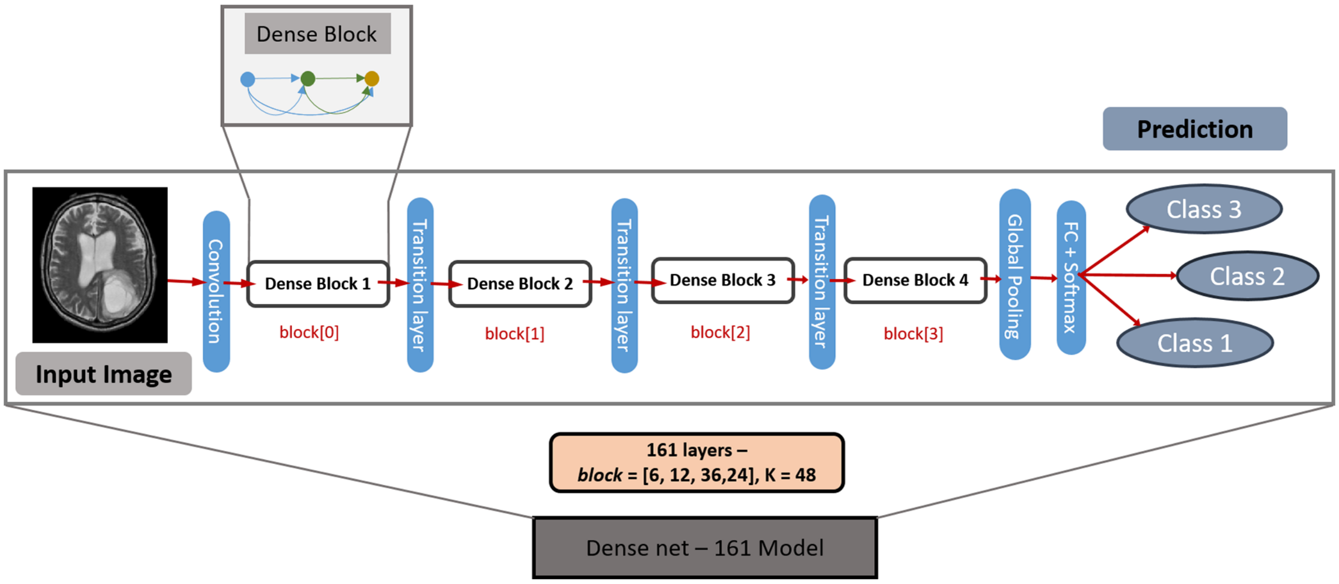

Fig. 3 Architecture of the DenseNet-161 (161 layers) model used for IDH classification.

Table 1 Slice-wise accuracy comparisons between the ResNet-50 Model, Inception-v4, and DenseNet-161 model averaged for fivefold cross validation.

Results averaged for fivefold cross validation			
Model	Training accuracy (%)	Validation accuracy (%)	Testing accuracy (%)
Inception-v4	64.8 ± 7.4	72.2 ± 6.9	1916/2522 (76.1 ± 3.7)
ResNet-50	97.9 ± 0.5	96.5 ± 0.6	2265/2522 (89.7 ± 1.1)
DenseNet-161	**97.9** ± **0.4**	**96.4** ± **0.6**	**2282/2522** (**90.5** ± **1.0**)

Note: The results from the best performing model "DenseNet-161" are highlighted in bold font.

Table 3 Slice-wise and subject-wise comparison of accuracy, precision, recall, F1 score, and AUC parameters for each of the fivefold cross validations for the DenseNet-161 model.

DenseNet-161 model						
Fold	Accuracy (%)	Precision (%)	Recall/ sensitivity (%)	Specificity (%)	F1 score (%)	AUC
			Slice-wise			
1	91.7	83.3	86.0	94.9	84.4	0.95
2	91.0	82.7	84.4	95.1	83.5	0.95
3	90.1	79.3	81.5	94.5	80.2	0.95
4	88.7	73.7	77.6	93.9	75.5	0.91
5	90.9	80.3	86.0	95.4	82.8	0.95
			Subject-wise			
1	84.6	84.8	84.0	84.0	84.2	0.84
2	86.5	87.2	87.5	87.5	86.5	0.87
3	78.8	79.1	77.9	77.9	78.2	0.78
4	82.7	83.1	81.8	81.8	82.2	0.82
5	86.5	86.3	86.6	86.6	86.4	0.87

For subject-wise IDH classification, accuracy, precision/positive predictive value, recall/sensitivity, specificity, F1 score, and AUC with the DenseNet-161 model were 83.8 ± 2.9%, 84.1 ± 2.9%, 83.5 ± 3.5%, 83.5 ± 7.3%, 83.5 ± 3.1%, and 0.84, respectively (Table 2). Subject-wise IDH classification, accuracy, precision/positive predictive value, recall/sensitivity, specificity, F1 score, and AUC with the ResNet-50 model were 81.4 ± 7.3%, 81.5 ± 7.2%, 81.5 ± 7.1%, 81.5 ± 7.1%, 81.4 ± 7.3%, and 0.81 ± 0.1% and for the Inception-v4 model were 64.2 ± 6.5%, 65.8 ± 8.2%, 65.1 ± 7.3%, 65.1 ± 3.5%, 64.0 ± 6.5%, and 0.65 ± 0.1%, respectively. Slice-wise and subject-wise comparisons of accuracy, precision, recall/sensitivity, specificity, F1 score, and AUC for each of the fivefold cross validations for the DenseNet-161 model are shown in Table 3. The DenseNet-161 model performed significantly better than the Inception-v4 model for slice-wise classification ($p = 0.008$) and for subject-wise classification ($p = 0.008$) using the rank sum test. There was not a statistically significant difference in performance between the DenseNet-161 and ResNet-50 models on either slice-wise performance ($p = 0.341$) or subject-wise performance ($p = 0.913$). However, the DenseNet-161 model provided higher mean cross-validation accuracy and less variability between the folds of the cross-validation procedure than the ResNet-50 model for subject-wise classification.

3.3 *Slice-Wise Comparison*

The precision for the DenseNet-161 model across fivefold cross validation was 97.7 ± 0.5% for the no tumor classification, 71.7 ± 6.8% for IDH mutation, and 70.3 ± 5.5% for IDH wild type. The precision for the ResNet-50 model across fivefold

Table 2 Slice-wise and subject-wise comparison of accuracy, precision, recall, F1 score, and AUC parameters averaged for fivefold cross validation for ResNet-50, Inception-v4, and DenseNet-161.

Results averaged for fivefold cross validation						
Parameters	Accuracy (%)	Precision (%)	Recall/sensitivity (%)	Specificity (%)	F1 score (%)	AUC
			Slice-wise			
Inception-v4	76.1 ± 3.7	59.4 ± 2.7	59.2 ± 2.6	84.5 ± 3.1	58.2 ± 2.1	0.86 ± 0.0
ResNet-50	89.7 ± 1.1	79.3 ± 3.3	81.7 ± 3.2	94.1 ± 0.8	80.2 ± 3.1	0.95 ± 0.0
DenseNet-161	**90.5** ± **1.0**	**79.9** ± **3.4**	**83.1** ± **3.2**	**94.8** ± **0.5**	**81.3** ± **3.2**	**0.95** ± **0.0**
			Subject-wise			
Inception-v4	64.2 ± 6.5	65.8 ± 8.2	65.1 ± 7.3	65.1 ± 3.5	64 ± 6.5	0.65 ± 0.1
ResNet-50	81.4 ± 7.3	81.5 ± 7.2	81.5 ± 7.1	81.5 ± 7.1	81.4 ± 7.3	0.81 ± 0.1
DenseNet-161	**83.8** ± **2.9**	**84.1** ± **2.9**	**83.5** ± **3.5**	**83.5** ± **7.3**	**83.5** ± **3.1**	**0.84** ± **0.0**

Note: The results from the best performing model "DenseNet-161" are highlighted in bold font.

cross validation was 97.0 ± 1.1% for the no tumor class, 72.8 ± 3.9% for IDH mutation, and 68.0 ± 7.5% for IDH wild type, and for the Inception-v4 model was 88.1 ± 6.3% for the no tumor class, 53.9 ± 7.8% for IDH mutation, and 36.3 ± 7.0% for IDH wild type.

3.4 *Data Leakage*

To demonstrate the effect of data leakage, we also performed training, validation, and testing using the DenseNet-161 model and T2w images with simple randomization of the slices without separating by subject. Mean slice-wise accuracies of 96.7% were achieved across the fivefold cross validation, which was 6.2% higher than when correctly separating slices by subject.

4 Discussion

The results from Tables 1 and 2 show that the ResNet-50 model performed better than the Inception-v4 model. The ResNet-50 architecture has residual connections that preserve information from the previous layer in the residual block. The DenseNet-161 model performed the best of all the three models tested. Unlike the ResNet-50 model, the DenseNet-161 model architecture carries the information from all previous layers and adds the information to the next layer. This helped in learning the information from different layers and transferring to the next layers. The slice-wise classification AUC results were 0.95 for DenseNet-161, 0.95 for ResNet-50, and 0.86 for Inception-v4. The DenseNet-161 model performed significantly better than the Inception-v4 model ($p = 0.008$). Although there was not a statistically significant difference in performance between the DenseNet-161 and ResNet-50 models ($p = 0.341$), the DenseNet-161 model provided higher mean cross-fold validation accuracy and less variability between folds for subject-wise classifications.

Chang et al.[14] demonstrated a high classification accuracy for IDH mutation status using T2w, FLAIR, T1w pre- and postcontrast images. Preprocessing steps included co-registration across multiple sequences, intensity normalization to zero mean and unit variance, segmentation of the brain tumor, cropping the images, and resizing slices to 32×32. A 94% mean accuracy on fivefold cross validation was reported. The approach to classification was slice-wise, similar to our model. In designing the slice-wise classification model, it is important to ensure that none of the slices of subjects from the testing set are inadvertently included in the training set. This can easily be overlooked in 2-D slice-wise models during the slice randomization process that generate the training slices, validation slices, and testing slices. This can introduce bias in the testing phase, artificially boosting accuracies by including slices from subjects in the training set that share considerable information with different slices but from the same subjects in the testing set. It is not clear in the previously reported 2-D models whether this caveat was adhered to.

An important methodological contribution that we make specifically to the radiologic deep learning literature is on the approach to data randomization for 2-D models. It is critical that imaging researchers are aware of the data leakage and subject duplication issue. This is perhaps unique to radiology where multiple slices of pathology are acquired in MRI or computed tomography, with considerable overlap in feature content from slice to slice. Widely used deep learning tools provide the ability to perform data randomization using a simple flag in the so-called routine (e.g., in Keras or Scikit-learn[29]). Use of this flag in 2-D imaging-based CNNs can lead to bias in the results by inadvertently including slices from the same subject in both training and testing cohorts. This is a significant concern, as it can lead to data leakage in which examples of the same subject (albeit different slices of the same tumor) can appear in the training set and the test set. The problem of data leakage in medical images was discussed by Wegmayr et. al.[30] and Feng et al.[20] and has been referred to as subject duplication in training and testing sets. When not accounting for data leakage, we were able to achieve slice-wise accuracies of 96.7% with the T2 images alone, slightly higher than that of Chang et al.[14]. When appropriately accounting for the data leakage issue, our slice-wise accuracies were reduced to 90.5% as reported here. One of the contributions of our work is in making the radiology community aware of the data leakage problem, as it is very easy to overlook when 2-D networks that use image slices as input are considered. We also demonstrate that T2w images can provide significant information for IDH image-based classification.

The majority of HGG tumors are IDH wild type (up to 90%). An algorithm that merely distinguishes between HGG and LGG for determination of IDH status is likely of limited value as this can be done subjectively with fairly high accuracy on the basis of contrast enhancement. For example, previous studies that used multiparametric MR data for determination of IDH status in HGG and LGG may have demonstrated high accuracy predominantly on the basis of contrast-enhancement features. The more valuable distinction from a clinical standpoint would be between IDH mutated and IDH wild type LGGs, in which contrast enhancement is usually absent. Our training and testing samples were weighted toward LGG, and there was a significant number of IDH wild type LGGs in both the training and validation samples (∼30%). Our testing accuracy for the LGG group was 78.6%. Additionally, our use of T2w-only images eliminates the potential for the algorithm being a contrast-enhancement discriminator.

Our method provides high accuracy with minimal preprocessing steps as compared to previous work. The preprocessing steps in our work only involve N4 bias field correction and intensity normalization. Our method also involves no tumor segmentation or ROI extraction as described by Chang et al.,[12] which helps in reducing the time, effort, and potential sources of error. Our method also does not require pre-engineered features to be extracted from the images or histopathological data as described by Delfanti et al.[11]. This general approach can be easily incorporated into an automated clinical workflow for IDH classification. The minimal preprocessing and the use of standard T2w images alone make it promising as a robust clinical tool for noninvasively determining IDH mutation status.

5 Limitations

This is a retrospective study applying several neural network architectures to the TCIA HGG-LGG database to generate a model predicting IDH genotype based only on T2w MR imaging. The dataset, especially at the subject level, is small in terms of deep learning applications and may not generalize well. Fluctuation of performance is also a concern with small datasets. However, the TCIA dataset is the largest curated brain tumor dataset publicly available, and it uses data from multiple sites using different imaging protocols. This database consisted of data from 10 different institutions, out of which 8 institutions contributed GBM/HGG datasets and 5 institutions contributed LGG datasets to the TCIA cohort. This provided a very heterogeneous dataset, and we believe this is perhaps even better than using data from a single source for deep learning applications. While our current study focused on the classification of T2w images into no tumor, IDH mutated, and IDH wild type, future studies can extend this approach to classify IDH1 and IDH2 subtypes. Accuracies may be further improved with the inclusion of multiparametric imaging data in the training model. Our approach, however, is much more straightforward using T2w images alone without the requirement of additional imaging sequences. Clinically, T2w images are typically acquired within 2 min and are robust to patient motion. The multisequence input required by previous approaches can be compromised due to patient motion from lengthier examination times and the need for gadolinium contrast, especially as the postcontrast images are typically acquired at the end of an already lengthy examination time. For a potential clinical solution, the use of T2w images is a significant strength, as these images are almost uniformly acquired without artifacts from patient motion.

6 Conclusion

We demonstrate a deep learning method to predict IDH mutation status using T2w MR images alone. The proposed model requires minimal preprocessing to obtain high accuracies without the need for tumor segmentation or extraction of ROI, making it promising for robust clinical implementation.

7 Appendix

Additional information on the models used is included here. Figure 4 provides an example of the three classes used in the models (IDH mutated tumor, IDH wild type tumor, and the no tumor class). Figure 5 provides an example of the Inception-v4 model architecture. Figure 6 provides receiver operating characteristic (ROC) curves for subject-wise classification of the three models. The DenseNet-161 model provided the highest mean AUC across the fivefold validation (0.84). Table 4 compares the accuracy, precision, and sensitivity for slice-wise classification using the DenseNet-161 model with and without data leakage. When not accounting for data leakage (e.g., random assignment of slices to training, validation, and testing), accuracy was ∼6% higher than when the data leakage was appropriately handled (separating slices by subject). Table 5 provides the class weights used for each fold in the IDH classification. Table 6 provides a list of the 260 subjects used from the TCIA database.

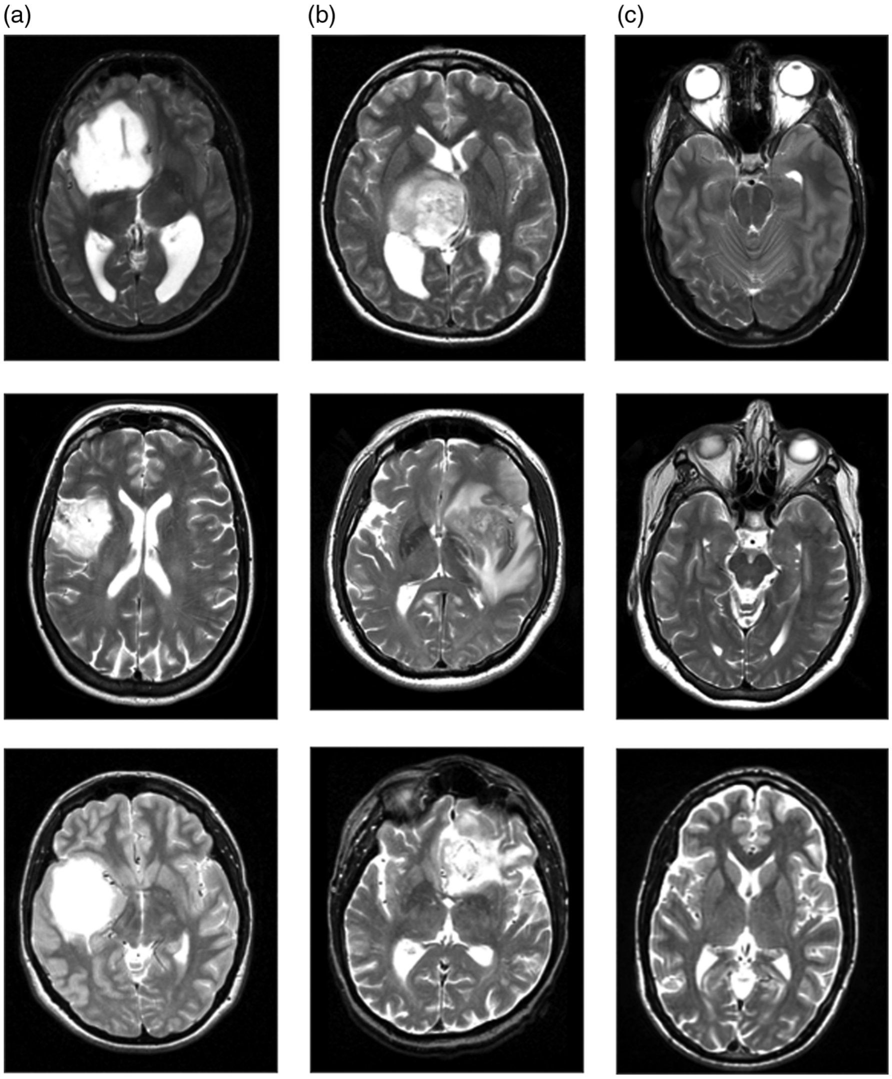

Fig. 4 Examples of the three classes used for IDH classification: (a) IDH mutated tumor, (b) IDH wild type tumor, and (c) no tumor.

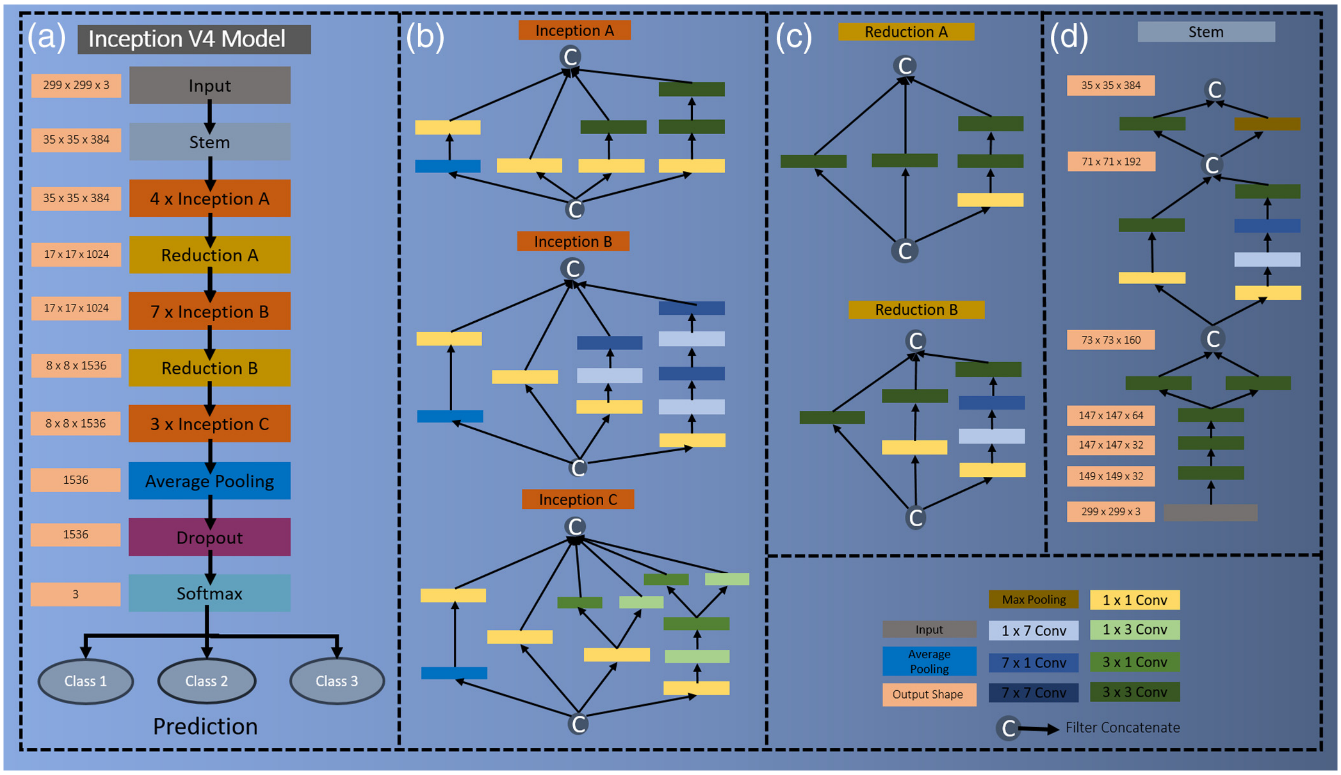

Fig. 5 Inception-v4 model architecture. This model uses (b) three different inception blocks, (c) two different reduction blocks, and (d) one stem block with convolution layers at different resolutions (color legend on the lower right). (a) The final prediction layer is modified to provide three class outputs.

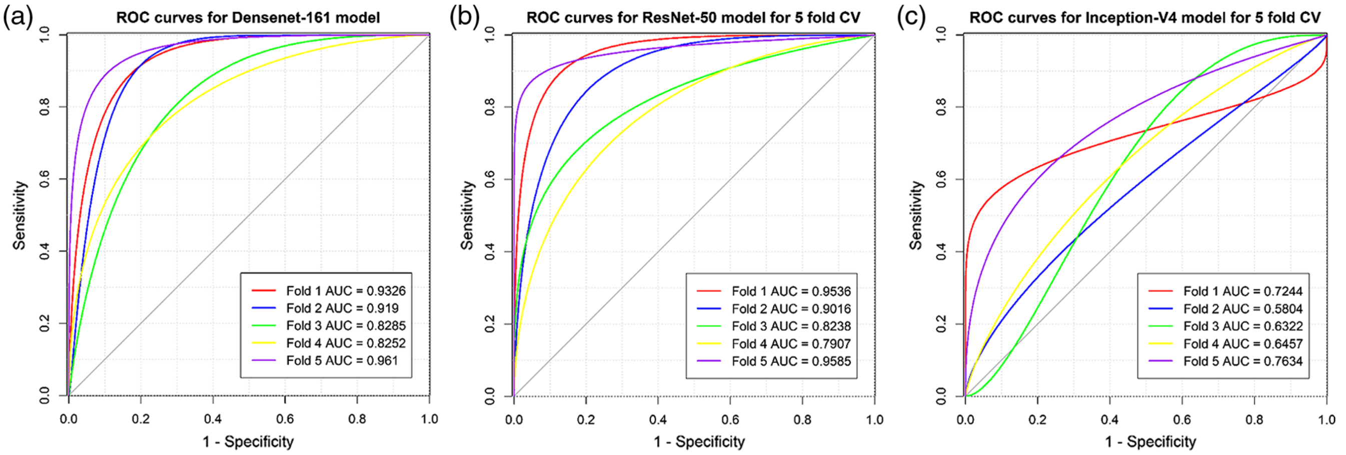

Fig. 6 Subject-wise classification ROC curves for the three models. (a) DenseNet-161, (b) ResNet-50, and (c) Inception-v4. DenseNet-161 provided the highest mean AUC across the fivefold cross validation (0.84 ± 0.0) compared to ResNet-50 (0.81 ± 0.1) and Inception-v4 (0.65 ± 0.1).

Table 4 DenseNet-161 model performance with no data leakage compared to with data leakage.

	DenseNet-161 model with and without data leakage					
	No data leakage			Data leakage		
Fold	Accuracy (%)	Precision (%)	Recall/sensitivity (%)	Accuracy (%)	Precision (%)	Recall/sensitivity (%)
			Slice-wise			
1	91.7	83.3	86.0	95.6	91.3	94.31

Table 4 (*Continued*).

	DenseNet-161 model with and without data leakage					
	No data leakage			Data leakage		
Fold	Accuracy (%)	Precision (%)	Recall/sensitivity (%)	Accuracy (%)	Precision (%)	Recall/sensitivity (%)
2	91.0	82.7	84.4	96.7	93.6	94.9
3	90.1	79.3	81.5	96.1	91.6	94.40
4	88.7	73.7	77.6	97.2	94.5	95.13
5	90.9	80.3	86.0	97.9	94.9	96.70
Average	90.5 ± 1.0	79.9 ± 3.4	83.1 ± 3.2	96.7 ± 0.8	93.2 ± 1.5	95.1 ± 0.9

Table 5 Class weights used for each fold in IDH classification.

	Class weights		
Fold	No tumor	IDH mutated	IDH wild type
1	0.32	0.36	0.32
2	0.32	0.36	0.32
3	0.33	0.34	0.33
4	0.33	0.35	0.32
5	0.32	0.35	0.33

Table 6 List of TCIA subjects used for IDH classification.

Subject no.	Subject ID	IDH status
1	TCGA-HT-7684	IDH mutated
2	TCGA-HT-7468	IDH mutated
3	TCGA-HT-8563	IDH mutated
4	TCGA-DU-5871	IDH mutated
5	TCGA-HT-7471	IDH mutated
6	TCGA-HT-A5RB	IDH mutated
7	TCGA-HT-8105	IDH mutated
8	TCGA-DU-7015	IDH mutated
9	TCGA-HT-A616	IDH mutated
10	TCGA-HT-7604	IDH mutated
11	TCGA-HT-7473	IDH mutated
12	TCGA-DU-A6S7	IDH mutated
13	TCGA-DU-6397	IDH mutated
14	TCGA-DU-A5TP	IDH mutated
15	TCGA-CS-6667	IDH mutated
16	TCGA-HT-8111	IDH mutated
17	TCGA-HT-7481	IDH mutated
18	TCGA-HT-7879	IDH mutated
19	TCGA-HT-A615	IDH mutated
20	TCGA-CS-4942	IDH mutated
21	TCGA-HT-7605	IDH mutated
22	TCGA-DU-7010	IDH mutated
23	TCGA-HT-A61A	IDH mutated
24	TCGA-27-1830	IDH wild type
25	TCGA-08-0353	IDH wild type
26	TCGA-CS-4941	IDH wild type
27	TCGA-02-0068	IDH wild type
28	TCGA-FG-6692	IDH wild type
29	TCGA-76-4926	IDH wild type
30	TCGA-02-0009	IDH wild type
31	TCGA-CS-6186	IDH wild type
32	TCGA-27-1838	IDH wild type
33	TCGA-12-1602	IDH wild type
34	TCGA-76-6662	IDH wild type
35	TCGA-02-0048	IDH wild type
36	TCGA-06-0138	IDH wild type
37	TCGA-08-0351	IDH wild type
38	TCGA-06-0216	IDH wild type
39	TCGA-14-0789	IDH wild type
40	TCGA-76-6280	IDH wild type

Table 6 (*Continued*).

Subject no.	Subject ID	IDH status
41	TCGA-08-0352	IDH wild type
42	TCGA-19-5953	IDH wild type
43	TCGA-19-1390	IDH wild type
44	TCGA-19-2631	IDH wild type
45	TCGA-02-0047	IDH wild type
46	TCGA-08-0360	IDH wild type
47	TCGA-06-0157	IDH wild type
48	TCGA-DU-A5TY	IDH wild type
49	TCGA-06-5412	IDH wild type
50	TCGA-06-0160	IDH wild type
51	TCGA-CS-6669	IDH wild type
52	TCGA-12-3650	IDH wild type
53	TCGA-HT-7692	IDH mutated
54	TCGA-DU-7008	IDH mutated
55	TCGA-DU-A6S3	IDH mutated
56	TCGA-HT-7686	IDH mutated
57	TCGA-DU-5872	IDH mutated
58	TCGA-FG-6691	IDH mutated
59	TCGA-CS-5396	IDH mutated
60	TCGA-HT-7877	IDH mutated
61	TCGA-HT-7603	IDH mutated
62	TCGA-FG-A87N	IDH mutated
63	TCGA-HT-8106	IDH mutated
64	TCGA-HT-7475	IDH mutated
65	TCGA-DU-7300	IDH mutated
66	TCGA-DU-6401	IDH mutated
67	TCGA-HT-7478	IDH mutated
68	TCGA-CS-6666	IDH mutated
69	TCGA-HT-7694	IDH mutated
70	TCGA-FG-7637	IDH mutated
71	TCGA-HT-7690	IDH mutated
72	TCGA-DU-5853	IDH mutated
73	TCGA-DU-A6S8	IDH mutated
74	TCGA-FG-6689	IDH mutated
75	TCGA-CS-5394	IDH mutated
76	TCGA-06-0645	IDH wild type
77	TCGA-06-5413	IDH wild type

Table 6 (*Continued*).

Subject no.	Subject ID	IDH status
78	TCGA-76-6657	IDH wild type
79	TCGA-FG-6688	IDH wild type
80	TCGA-02-0037	IDH wild type
81	TCGA-06-0881	IDH wild type
82	TCGA-76-6286	IDH wild type
83	TCGA-08-0359	IDH wild type
84	TCGA-06-0184	IDH wild type
85	TCGA-06-0646	IDH wild type
86	TCGA-FG-A4MU	IDH wild type
87	TCGA-19-2620	IDH wild type
88	TCGA-76-6282	IDH wild type
89	TCGA-CS-6188	IDH wild type
90	TCGA-06-0210	IDH wild type
91	TCGA-27-1836	IDH wild type
92	TCGA-19-5954	IDH wild type
93	TCGA-19-1791	IDH wild type
94	TCGA-08-0389	IDH wild type
95	TCGA-19-2624	IDH wild type
96	TCGA-06-0648	IDH wild type
97	TCGA-02-0011	IDH wild type
98	TCGA-06-0137	IDH wild type
99	TCGA-06-0644	IDH wild type
100	TCGA-12-1598	IDH wild type
101	TCGA-06-0143	IDH wild type
102	TCGA-HT-7680	IDH wild type
103	TCGA-06-0173	IDH wild type
104	TCGA-76-4929	IDH wild type
105	TCGA-DU-8168	IDH mutated
106	TCGA-DU-7019	IDH mutated
107	TCGA-DU-A5TW	IDH mutated
108	TCGA-HT-7606	IDH mutated
109	TCGA-HT-7608	IDH mutated
110	TCGA-HT-7880	IDH mutated
111	TCGA-FG-6690	IDH mutated
112	TCGA-FG-A6IZ	IDH mutated
113	TCGA-DU-6395	IDH mutated
114	TCGA-DU-5851	IDH mutated

Table 6 (*Continued*).

Subject no.	Subject ID	IDH status
115	TCGA-HT-7693	IDH mutated
116	TCGA-DU-6399	IDH mutated
117	TCGA-DU-8164	IDH mutated
118	TCGA-DU-7309	IDH mutated
119	TCGA-HT-8010	IDH mutated
120	TCGA-HT-7616	IDH mutated
121	TCGA-HT-7855	IDH mutated
122	TCGA-DU-A5TU	IDH mutated
123	TCGA-FG-5964	IDH mutated
124	TCGA-DU-7298	IDH mutated
125	TCGA-HT-A61B	IDH mutated
126	TCGA-HT-7902	IDH mutated
127	TCGA-DU-8163	IDH mutated
128	TCGA-HT-8107	IDH wild type
129	TCGA-02-0102	IDH wild type
130	TCGA-HT-7860	IDH wild type
131	TCGA-HT-7882	IDH wild type
132	TCGA-06-0158	IDH wild type
133	TCGA-19-1388	IDH wild type
134	TCGA-08-0357	IDH wild type
135	TCGA-06-0142	IDH wild type
136	TCGA-02-0006	IDH wild type
137	TCGA-HT-7469	IDH wild type
138	TCGA-76-4925	IDH wild type
139	TCGA-DU-8165	IDH wild type
140	TCGA-06-0165	IDH wild type
141	TCGA-DU-5854	IDH wild type
142	TCGA-76-4935	IDH wild type
143	TCGA-14-1794	IDH wild type
144	TCGA-06-0185	IDH wild type
145	TCGA-02-0069	IDH wild type
146	TCGA-02-0070	IDH wild type
147	TCGA-76-6663	IDH wild type
148	TCGA-19-5960	IDH wild type
149	TCGA-06-0176	IDH wild type
150	TCGA-76-4928	IDH wild type
151	TCGA-HT-8564	IDH wild type
152	TCGA-08-0349	IDH wild type
153	TCGA-06-0174	IDH wild type
154	TCGA-12-0829	IDH wild type
155	TCGA-06-0241	IDH wild type
156	TCGA-HT-A4DS	IDH wild type
157	TCGA-DU-7294	IDH mutated
158	TCGA-HT-8013	IDH mutated
159	TCGA-HT-7677	IDH mutated
160	TCGA-DU-7299	IDH mutated
161	TCGA-DU-7306	IDH mutated
162	TCGA-DU-7301	IDH mutated
163	TCGA-HT-7695	IDH mutated
164	TCGA-06-0128	IDH mutated
165	TCGA-HT-7472	IDH mutated
166	TCGA-HT-8113	IDH mutated
167	TCGA-DU-5849	IDH mutated
168	TCGA-DU-A5TR	IDH mutated
169	TCGA-FG-A4MT	IDH mutated
170	TCGA-HT-8108	IDH mutated
171	TCGA-CS-4943	IDH mutated
172	TCGA-HT-7602	IDH mutated
173	TCGA-DU-6407	IDH mutated
174	TCGA-DU-8166	IDH mutated
175	TCGA-DU-7018	IDH mutated
176	TCGA-FG-8186	IDH mutated
177	TCGA-DU-7304	IDH mutated
178	TCGA-HT-8018	IDH mutated
179	TCGA-DU-A6S2	IDH mutated
180	TCGA-DU-7013	IDH wild type
181	TCGA-08-0244	IDH wild type
182	TCGA-HT-7854	IDH wild type
183	TCGA-HT-8110	IDH wild type
184	TCGA-CS-5397	IDH wild type
185	TCGA-02-0075	IDH wild type
186	TCGA-08-0348	IDH wild type
187	TCGA-02-0060	IDH wild type
188	TCGA-HT-A5RC	IDH wild type

Table 6 (*Continued*).

Subject no.	Subject ID	IDH status
189	TCGA-02-0003	IDH wild type
190	TCGA-76-4927	IDH wild type
191	TCGA-06-0168	IDH wild type
192	TCGA-06-0189	IDH wild type
193	TCGA-08-0354	IDH wild type
194	TCGA-HT-A617	IDH wild type
195	TCGA-19-1789	IDH wild type
196	TCGA-12-1601	IDH wild type
197	TCGA-76-4934	IDH wild type
198	TCGA-06-0166	IDH wild type
199	TCGA-02-0085	IDH wild type
200	TCGA-06-0139	IDH wild type
201	TCGA-76-6656	IDH wild type
202	TCGA-12-1093	IDH wild type
203	TCGA-08-0356	IDH wild type
204	TCGA-76-6664	IDH wild type
205	TCGA-08-0355	IDH wild type
206	TCGA-08-0350	IDH wild type
207	TCGA-06-0145	IDH wild type
208	TCGA-02-0054	IDH wild type
209	TCGA-FG-A713	IDH mutated
210	TCGA-DU-5855	IDH mutated
211	TCGA-CS-6668	IDH mutated
212	TCGA-HT-7884	IDH mutated
213	TCGA-FG-8189	IDH mutated
214	TCGA-FG-7634	IDH mutated
215	TCGA-HT-7856	IDH mutated
216	TCGA-CS-6665	IDH mutated
217	TCGA-HT-A5R5	IDH mutated
218	TCGA-FG-A6J1	IDH mutated
219	TCGA-DU-8167	IDH mutated
220	TCGA-DU-7302	IDH mutated
221	TCGA-06-2570	IDH mutated
222	TCGA-06-5417	IDH mutated
223	TCGA-DU-5874	IDH mutated
224	TCGA-HT-7476	IDH mutated
225	TCGA-DU-A6S6	IDH mutated
226	TCGA-14-1456	IDH mutated
227	TCGA-DU-6408	IDH mutated
228	TCGA-06-6389	IDH mutated
229	TCGA-HT-7874	IDH mutated
230	TCGA-HT-8114	IDH mutated
231	TCGA-CS-5390	IDH mutated
232	TCGA-06-0133	IDH wild type
233	TCGA-02-0046	IDH wild type
234	TCGA-08-0380	IDH wild type
235	TCGA-06-0154	IDH wild type
236	TCGA-12-0616	IDH wild type
237	TCGA-DU-A5TT	IDH wild type
238	TCGA-08-0390	IDH wild type
239	TCGA-HT-8558	IDH wild type
240	TCGA-08-0246	IDH wild type
241	TCGA-FG-5963	IDH wild type
242	TCGA-76-6285	IDH wild type
243	TCGA-08-0392	IDH wild type
244	TCGA-08-0385	IDH wild type
245	TCGA-DU-6404	IDH wild type
246	TCGA-02-0034	IDH wild type
247	TCGA-FG-7643	IDH wild type
248	TCGA-06-0649	IDH wild type
249	TCGA-14-1829	IDH wild type
250	TCGA-27-1835	IDH wild type
251	TCGA-02-0033	IDH wild type
252	TCGA-14-3477	IDH wild type
253	TCGA-06-0213	IDH wild type
254	TCGA-DU-5852	IDH wild type
255	TCGA-06-0192	IDH wild type
256	TCGA-19-5958	IDH wild type
257	TCGA-06-0238	IDH wild type
258	TCGA-02-0064	IDH wild type
259	TCGA-06-0122	IDH wild type
260	TCGA-HT-8019	IDH wild type

Disclosures

No conflicts of interest.

Acknowledgments

This research was supported by the U.S. National Cancer Institute under Grant No. U01CA207091 (A.J.M. and J.A.M.). We would like to thank statistician Dr. Yin Xi for help with the ROC and AUC analyses.

References

1. D. W. Parsons et al., "An integrated genomic analysis of human glioblastoma multiforme," *Science* **321**(5897), 1807–1812 (2008).
2. H. Yan et al., "IDH1 and IDH2 mutations in gliomas," *N. Engl. J. Med.* **360**(8), 765–773 (2009).
3. D. N. Louis et al., "The 2016 World Health Organization classification of tumors of the central nervous system: a summary," *Acta Neuropathol.* **131**(6), 803–820 (2016).
4. W. B. Pope et al., "Non-invasive detection of 2-hydroxyglutarate and other metabolites in IDH1 mutant glioma patients using magnetic resonance spectroscopy," *J. Neuro-Oncol.* **107**(1), 197–205 (2012).
5. C. Choi et al., "2-hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated patients with gliomas," *Nat. Med.* **18**(4), 624–629 (2012).
6. M. I. de la Fuente et al., "Integration of 2-hydroxyglutarate-proton magnetic resonance spectroscopy into clinical practice for disease monitoring in isocitrate dehydrogenase-mutant glioma," *Neuro-Oncology* **18**(2), 283–290 (2015).
7. A. Tietze et al., "Noninvasive assessment of isocitrate dehydrogenase mutation status in cerebral gliomas by magnetic resonance spectroscopy in a clinical setting," *J. Neurosurg.* **128**(2), 391–398 (2017).
8. C. Choi et al., "Prospective longitudinal analysis of 2-hydroxyglutarate magnetic resonance spectroscopy identifies broad clinical utility for the management of patients with IDH-mutant glioma," *J. Clin. Oncol.* **34**(33), 4030–4039 (2016).
9. C. Choi et al., "A comparative study of short- and long-TE 1H MRS at 3 T for *in vivo* detection of 2-hydroxyglutarate in brain tumors," *NMR Biomed.* **26**(10), 1242–1250 (2013).
10. S. K. Ganji et al., "In vivo detection of 2-hydroxyglutarate in brain tumors by optimized point-resolved spectroscopy (PRESS) at 7T," *Magn. Reson. Med.* **77**(3), 936–944 (2017).
11. R. L. Delfanti et al., "Imaging correlates for the 2016 update on WHO classification of grade II/III gliomas: implications for IDH, 1p/19q and ATRX status," *J. Neuro-Oncol.* **135**(3), 601–609 (2017).
12. K. Chang et al., "Residual convolutional neural network for the determination of IDH status in low- and high-grade gliomas from MR imaging," *Clin. Cancer Res.* **24**(5), 1073–1081 (2018).
13. X. Zhang et al., "Radiomics strategy for molecular subtype stratification of lower-grade glioma: detecting IDH and TP53 mutations based on multimodal MRI," *J. Magn. Reson. Imaging* **48**(4), 916–926 (2018).
14. P. Chang et al., "Deep-learning convolutional neural networks accurately classify genetic mutations in gliomas," *Am. J. Neuroradiol.* **39**(7), 1201–1207 (2018).
15. Z. Akkus et al., "Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence," *J Digital Imaging* **30**(4), 469–476 (2017).
16. K. Clark et al., "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *J. Digital Imaging* **26**(6), 1045–1057 (2013).
17. L. Scarpace et al., "Radiology data from the cancer genome atlas glioblastoma multiforme [TCGA-GBM] collection," *Cancer Imaging Arch.* **11**, 4 (2016).
18. N. Pedano, A. Flanders, and L. Scarpace, "Radiology data from the Cancer Genome Atlas Low Grade Glioma [TCGA-LGG] collection," *The Cancer Imaging Archive*, 2016, https://wiki.cancerimagingarchive.net/display/Public/TCGA-LGG#64c2b0756f974ab5b574ca3888851202.
19. M. Ceccarelli et al., "Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma," *Cell* **164**(3), 550–563 (2016).
20. X. Feng et al., "Deep learning on MRI affirms the prominence of the hippocampal formation in Alzheimer's disease classification," *bioRxiv* (2018).
21. D. A. Bluemke, "Editor's note: publication of AI research in radiology," *Radiology* **289**(3), 579–580 (2018).
22. X. Li et al., "The first step for neuroimaging data analysis: DICOM to NIfTI conversion," *J. Neurosci. Methods* **264**, 47–56 (2016).
23. C. Szegedy, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *AAAI*, Vol. **4** (2017).
24. C. Szegedy, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.* (2015).
25. G. Huang et al., "Densely connected convolutional networks," in *Comput. Vision and Pattern Recognit. (CVPR)* (2017).
26. K. He et al., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.* (2016).
27. H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat.* **22**, 400–407 (1951).
28. T. Zhang, "Solving large scale linear prediction problems using stochastic gradient descent algorithms," in *Proc. Twenty-First Int. Conf. Mach. Learn.*, ACM (2004).
29. F. Pedregosa et al., "Scikit-learn: machine learning in Python," *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
30. V. Wegmayr, S. Aitharaju, and J. Buhmann, "Classification of brain MRI with big data and deep 3D convolutional neural networks," *Proc. SPIE* **10575**, 105751S.

Sahil Nalawade is a research associate at the University of Texas, Southwestern Medical Center. He received his MS degree in biomedical engineering from the University of Texas, Arlington, in May 2017.

Gowtham K. Murugesan is currently pursuing his PhD in biomedical engineering at the University of Texas, Southwestern Medical Center, Arlington. He received his MS degree in biomedical engineering from the University of Texas, Arlington, in May 2016.

Joseph A. Maldjian is a professor of radiology at UT Southwestern Medical Center, the chief of neuroradiology, and holds the Lee R. and Charlene B. Raymond Distinguished Chair in Brain Research. An expert in advanced neuroimaging clinical and research applications, he has authored more than 150 peer-reviewed papers and has served on a number of review panels.

Biographies of the other authors are not available.