

# Brain tumor IDH, 1p/19q, and MGMT molecular classification using MRI-based deep learning: an initial study on the effect of motion and motion correction

Sahil S. Nalawade<sup>1b</sup>,<sup>a</sup> Fang F. Yu,<sup>a</sup>

Chandan Ganesh Bangalore Yogananda,<sup>a</sup> Gowtham K. Murugesan<sup>1b</sup>,<sup>a</sup>  
Bhavya R. Shah,<sup>a</sup> Marco C. Pinho,<sup>a</sup> Benjamin C. Wagner<sup>1b</sup>,<sup>a</sup> Yin Xi<sup>1b</sup>,<sup>a</sup>  
Bruce Mickey,<sup>b</sup> Toral R. Patel<sup>1b</sup>,<sup>b</sup> Baowei Fei<sup>1b</sup>,<sup>c</sup>

Ananth J. Madhuranthakam<sup>1b</sup>,<sup>a</sup> and Joseph A. Maldjian<sup>1b</sup><sup>a,\*</sup>

<sup>a</sup>University of Texas Southwestern Medical Center, Advanced Neuroscience Imaging  
Research Lab, Department of Radiology, Dallas, Texas, United States

<sup>b</sup>University of Texas Southwestern Medical Center, Department of Neurological Surgery,  
Dallas, Texas, United States

<sup>c</sup>University of Texas at Dallas, Department of Bioengineering, Richardson, Texas,  
United States

## Abstract

**Purpose:** Deep learning has shown promise for predicting the molecular profiles of gliomas using MR images. Prior to clinical implementation, ensuring robustness to real-world problems, such as patient motion, is crucial. The purpose of this study is to perform a preliminary evaluation on the effects of simulated motion artifact on glioma marker classifier performance and determine if motion correction can restore classification accuracies.

**Approach:** T2w images and molecular information were retrieved from the TCIA and TCGA databases. Simulated motion was added in the k-space domain along the phase encoding direction. Classifier performance for IDH mutation, 1p/19q co-deletion, and MGMT methylation was assessed over the range of 0% to 100% corrupted k-space lines. Rudimentary motion correction networks were trained on the motion-corrupted images. The performance of the three glioma marker classifiers was then evaluated on the motion-corrected images.

**Results:** Glioma marker classifier performance decreased markedly with increasing motion corruption. Applying motion correction effectively restored classification accuracy for even the most motion-corrupted images. For isocitrate dehydrogenase (IDH) classification, 99% accuracy was achieved, exceeding the original performance of the network and representing a new benchmark in non-invasive MRI-based IDH classification.

**Conclusions:** Robust motion correction can facilitate highly accurate deep learning MRI-based molecular marker classification, rivaling invasive tissue-based characterization methods. Motion correction may be able to increase classification accuracy even in the absence of a visible artifact, representing a new strategy for boosting classifier performance.

© 2022 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.9.1.016001](https://doi.org/10.1117/1.JMI.9.1.016001)]

**Keywords:** isocitrate dehydrogenase; magnetic resonance imaging; deep learning; motion correction; motion artifact simulation.

Paper 21152R received Jun. 17, 2021; accepted for publication Jan. 3, 2022; published online Jan. 27, 2022.

---

\*Address all correspondence to Joseph A. Maldjian, [Joseph.Maldjian@UTSouthwestern.edu](mailto:Joseph.Maldjian@UTSouthwestern.edu)

## 1 Introduction

Primary brain neoplasms demonstrate broad variations in imaging features, response to therapy, and prognosis. It has become evident that this heterogeneity is associated with specific molecular and genetic profiles. For example, isocitrate dehydrogenase 1 and 2 (IDH 1/2) mutated gliomas<sup>1</sup> demonstrate increased survival compared with wild-type gliomas with the same histologic grade. Additionally, 1p/19q codeletion<sup>2</sup> and O6-methyl guanine-DNA methyltransferase (MGMT) promoter methylation<sup>3</sup> are associated with differences in response to specific chemoradiation regimens.

Although the molecular profiling of gliomas is now a routine part of the evaluation of specimens obtained at biopsy or tumor resection, it would be helpful in many situations to have this information prior to surgery. In some cases, the information would aid in planning the extent of tumor resection. In others, for tumors in locations where resection is not possible and the risk of a biopsy is high, accurate delineation of the molecular and genetic profile of the tumor might be used to guide empiric treatment with radiation and/or chemotherapy. Recently, there have been advances in classifying tumor profiles using non-invasive imaging.<sup>4,5</sup> These classification algorithms have been designed based on linear regression models, classical machine learning,<sup>6–8</sup> and, more recently, deep learning networks<sup>9</sup> and have shown particular promise by outperforming other approaches.

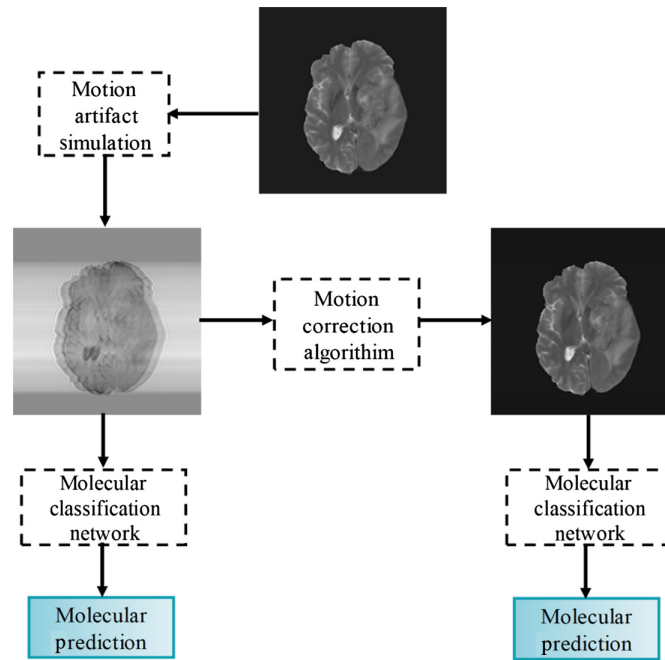
Our group recently achieved 97%, 93%, and 95% accuracy for classifying IDH mutation status,<sup>5</sup> 1p/19q,<sup>10</sup> and MGMT,<sup>11</sup> respectively, in primary brain tumors utilizing T2w MR images alone. An important caveat is that the effects of degradation on the input images, such as motion artifact, and, in turn, on the performance of deep learning-based classifiers, have not been systematically studied. Motion artifacts are an especially pervasive source of MR image quality degradation and can be due to gross patient movements, as well as cardiac and respiratory motion.<sup>12,13</sup> In clinical practice, these artifacts can interfere with diagnostic interpretation,<sup>14</sup> necessitating repeat imaging in up to 20% of cases.<sup>15</sup> Pei et al.<sup>16</sup> applied physical models of motion blurring to non-medical images and tested the classification performance of two deep learning networks, demonstrating decreased accuracies. It is likely that motion corruption will also lead to reduced performance of deep-learning algorithms in classifying brain tumor images.

The goals of our study were (1) to perform a preliminary evaluation on the effect of motion corruption on deep learning-based molecular marker classification accuracy in gliomas and (2) to determine if a rudimentary deep learning motion correction algorithm can recover classification accuracies to levels similar to non-corrupted images. In assessing the effects of motion artifact corruption on classification accuracies, we utilized our previously developed top-performing deep learning networks for determination of IDH mutation status,<sup>5</sup> 1p/19q codeletion,<sup>10</sup> and MGMT methylation. These networks use only T2w images and have provided the highest MRI-based classification accuracies reported to date, approaching those of invasive tissue-based methods.

## 2 Materials and Methods

Our study design consisted of (1) simulating motion in the original T2w glioma images, (2) training a network on the motion-simulated images to generate artifact-free images using the original non-distorted images as ground truth, (3) evaluating the motion correction performance of the network on the held-out subjects, and (4) comparing the performance of our previously trained glioma molecular classification networks<sup>5,10,11</sup> using the motion-corrupted and motion-corrected images, respectively.

The design, training, and performance of our previously trained molecular marker classification networks for IDH, 1p/19q, and MGMT<sup>5,10,11</sup> are briefly summarized here. These networks were designed with a 3D-Dense-UNet architecture and trained using a  $32 \times 32 \times 32$  patch-based approach for voxelwise dual-class segmentation of the whole tumor. The networks consisted of (i) an initial convolution layer, (ii) an encoder with three dense blocks and three transition down blocks, (iii) a decoder with three dense blocks and three transition up blocks, (iv) a bottleneck dense block, and (v) a final convolution layer followed by an activation layer. T2w MR images from 214 subjects in the TCIA were used to train and test the IDH network. The IDH network



**Fig. 1** Overview of the study design.

was trained with classes 1 and 2 representing IDH mutated and IDH wild-type, respectively. The network yields 2 segmentation volumes, which are combined to provide a single tumor segmentation map. A majority voting scheme provides a single IDH classification for each subject. The trained IDH network demonstrated a 97% mean cross-validation accuracy for IDH prediction on the TCIA data, representing the highest accuracy to date for noninvasive IDH prediction. A transfer learning approach was implemented on the IDH network to develop both 1p/19q and MGMT classification networks, with fine-tuning only of the decoder part of the IDH-network. For the 1p/19q-network, 368 subjects from the TCIA were used to fine-tune and test the network. Classes 1 and 2 represented 1p/19q co-deleted and 1p/19q non-co-deleted types, respectively. A mean cross-validation accuracy of 93% was obtained for the final 1p/19q network on the TCIA data representing the highest accuracy to date for non-invasive 1p/19q prediction. For the MGMT-network, 247 subjects from the TCIA were used to fine-tune and test the network. Classes 1 and 2 represented methylated and unmethylated MGMT promoter types, respectively. A mean cross-validation accuracy of 95% was obtained for the final MGMT network on the TCIA data representing the highest accuracy to date for noninvasive MGMT prediction. The entire procedure of molecular marker classification using these networks is fully automated, and a tumor segmentation map is a natural output of the voxelwise classification approach. For the current work, we evaluated the effects of motion and motion correction on these pretrained high-performing molecular classification networks.

For this work, motion simulation was introduced only in-plane along the phase encoding direction as a proof-of-principle rather than as an exhaustive study of all motion artifacts. Similarly, rudimentary motion correction networks were trained for recovery of in-plane motion, again as proof-of-principle. Figure 1 shows an overview of the study design.

## 2.1 Dataset and Preprocessing

Individual subject imaging data were retrieved from the TCIA database,<sup>17</sup> and corresponding tumor genomic information was obtained from the TCGA database.<sup>18</sup> Only preoperative cases with T2w images were included. The final IDH dataset consisted of 214 subjects (94 IDH-mutated and 120 IDH wild-type). Imaging and genomic data from 368 subjects with a 1p/19q co-deletion status (130 1p/19q co-deleted and 238 non-co-deleted) and 247 subjects with a MGMT methylation status (163 MGMT methylated and 84 unmethylated) were also obtained from the TCIA and TCGA databases.

Minimal preprocessing was applied to the imaging datasets and consisted of (1) co-registration of the T2w images to the SR124 T2w template<sup>19</sup> using Advanced Normalization Tools software,<sup>20</sup> (2) skull stripping of the T2w images using the Brain Extraction Toolkit (FMRIB software library),<sup>21</sup> (3) N4 bias field correction to remove radiofrequency pulse inhomogeneity, and (4) image intensity normalization by performing zero mean and unit variance. The preprocessing steps required <5 min per subject.

## 2.2 Motion Simulation

K-space data were obtained after applying a 2D inverse Fourier transformation to the T2w image. Motion artifacts were simulated by incorporating additional phase to the k-space data along the phase encoding direction according to Eq. (1) as<sup>22</sup>

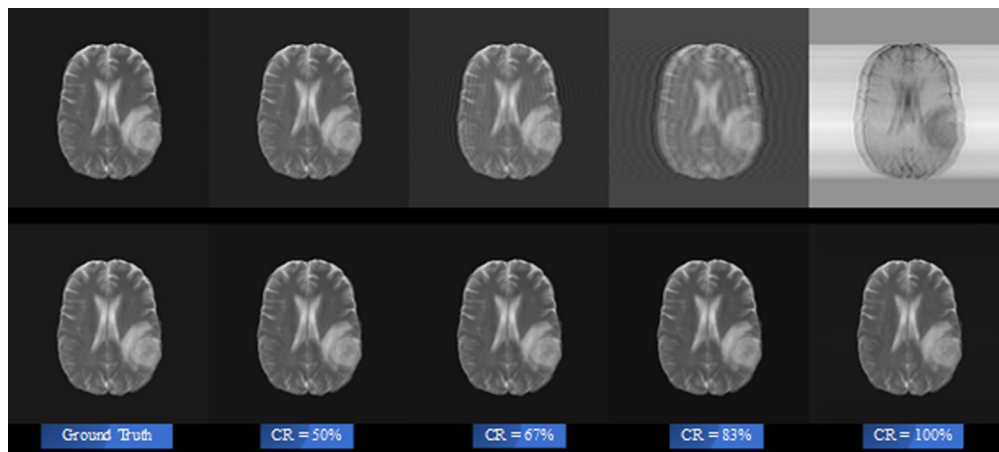
$$SM_{xy}(ky) = M_{xy}(ky) \times e^{(-i2\pi k_y \theta(k_y))}; \quad \begin{cases} k_y = -\frac{N_y}{2} \text{ to } -\frac{(N_y-N)}{2} \\ k_y = +\frac{(N_y-N)}{2} \text{ to } +\frac{N_y}{2} \end{cases} \quad (1)$$

where  $k_y$  represents the k-space data along the phase encoding direction,  $M_{xy}(k_y)$  is the original k-space,  $SM_{xy}(k_y)$  is the motion simulated k-space, and  $\theta(k_y)$  is the phase induced by motion. This approach closely simulates the additional phase induced by translational patient movements.<sup>23</sup>

The total number of corrupted k-space lines is given by  $N$ , such that the outermost  $N/2$  lines on either side of k-space are corrupted. The corruption rate (CR) represents the percentage of corrupted k-space (Fig. 2), where  $CR = N/N_y$  with  $N_y$  being the total number of phase encoding lines (e.g.,  $N_y = 240$ ). In this study, the number of corrupted k-space lines ( $N$ ) is 10, 20, 60, 80, 100, 120, 140, 150, 160, 180, 200, 220, and 240, which corresponds to CRs of 4%, 8%, 25%, 33%, 42%, 50%, 58%, 63%, 67%, 75%, 83%, 92%, and 100%, respectively. These CR values were selected to represent a broad range of motion artifacts, from minimal to highly corrupted images.

## 2.3 Network Architectures

For the motion correction networks, we chose to evaluate three networks. These included a 2D Dense U-Net (Blur-Net), SE-Net154, and SE-Net154 with a perceptive blur metric. These networks are fully described in the [Appendix](#). Briefly, Model-1 (Blur-Net) was adapted from a 2D Dense-UNet architecture<sup>24</sup> consisting of four transition down blocks and four transition up



**Fig. 2** Example of simulated motion and motion correction. Top Row: Simulated motion. From left to right, ground truth T2w image (column 1) and corrupted images for CR=50%, 67%, 83%, and 100% (columns 2-5). Bottom Row: Motion correction using Model-1. From left to right, corrected output images for CR = 0%, 50%, 67%, 83%, and 100%.

blocks with an initial and final convolution layer. Model-2 was based on a modified squeeze and excitation network (SE-Net).<sup>25</sup> SE-Net has shown promising results in image classification, winning the 2017 ImageNet Large Scale Visual Recognition and Classification challenge.<sup>26</sup> It consists of an input block, four transition down blocks, and four transition up blocks. Model-3 was largely the same as Model-2, with the only difference being the use of a different loss function that incorporated the perceptive blur metric to help the algorithm learn and reduce errors in its predictions, particularly with respect to image sharpness.<sup>27</sup>

## 2.4 Training

Training of the motion correction networks was performed in two phases. First, the three motion correction models were trained using 213 subjects from the IDH dataset. The subjects were randomly shuffled into three groups for training, in-training validation, and held-out testing (~71 subjects per set). The performance of all three motion correction networks was compared using the held-out testing set, which was not used during the training steps. The best performing network was then selected and retrained using a larger combined dataset of 446 unique subjects from all three molecular marker groups (IDH, 1p/19q, and MGMT). This combined dataset was randomly shuffled into three groups to perform three-fold cross-validation. For each of the three folds, the groups were alternated between training, in-training validation, and held-out testing sets (~149 subjects per set). The 2D slices were separated by subject for each of the cross-validation folds to eliminate the problems of subject duplication and data leakage.<sup>28,29</sup>

Data augmentation was performed on the input T2w images, including horizontal and vertical flipping, to increase training quality and diversity, which helps for training models with limited data. The networks were implemented on NVIDIA Tesla V100 GPUs using Keras, a python package with Tensorflow<sup>30</sup> as the backend, with an adaptive moment optimizer. The initial learning rate of the optimizer was set at  $1 \times 10^{-5}$ . Model-1 and Model-2 were trained using a combined loss function of structural similarity index (SSIM) loss, peak signal-to-noise ratio (PSNR) loss, mean absolute error (MAE) loss, and perceptual loss, with equal weighting for the structural components, noise level, and perceived image quality. MAE loss was selected over MSE because MAE is believed to be more robust against the outliers and the network convergence is faster with MAE.<sup>31</sup> The loss function for Model-3 differed in its use of blur loss<sup>27</sup> instead of MAE and perceptual loss. All networks were trained from scratch with a batch size of 4, and training times ranged between 96 and 120 h.

## 2.5 Testing

The three trained motion correction networks were evaluated on the held-out testing set. The motion corrupted input images and predicted motion-corrected output images were compared with the ground truth non-corrupted reference images. The performance of the models was evaluated using SSIM, PSNR, and normalized mean squared error (NMSE). The best performing motion correction network was then retrained on the combined dataset and evaluated on the held-out testing set for each of the three cross-validation folds. The results from each fold were averaged across all subjects for each corruption level. The testing time for each subject was <60 s.

The classification accuracy for the IDH mutation status was initially evaluated for the three motion correction networks to further corroborate the selection of the best-performing network. Molecular classification accuracies for IDH, 1p/19q, and MGMT promoter were then evaluated using the retrained best-performing motion correction network. Accuracies were determined using the ground truth uncorrupted images and at each of the image corruption levels (from 4% to 100% CR) for each cross-validation fold using our previously trained deep learning molecular classification networks.<sup>5,10,11</sup> The results were averaged across folds to provide a mean classification accuracy for each molecular marker at each corruption level. This process was then repeated on the motion-corrected images for IDH, 1p/19q, and MGMT at each corruption level to determine if the ground truth accuracies could be recovered. It is worth noting that the cross-validation folds for the motion correction training network were designed to exclude the subjects in the molecular marker testing folds to avoid bias in determining accuracy recovery.

**Table 1** Motion correction model performance for selecting the best network.

	SSIM	PSNR	NMSE	SSIM	PSNR	NMSE	SSIM	PSNR	NMSE
Model	Output for CR = 100%			Output for CR = 92%			Output for CR = 83%		
Model-1	99.47	44.39	0.01	99.72	49.62	0.00	99.76	50.95	0.00
Model-2	99.23	42.53	0.01	99.23	42.53	0.01	99.23	42.53	0.01
Model-3	99.03	41.60	0.02	99.18	42.35	0.01	99.28	42.79	0.01

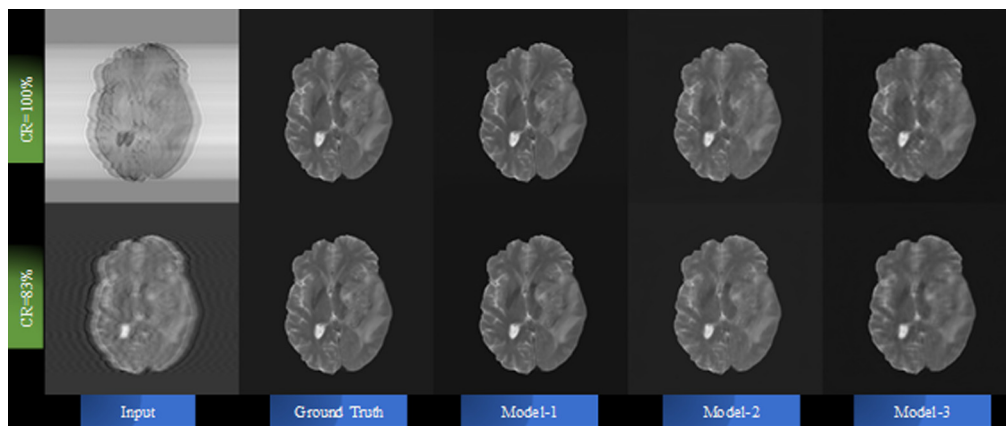
### 3 Experimental Results

#### 3.1 Comparison of Motion Correction Algorithms

Table 1 shows the SSIM, PSNR, and NMSE metrics for our three motion correction networks on the held-out testing dataset at the three highest motion corruption levels. Model-1 achieved the best performance across all three metrics, most notably PSNR. Figure 2 shows an example of the motion corruption and motion correction performance for Model-1. Figure 3 shows the motion-corrected output images generated by the three networks for a single subject at high levels of motion corruption (CR = 83% and 100%). Although the three models generated similar results up to CR = 50%, at higher corruption levels (CR = 92%), Model-1 generated sharper images that were visually indistinguishable from the ground truth image (Fig. 4), surpassing the other models in terms of quantitative metrics as well as perceived visual quality.

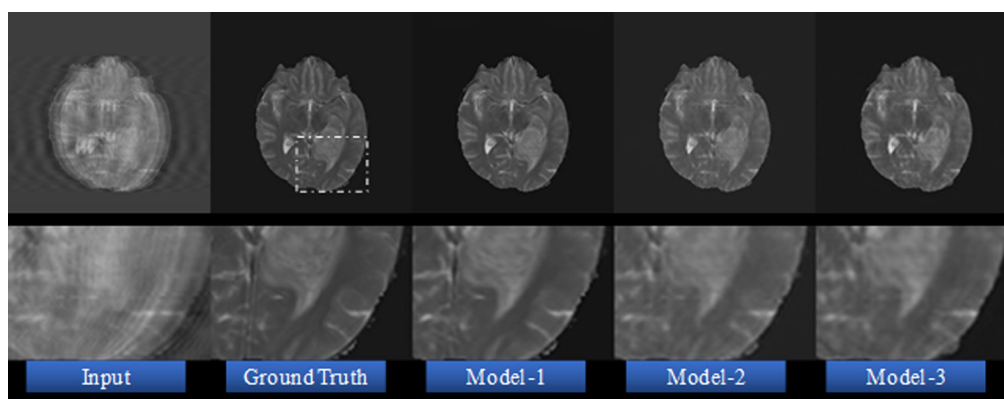
#### 3.2 Effects of Motion and Motion Correction on IDH/1p19q/MGMT Classification Accuracy

Figure 5 compares IDH classification accuracy for the three motion correction networks using uncorrected motion corrupted images and motion-corrected images at increasing motion corruption levels. The IDH classification began to fail on the motion corrupted images at a CR of 42% and progressively deteriorated with higher corruption levels. Model-1 achieved the best results, maintaining a 97% IDH classification accuracy through a CR of 92%. Although the other two models were able to improve classification accuracy over the full range of image corruption levels, they were unable to match the performance of Model-1.

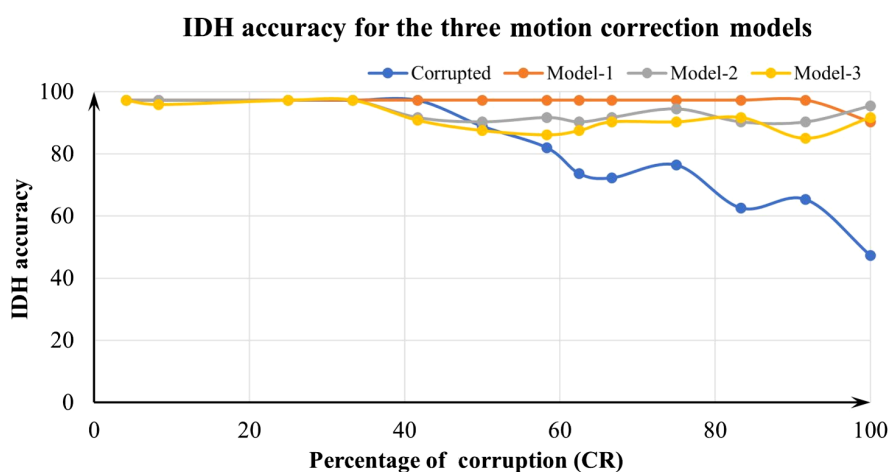


**Fig. 3** Examples of motion correction at high CR levels for the three models. Input corrupted image (column 1) at CR of 100% (top row) and 83% (bottom row), ground truth (column 2), Model-1 output (column 3), Model-2 output (column 4), and Model-3 output (column 5). Model-1 provided visually obvious improved performance over the other two models.



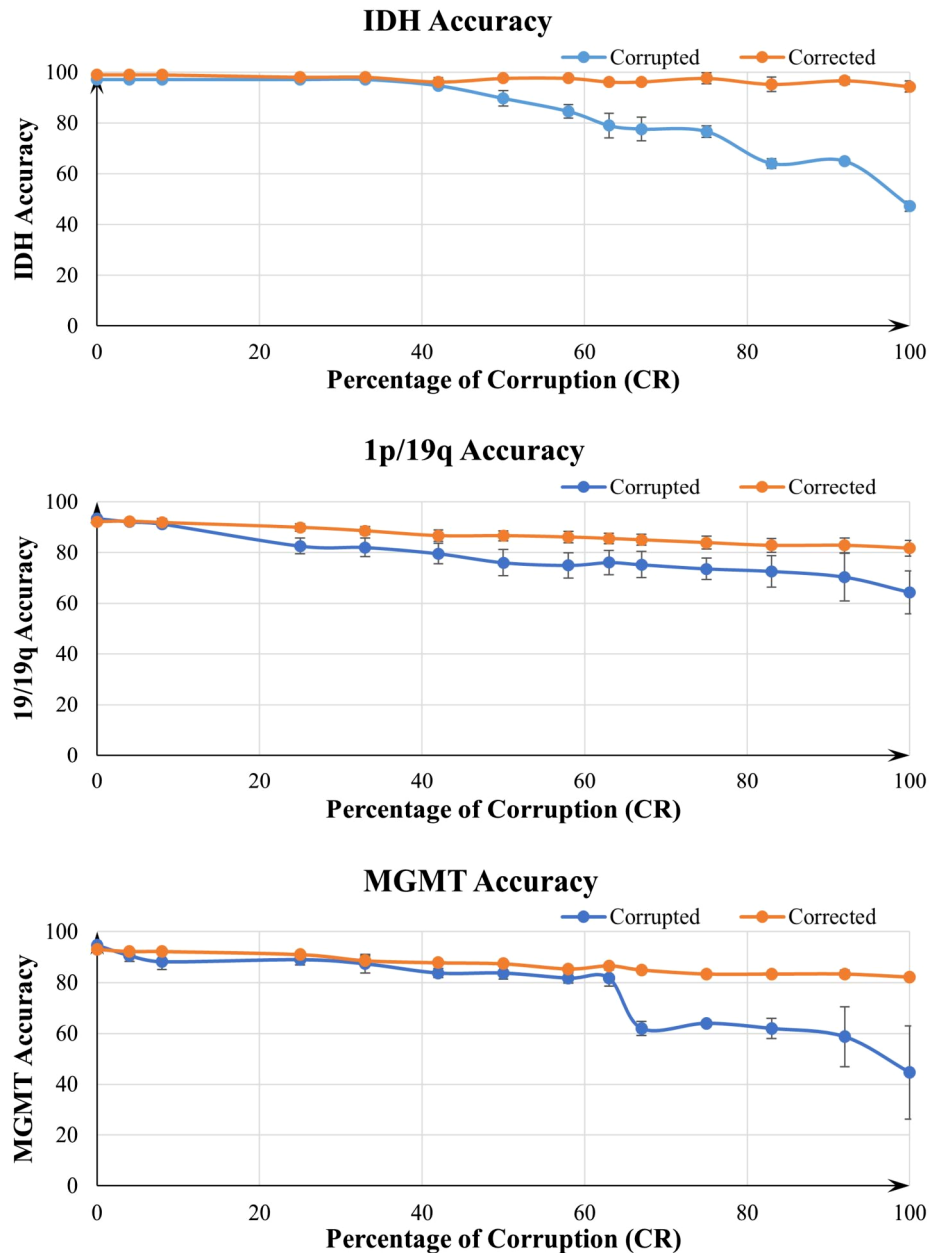


**Fig. 4** Example motion correction at high corruption level (CR = 92%) for the three models. Input corrupted image (column one), ground truth (column 2), Model-1 output (column 3), Model-2 output (column 4), and Model-3 output (column 5). Model-1 provided visually obvious improved performance over the other two models as evidenced by the sharpness of the sulci and tumor borders. Bottom row provides view of inset for input, ground truth, and each model.



**Fig. 5** IDH classification accuracy with respect to percent corruption for motion-corrupted images and motion-corrected images for the three correction networks. Motion corrupted accuracies (blue), as well as accuracies following motion correction for Model-1 (orange), Model-2 (gray), and Model-3 (yellow), are shown. A progressive decrease in classification accuracy for the corrupted images is demonstrated beyond 42% CR (blue line). Model-1 performed best (orange line), recovering the original 97% classification accuracy out to 92% CR.

Figure 6 shows the IDH, 1p/19q co-deletion, and MGMT methylation status classification performance on the motion corrupted images and the recovery of accuracy using the best performing network (Model-1) after it was retrained on the larger combined dataset. The classification accuracy on the corrupted images declined at 42% CR for both IDH and 1p/19q, whereas MGMT performance declined at 63% CR. For the corrected images, IDH classification was maintained at 97% accuracy out to 92% CR and recovered to 94% accuracy even at 100% CR. More remarkably, for correction of the native images and at lower levels of image corruption (0% to 33%), IDH classification accuracy exceeded the performance of the uncorrupted images achieving up to 99% accuracy. For both 1p/19q and MGMT, 82% accuracy was recovered out to 100% CR.



**Fig. 6** IDH, 1p/19q, and MGMT classification accuracies for motion corrupted (blue lines) and Model-1 corrected images (orange lines) averaged across the three cross-validation folds for each molecular marker. Recovery of accuracy was best for IDH classification, boosting the accuracy to 99% for the baseline uncorrupted images and low-levels of motion corruption and recovering the original 97% accuracy out to 92% CR.

## 4 Discussion

We demonstrate that the performance of previously trained top-performing glioma molecular classifiers (IDH mutation, 1p/19q co-deletion, and MGMT methylation) was adversely affected by simulated motion corruption, with progressive loss in accuracy with increasing motion corruption. We developed and evaluated three motion correction algorithms that were able to handle a broad range of translational motion corruption levels for glioma T2w images. Importantly, classification accuracies could be recovered or significantly improved after applying motion correction, even at very high levels of motion corruption. In the case of IDH classification,



99% accuracy was achieved following motion correction, exceeding the performance on the native ground truth images.<sup>5</sup>

Although we expected that molecular classification accuracy would be degraded by motion corruption, we were surprised by the relative resilience of the networks to motion corruption, retaining accuracies up to corruption levels of ~40%. The algorithms that we previously developed for brain tumor molecular classification were trained on relatively motion-artifact free images from the TCIA database.<sup>5,10,11</sup> A baseline level of motion artifact is likely present when averaged throughout this data and could partly account for the robustness of classification accuracy. However, the performance declined sharply at image corruption levels beyond CR = 42%. This work also provided a serendipitous observation, wherein the application of the motion correction network boosted the IDH classification up to an astounding 99% accuracy for the native images without any added simulated motion, representing a new benchmark for IDH classification. This may reflect the presence of latent image artifacts within some of the ground truth images that obscured image features important for classification of IDH status and were then removed by the motion correction algorithm. This also points to a potential new strategy for boosting deep learning classifier performance with the use of motion or artifact correction networks, even when there is no visible motion.

Of the three motion correction algorithms, Model-1 performed the best. This model was based on a 2D Dense-UNet architecture, which may account for its superior performance. With its densely connected design, all feature maps are reused, such that each layer in the architecture received a direct supervision signal. In addition, the Dense-UNet architecture alleviates the vanishing gradient problem in machine learning, which can prevent the network from further improvement. Other advantages include feature propagation and feature reuse, as previously described.<sup>32</sup>

All three models achieved excellent performance with SSIM of over 0.99 and RMSE of less than 0.03 for all motion corruption levels. For comparison, Duffy et al.<sup>23</sup> achieved SSIM and RMSE of 0.97 and 0.04, respectively, whereas Sommer et al.<sup>33</sup> reported SSIM of 0.86 to 0.924 when comparing motion-corrected brain MRIs with uncorrupted ground-truth images. In contrast to prior studies that used limited sets of motion corruption levels (Duffy et al. applied corruption to 30 lines of k-space), we trained our models using a broad range of motion corruption levels, from a minimum (4%) to 100% of k-space lines affected. This approach better captures real-world conditions in which there is a mixture of motion artifacts, from mild motion that largely preserves diagnostic information to more severe cases that are uninterpretable diagnostically.

In the setting of large amounts of motion, the MRI acquisition would typically be repeated. However, in the clinical scenario, time constraints can preclude reacquisition of heavily motion corrupted images, or there can be circumstances in which patient motion cannot be overcome without sedation. Although deep-learning motion correction can be regarded as a preprocessing step in the classification pipeline, an alternative approach would be to intentionally train the molecular classifier networks using corrupted imaging data. Motion (and other image artifacts) could be directly incorporated into the molecular classification network as augmentation steps during network training. A potential caveat for such a strategy is that it could lead to the networks erroneously learning incorrect imaging features (in the form of motion corrupted imaging features or the motion artifacts themselves) as the basis for classifying the molecular markers. As such, deep learning image-based classification studies have excluded data with significant artifacts from their training database.<sup>4,6,34</sup> Alternatively, conventional non-machine-learning-based motion correction strategies could be utilized before applying the molecular classification algorithms. However, a key advantage to our deep learning approach is that it may be applied retrospectively to any previously acquired image without the need for any additional acquisition time, special scanner preparatory steps, or additional input data.

The TCIA dataset has a variety of gliomas with different biological behaviors, including glioblastoma, anaplastic astrocytoma, low-grade glioma, and oligodendroglioma, with their associated variations in IDH mutation, 1p/19q co-deletion, and MGMT methylation status. We performed an analysis of classification accuracy of our networks with respect to tumor grade and found no significant differences in network performance. Importantly, our motion correction networks, particularly Model-1 (Blur-Net), were able to not only remove the motion artifacts but

also preserve the key MR imaging features of the tumors necessary for accurate classification. This was evidenced by the full recovery of classification accuracy for the IDH network extending out to a corruption level of 92% and markedly improved accuracies for 1p/19q co-deletion and MGMT methylation networks following the application of the Blur-Net motion correction algorithm. These compelling results support the routine use of a deep learning-based image artifact removal step for imaging-based deep learning applications to classify glioma molecular profiles. We demonstrated that this implementation enhances the robustness of the classification pipeline to real-world challenges, which facilitates its potential clinical feasibility and implementation.

## 5 Limitations and Future Directions

It is important to note that this study was an initial evaluation and not meant to be an exhaustive study of artifacts and artifact correction networks on molecular classifier performance. For this initial proof-of-principle, we focused on the effect of translational motion artifacts in MR images on deep learning molecular classification, although we recognize that other artifacts such as rotational motion artifacts, magnetic field inhomogeneity, Gaussian noise, and radiofrequency spikes can also affect MR image quality. Similarly, our motion correction networks were only trained to recover translational motion for this proof-of-principle. We recognize that network architectures and artifact correction methods are constantly evolving, and much more elaborate motion correction strategies can be devised. Our approach, however, does provide a framework for training, evaluating, and benchmarking artifact-correction architectures for potential insertion into a workflow. Although we achieved excellent performance for recovering classification accuracy for glioma molecular profiles, our study was confined to the TCIA database. Before implementation in the clinical environment, it will be essential to train and validate using additional independent datasets.

Although our molecular classification networks performed better using motion-corrected images than motion-corrupted images, we could not fully restore the classification accuracies of the 1p/19q and MGMT networks achieved using uncorrupted images. These findings indicate that the three classification algorithms differed in terms of their resilience to motion artifact. Both the 1p/19q and MGMT networks were based on the pre-trained IDH classifier network, with fine-tuning to the decoder part of the network to adjust classification weights without changes to the encoder part. This led to faster training and resultant excellent classification accuracies using uncorrupted images but may have also rendered the networks less robust to image corruption than the fully-trained IDH network. Although we achieved superior motion correction results compared with previous studies, subtle residual artifacts within the image appear to have been sufficient to affect molecular classification performance. It is also possible that performance could be enhanced with modifications to the motion correction network architecture. We chose to use a 2D network design for the associated lower computational demands, as well as the fact that the TCIA database contained 2D T2w images. However, 3D architectures can be adapted in the future.

## 6 Conclusion

We evaluated the effect of simulated translational motion artifacts on glioma molecular classification networks and the ability of rudimentary motion correction networks to recover classification accuracy. We demonstrate that high-performing classification networks for IDH mutation status, 1p/19q co-deletion, and MGMT methylation progressively lose accuracy with increasing motion-related image degradation. However, by incorporating motion correction prior to the classification step, recovery of classification network accuracy was possible even at the highest degrees of motion disruption, indicating that the network was successful not only at removing artifacts but also in recovering crucial imaging features of the tumors. After training the motion correction network on a larger dataset composed of all three glioma markers, we achieved 99% accuracy for IDH classification, representing a new benchmark in non-invasive image-based IDH classification performance. More remarkably, classification accuracy was boosted even in the absence of added simulated motion in the native images. This provides

a potential new strategy for boosting deep learning classifier performance by including the use of motion or artifact correction networks, even when there is no visible motion.

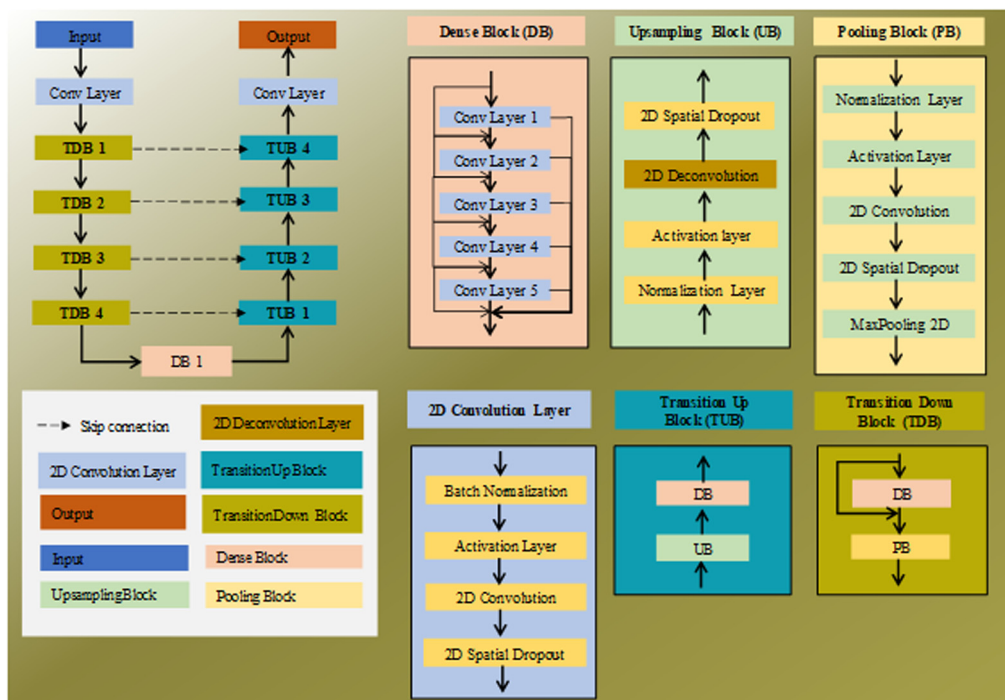
## 7 Appendix

### 7.1 Network Architecture

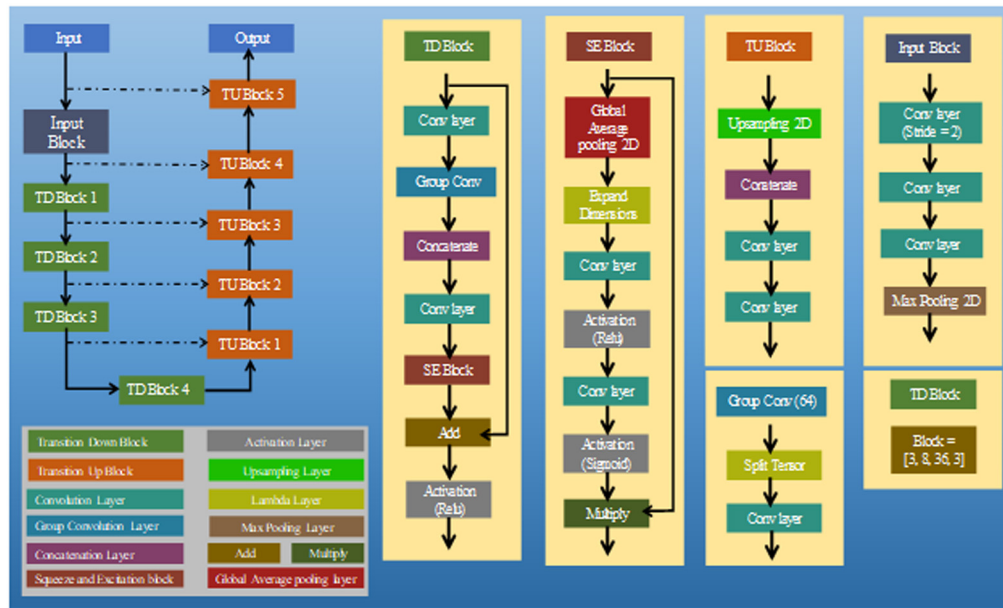
#### 7.1.1 Model-1 (Blur-Net)

Figure 7 shows the network architecture for Model-1 (Blur-Net), which is adapted from a 2D Dense-UNet architecture.<sup>24</sup> It consists of four transition down blocks, and four transition up blocks with an initial and a final convolution layer. Each transition down block consists of a dense block and a pooling block, and each transition up block consists of an up-sampling block and a dense block. Each dense block has five densely connected convolutional layers, where each layer is connected to every other layer. The feature maps of all of the convolutional layers in the dense block were concatenated to the output of the block, providing a dense connection. The output of the dense block was also concatenated with the input.

The encoder part of the network has four transition down blocks, which are comprised of 4 dense blocks and a pooling block. Each pooling block contains a batch normalization layer, activation layer, convolution layer, spatial dropout layer, and maximum pooling layer. The decoder part of the network has four transition up blocks, which are comprised of 4 dense blocks, each of which is preceded by an up-sampling block. Up-sampling blocks are comprised of a batch normalization layer, activation layer, deconvolution layer, and spatial dropout. The activation layer used was a rectified linear unit. Dense block 1 was used as a bottleneck in the network, which helps reduce the number of feature maps. This reduction can assist with memory optimization, allowing the network to operate within the resource limits of the GPU. Additionally, as the number of features was reduced, the network was trained in less time. A total of 50 2D densely connected convolution layers were implemented in the Blur-Net architecture.



**Fig. 7** Architecture of the Blur-Net network (Model-1).



**Fig. 8** Architecture of SE-Net 154 (Model-2 and Model-3). For Model-3, the perceptible blur metric (blur loss) was substituted as the loss function.

### 7.1.2 Model-2 (SE-Net 154)

The network architecture of Model-2 (Fig. 8) was based on a modified SE-Net.<sup>25</sup> SE-Net has shown promising results in image classification, winning the 2017 ImageNet Large Scale Visual Recognition Challenge classification challenge.<sup>26</sup> The SE-Net 154-based motion correction architecture consists of an input block, four transition down blocks, and four transition up blocks. The input block consists of a three-convolution layer and a maximum pooling layer. Each transition down block consists of two convolution layers as well as a group convolution layer, concatenation layer, squeeze and excitation block, addition layer, and activation layer. The group convolution layer split the input tensor into the number of groups, and then each group ran through the convolution layer. In our network, the input tensor was split into 64 groups. The final outputs of all group convolutions were then concatenated. The squeeze and excitation block is comprised of a global average pooling layer, lambda layer, two convolution layers, rectified linear unit activation layer, sigmoid activation layer, and multiplication layer. The lambda layer was used for expanding the dimensions of the input tensor. Each transition up block consists of an up-sampling layer, concatenation layer, and two convolution layers. Each transition down block was iterated sequentially. Transition down blocks 1, 2, 3, and 4 were iterated for 3, 8, 36, and 3, respectively. A total of 3407 convolution layers were implemented in the SE-Net 154 architecture.

### 7.1.3 Model-3 (SE-Net 154) with blur loss

The network architecture for Model-3 is largely the same as Model-2, with the only difference being the use of a different loss function that incorporated the perceptible blur metric to evaluate image blurriness.<sup>27</sup> The use of this metric was intended to help the algorithm learn and reduce errors in its predictions, particularly with respect to image sharpness.

## Disclosures

No conflicts of interest.

## Acknowledgments

Support for this research was provided by the National Cancer Institute (NCI), Grant Nos. U01CA207091 (AJM, JAM) and R01CA260705 (JAM).

## References

1. D. W. Parsons et al., "An integrated genomic analysis of human glioblastoma multiforme," *Science* **321**, 1807–1812 (2008).
2. M. J. Van Den Bent et al., "Adjuvant procarbazine, lomustine, and vincristine chemotherapy in newly diagnosed anaplastic oligodendroglioma: long-term follow-up of EORTC Brain Tumor Group Study 26951," *J. Clin. Oncol.* **31**, 344–350 (2013).
3. M. E. Hegi et al., "MGMTGene silencing and benefit from temozolomide in glioblastoma," *N. Engl. J. Med.* **352**, 997–1003 (2005).
4. S. Nalawade et al., "Classification of brain tumor isocitrate dehydrogenase status using MRI and deep learning," *J. Med. Imaging* **6**(4), 046003 (2019).
5. C. G. Bangalore Yogananda et al., "A novel fully automated MRI-based deep-learning method for classification of IDH mutation status in brain gliomas," *Neuro Oncol.* **22**, 402–411 (2020).
6. 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**, 916–926 (2018).
7. P. Korfiatis et al., "MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas," *Med. Phys.* **43**, 2835–2844 (2016).
8. T. Meraj et al., "Lung nodules detection using semantic segmentation and classification with optimal features," *Neural Comput. Appl.* **33**, 10737–10750 (2020).
9. J. E. Villanueva-Meyer et al., "MRI features and IDH mutational status of grade II diffuse gliomas: impact on diagnosis and prognosis," *Am. J. Roentgenol.* **210**, 621–628 (2018).
10. C. G. B. Yogananda et al., "A novel fully automated MRI-based deep-learning method for classification of 1p/19q co-deletion status in brain gliomas," *Neurooncol. Adv.* **2**, vdaa066 (2020).
11. C. Yogananda et al., "MRI-based deep-learning method for determining glioma MGMT promoter methylation status," *Am. J. Neuroradiol.* **42**, 845–852 (2021).
12. D. R. Enzmann et al., "CSF pulsations within nonneoplastic spinal cord cysts," *Am. J. Roentgenol.* **149**, 149–157 (1987).
13. B. O. Kjos et al., "Reproducibility of relaxation times and spin density calculated from routine MR imaging sequences: clinical study of the CNS," *Am. J. Roentgenol.* **144**, 1165–1170 (1985).
14. M. Zaitsev, J. Maclaren, and M. Herbst, "Motion artifacts in MRI: a complex problem with many partial solutions," *J. Magn. Reson. Imaging* **42**, 887–901 (2015).
15. J. Maclaren et al., "Prospective motion correction in brain imaging: a review," *Magn. Reson. Med.* **69**, 621–636 (2013).
16. Y. Pei et al., "Effects of image degradation and degradation removal to CNN-based image classification," *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 1239–1253 (2021).
17. K. Clark et al., "The cancer imaging archive (TCIA): maintaining and operating a public information repository," *J. Digit. Imaging* **26**, 1045–1057 (2013).
18. M. Ceccarelli et al., "Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma," *Cell* **164**, 550–563 (2016).
19. T. Rohlfing et al., "The SRI24 multichannel atlas of normal adult human brain structure," *Hum. Brain Mapp.* **31**, 798–819 (2010).
20. B. B. Avants et al., "A reproducible evaluation of ANTs similarity metric performance in brain image registration," *Neuroimage* **54**, 2033–2044 (2011).
21. S. M. Smith, "Fast robust automated brain extraction," *Hum. Brain Mapp.* **17**, 143–155 (2002).
22. T. A. Gallagher, A. J. Nemeth, and L. Hacin-Bey, "An introduction to the Fourier transform: relationship to MRI," *Am. J. Roentgenol.* **190**, 1396–1405 (2008).
23. B. A. Duffy et al., "Retrospective correction of motion artifact affected structural MRI images using deep learning of simulated motion," in *1st Conf. Med. Imaging Deep Learn.*, Amsterdam, The Netherlands (2018).
24. S. Jégou et al., "The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit., Workshops*, pp. 11–19 (2017).



25. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 7132–7141 (2018).
26. O. Russakovsky et al., "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.* **115**, 211–252 (2015).
27. F. Crete et al., "The blur effect: perception and estimation with a new no-reference perceptual blur metric," *Proc. SPIE* **6492**, 64920I (2007).
28. V. Wegmayr, S. Aitharaju, and J. Buhmann, "Classification of brain MRI with big data and deep 3D convolutional neural networks," *Proc. SPIE* **10575**, 105751S (2018).
29. X. Feng et al., "Deep learning on MRI affirms the prominence of the hippocampal formation in Alzheimer's disease classification," *bioRxiv* **2018**, 456277 (2018).
30. M. Abadi et al., "Tensorflow: a system for large-scale machine learning," in *12th USENIX Symp. Oper. Syst. Des. and Implementation*, pp. 265–283 (2016).
31. W. Yang et al., "Deep learning for single image super-resolution: a brief review," *IEEE Trans. Multimedia* **21**, 3106–3121 (2019).
32. G. Huang et al., "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 4700–4708 (2017).
33. K. Sommer et al., "Correction of motion artifacts using a multiscale fully convolutional neural network," *Am. J. Neuroradiol.* **41**, 416–423 (2020).
34. N. Bahrami et al., "Molecular classification of patients with grade II/III glioma using quantitative MRI characteristics," *J. Neuro Oncol.* **139**, 633–642 (2018).

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

**Fang F. Yu**, MD, is an assistant professor of radiology at UT Southwestern Medical Center and Advanced Imaging Research Center. He is focused on the development and clinical translation of advanced neuroimaging techniques toward the goal of *in vivo* histology.

**Chandan Ganesh Bangalore Yogananda**, PhD, is a neuroimaging systems engineer at UT Southwestern Medical Center. He received his PhD in biomedical engineering from the University of Texas at Arlington in 2021.

**Bhavya R. Shah**, MD, is an assistant professor of neuroradiology and neurological surgery. He is director of the transcranial focused ultrasound program and co-director of the focused ultrasound lab at UT Southwestern Medical Center. His lab is focused on developing and improving the use of focused.

**Bruce Mickey** served on the neurosurgery faculty at UT Southwestern Medical Center from 1984 through 2021, supporting the institution's brain tumor research effort, including multiple projects involving glioma metabolism and migration, and the detection of 2-hydroxyglutarate using MR spectroscopy in patients with IDH mutated gliomas. As a professor emeritus of neurosurgery, he remains involved in several projects, including the investigation of the role of deep learning in the interpretation of MR images of gliomas.

**Toral R. Patel** is an associate professor of neurological surgery at UT Southwestern. She is a neurosurgical oncologist by training; her clinical practice is focused on the surgical management of adult brain tumor patients. Academically, she is interested in drug delivery and advanced imaging techniques for malignant gliomas.

**Baowei Fei** is a professor of bioengineering and Cecil H. and Ida Green Chair in Systems Biology Science at the University of Texas at Dallas. He is also a professor of radiology at UT Southwestern Medical Center. He is director of the Quantitative Bioimaging Laboratory ([www.fei-lab.org](http://www.fei-lab.org)) and director of the Center for Imaging and Surgical Innovation. He is a fellow of the International Society for Optics and Photonics (SPIE) and a fellow of the American Institute for Medical and Biological Engineering (AIMBE).



**Ananth J. Madhuranthakam**, PhD, is an associate professor of radiology, Advanced Imaging Research Center, and biomedical engineering at UT Southwestern Medical Center. He is a director of magnetic resonance (MR) research and his lab ([utsouthwestern.edu/labs/madhuranthakam/](https://utsouthwestern.edu/labs/madhuranthakam/)) is focused on the technical development and clinical translation of advanced MR imaging techniques.

**Joseph A. Maldjian**, MD, is a professor of radiology at UT Southwestern Medical Center and chief of neuroradiology, and he 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 in a variety of applications and ultrasound therapies for diseases of the central nervous system.

Biographies of the other authors are not available.