Learning-based analysis of amide proton transfer-weighted MRI to identify true progression in glioma patients

Pengfei Guo\textsuperscript{a, b, 1}, Mathias Unberath\textsuperscript{b, 1}, Hye-Young Heo\textsuperscript{a}, Charles G. Eberhart\textsuperscript{c}, Michael Lim\textsuperscript{d}, Jaishri O. Blakeley\textsuperscript{a, \*}, Shanshan Jiang\textsuperscript{b, 1}

\textsuperscript{a} Department of Radiology, Johns Hopkins University, Baltimore, MD, USA
\textsuperscript{b} Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA
\textsuperscript{c} Department of Pathology, Johns Hopkins University, Baltimore, MD, USA
\textsuperscript{d} Department of Neurosurgery, Johns Hopkins University, Baltimore, MD, USA
\textsuperscript{\*} Corresponding author at: Department of Radiology, Johns Hopkins University, 600 N. Wolfe Street, Park 332, Baltimore, MD 21287, USA.
\textsuperscript{1} Contributed equally to this work.

E-mail address: sjiang21@jhmi.edu (S. Jiang).

ARTICLE INFO
Keywords:
Amide proton transfer-weighted MRI
Convolutional neural network
Glioma
Treatment effect

ABSTRACT
The purpose of this study was to develop and verify a convolutional neural network (CNN)-based deep-learning algorithm to identify tumor progression versus response by adding amide proton transfer-weighted (APTw) MRI data to structural MR images as the proposed model input. 145 scans with 2175 MR instances from 98 patients with malignant glioma (acquired between April 2010 and February 2018) were re-analyzed. An end-to-end classification framework based on a ResNet backbone was developed. The architecture includes a learnable subtraction layer and a hierarchical classification paradigm, and synthesizes information over multiple MR slices using a long short-term memory. Areas under the receiver-operating-characteristic curves (AUCs) were used to assess the impact of adding APTw MRI to structural MRI (T1w, T2w, FLAIR, and GdT1w) on classification of tumor response vs. progression, both on the slice- and scan-level. With both APTw and structural MRI data, adding a learnable subtraction layer and a hierarchical classification paradigm to the backbone ResNet model improved the slice-level classification performance from an AUC of 0.85 to 0.90. Adding APTw data to structural MR images as input to our proposed CNN classification framework led to an increase in AUCs from 0.88 to 0.90 for the slice-level classification (P < 0.001), and from 0.85 to 0.90 for the scan-level classification (P < 0.05). Generated saliency maps highlighted the vast majority of lesions. Complementing structural MRI sequences with protein-based APTw MRI enhanced CNN-based classification of recurrent glioma at the slice and scan levels. Addition of APTw MRI to structural MRI sequences enhanced CNN-based classification of recurrent glioma at the slice and scan levels.

1. Introduction

Despite maximum feasible surgical resection followed by radiotherapy with concurrent chemotherapy, malignant gliomas eventually progress, with a median survival of 12–15 months for glioblastoma (Stupp et al., 2005). Radiographic evaluation plays a critical role in the management of post-treatment malignant gliomas, in which magnetic resonance imaging (MRI) following the response-assessment-in-neuro-oncology (RANO) criteria remains the standard (Wen et al., 2010). These limitations have immediate clinical consequences confounding post-treatment diagnostics and treatment planning and complicating the procedures for new therapy development. Therefore, reliable, automated imaging diagnostic tools to assess malignant glioma response to therapies are urgently needed.

Amide proton transfer-weighted (APTw) imaging, based on chemical exchange saturation transfer (CEST) MRI contrast mechanism (Ward et al., 2000), is an emerging molecular MRI technique that was designed to detect endogenous cellular proteins and peptides in tissue (Zhou et al., 2010). Despite maximum feasible surgical resection followed by radiotherapy with concurrent chemotherapy, malignant gliomas eventually progress, with a median survival of 12–15 months for glioblastoma (Stupp et al., 2005). Radiographic evaluation plays a critical role in the management of post-treatment malignant gliomas, in which magnetic resonance imaging (MRI) following the response-assessment-in-neuro-oncology (RANO) criteria remains the standard (Wen et al., 2010). However, structural MR images used in the clinical setting, including T1-weighted (T1w), T2-weighted (T2w), fluid-attenuated inversion recovery (FLAIR), and gadolinium-enhanced T1w (GdT1w) MR images, are not sufficiently tissue-specific to guide treatment decisions (Wen et al., 2010). These limitations have immediate clinical consequences confounding post-treatment diagnostics and treatment planning and complicating the procedures for new therapy development. Therefore, reliable, automated imaging diagnostic tools to assess malignant glioma response to therapies are urgently needed.

Amide proton transfer-weighted (APTw) imaging, based on chemical exchange saturation transfer (CEST) MRI contrast mechanism (Ward et al., 2000), is an emerging molecular MRI technique that was designed to detect endogenous cellular proteins and peptides in tissue (Zhou et al., 2010). Despite maximum feasible surgical resection followed by radiotherapy with concurrent chemotherapy, malignant gliomas eventually progress, with a median survival of 12–15 months for glioblastoma (Stupp et al., 2005). Radiographic evaluation plays a critical role in the management of post-treatment malignant gliomas, in which magnetic resonance imaging (MRI) following the response-assessment-in-neuro-oncology (RANO) criteria remains the standard (Wen et al., 2010). However, structural MR images used in the clinical setting, including T1-weighted (T1w), T2-weighted (T2w), fluid-attenuated inversion recovery (FLAIR), and gadolinium-enhanced T1w (GdT1w) MR images, are not sufficiently tissue-specific to guide treatment decisions (Wen et al., 2010). These limitations have immediate clinical consequences confounding post-treatment diagnostics and treatment planning and complicating the procedures for new therapy development. Therefore, reliable, automated imaging diagnostic tools to assess malignant glioma response to therapies are urgently needed.

Amide proton transfer-weighted (APTw) imaging, based on chemical exchange saturation transfer (CEST) MRI contrast mechanism (Ward et al., 2000), is an emerging molecular MRI technique that was designed to detect endogenous cellular proteins and peptides in tissue (Zhou et al., 2010). Despite maximum feasible surgical resection followed by radiotherapy with concurrent chemotherapy, malignant gliomas eventually progress, with a median survival of 12–15 months for glioblastoma (Stupp et al., 2005). Radiographic evaluation plays a critical role in the management of post-treatment malignant gliomas, in which magnetic resonance imaging (MRI) following the response-assessment-in-neuro-oncology (RANO) criteria remains the standard (Wen et al., 2010). However, structural MR images used in the clinical setting, including T1-weighted (T1w), T2-weighted (T2w), fluid-attenuated inversion recovery (FLAIR), and gadolinium-enhanced T1w (GdT1w) MR images, are not sufficiently tissue-specific to guide treatment decisions (Wen et al., 2010). These limitations have immediate clinical consequences confounding post-treatment diagnostics and treatment planning and complicating the procedures for new therapy development. Therefore, reliable, automated imaging diagnostic tools to assess malignant glioma response to therapies are urgently needed.

Amide proton transfer-weighted (APTw) imaging, based on chemical exchange saturation transfer (CEST) MRI contrast mechanism (Ward et al., 2000), is an emerging molecular MRI technique that was designed to detect endogenous cellular proteins and peptides in tissue (Zhou et al., 2010). Despite maximum feasible surgical resection followed by radiotherapy with concurrent chemotherapy, malignant gliomas eventually progress, with a median survival of 12–15 months for glioblastoma (Stupp et al., 2005). Radiographic evaluation plays a critical role in the management of post-treatment malignant gliomas, in which magnetic resonance imaging (MRI) following the response-assessment-in-neuro-oncology (RANO) criteria remains the standard (Wen et al., 2010). However, structural MR images used in the clinical setting, including T1-weighted (T1w), T2-weighted (T2w), fluid-attenuated inversion recovery (FLAIR), and gadolinium-enhanced T1w (GdT1w) MR images, are not sufficiently tissue-specific to guide treatment decisions (Wen et al., 2010). These limitations have immediate clinical consequences confounding post-treatment diagnostics and treatment planning and complicating the procedures for new therapy development. Therefore, reliable, automated imaging diagnostic tools to assess malignant glioma response to therapies are urgently needed.

Amide proton transfer-weighted (APTw) imaging, based on chemical exchange saturation transfer (CEST) MRI contrast mechanism (Ward et al., 2000), is an emerging molecular MRI technique that was designed to detect endogenous cellular proteins and peptides in tissue (Zhou et al., 2010). Despite maximum feasible surgical resection followed by radiotherapy with concurrent chemotherapy, malignant gliomas eventually progress, with a median survival of 12–15 months for glioblastoma (Stupp et al., 2005). Radiographic evaluation plays a critical role in the management of post-treatment malignant gliomas, in which magnetic resonance imaging (MRI) following the response-assessment-in-neuro-oncology (RANO) criteria remains the standard (Wen et al., 2010). However, structural MR images used in the clinical setting, including T1-weighted (T1w), T2-weighted (T2w), fluid-attenuated inversion recovery (FLAIR), and gadolinium-enhanced T1w (GdT1w) MR images, are not sufficiently tissue-specific to guide treatment decisions (Wen et al., 2010). These limitations have immediate clinical consequences confounding post-treatment diagnostics and treatment planning and complicating the procedures for new therapy development. Therefore, reliable, automated imaging diagnostic tools to assess malignant glioma response to therapies are urgently needed.
2019). The APTw hyperintensity observed in malignant glioma is most likely associated with increased cytosolic protein content due to higher cellularity and slightly increased intracellular pH (Lee et al., 2017; Ray et al., 2019; Yan et al., 2015). Numerous research groups worldwide have confirmed that the hypertensity on APTw images is a reliable imaging marker of malignant glioma before (Jiang et al., 2017; Jiang et al., 2022; Sotiriou et al., 2020) and after (Liu et al., 2020; Park et al., 2020a; Park et al., 2020b) treatment. Notably, consensus recommendations on clinical APTw imaging approaches at 3 T for brain tumors have recently been published (Zhou et al., 2022).

On the other hand, advances in artificial intelligence and computer vision have achieved powerful solutions for improving medical imaging techniques and automatic diagnosis (Elglaera et al., 2020; Hu et al., 2021; Verma et al., 2020), including CEST MRI (Cohen et al., 2018; Glang et al., 2020; Goldenberg et al., 2019). Convolutional neural networks (CNNs) have recently been applied successfully to neuro-oncological imaging (Buda et al., 2020; Chang et al., 2018; Choi et al., 2019). However, the number of studies on post-treatment image analysis to predict true progression for patients with malignant gliomas (Bacchi et al., 2019; Li et al., 2020) is still limited. This study aims to develop and verify a CNN-based deep-learning algorithm to identify tumor recurrence through a cross-sectional, multimodal MRI exam. Our CNN analysis results demonstrate that adding protein-based APTw MRI to traditional structural MR images can significantly increase the accuracy of treatment response assessment compared to using traditional MR images only.

2. Materials and methods

2.1. Patient Enrollment and annotation

This study is a secondary analysis of previously collected data, and part of the data used here have been published (Guo et al., 2021; Jiang et al., 2019; Ma et al., 2016). This study, based on the de-identified data, was approved by the Institutional Review Board (IRB) and the need for consent for this de-identified data reanalysis was waived. Patient inclusion criteria were as follows: ≥20 years old; diagnosis of WHO grade III or IV glioma; status-post initial surgery and chemoradiation or radiotherapy alone; suspected tumor recurrence and completed APTw imaging (in addition to structural MRI sequences) study after the completion of therapy; and an integrated clinical diagnosis of tumor recurrence or treatment effect.

Each lesion was annotated as “response to treatment” (including complete response, partial response, stable disease, radiation necrosis, and pseudoprogression) or “progressive disease” (including progression and pseudoresponse), according to the RANO criteria for two-dimensional (2D) images for each instance (Vogelbaum et al., 2012; Wen et al., 2010). Here, one instance indicates a set of APTw, T2w, T2w, FLAIR, and GdT1w MR images acquired at the same slice level. For the scan-level classification, scans with one or more slices of “progressive disease” were assigned as “progression”, and all other scans assigned as “response”. Notably, if patients underwent surgery within four weeks after the observed APTw-MRI scan, histopathologic diagnosis took priority over the longitudinal MRI analysis.

2.2. MRI data collection

APTw images were obtained on a 3 T human MRI scanner (Achieva; Philips Medical Systems) using body coil excitation and a 32-channel phased-array coil for reception. Three-dimensional (3D) APTw imaging was based on a previously published sequence (Jiang et al., 2019; Ma et al., 2016), with the following imaging parameters: radiofrequency saturation duration, 830 ms; saturation power, 2 µT; field of view (FOV), 212 × 186 × 66 mm³; resolution, 0.82 × 0.82 × 4.4 mm³ (reconstructed); and matrix, 256 × 256 × 15 (reconstructed). T2w was acquired with imaging parameters: TR, 4 sec; echo time (TE), 80 ms; 60 slices; thickness, 2.2 mm, and FLAIR was acquired with imaging parameters: TR, 11 sec; TE, 120 ms; inversion recovery time, 2.8 s; 60 slices; thickness, 2.2 mm. T1w and GdT1w images were acquired with the following parameters: 3D magnetization-prepared-rapid-gradient-echo sequence; TR, 3 s; TE, 3.7 ms; inversion recovery time, 843 ms; flip angle, 8; 150 slices; isotropic voxel, 1.1 mm³), and the dose of Gd contrast agents was 0.2 mL/kg body weight. The anatomic MRI sequences (T1w, T2w, FLAIR, and GdT1w) had the image parameters: FOV, 212 × 172 × 165 (or 212 × 189 × 132) mm³; resolution, 0.41 × 0.41 × 1.1 mm³ (reconstructed); and matrix, 512 × 512 × 150 (reconstructed). For each scan, due to the fact that the 3D APTw MRI protocol provided 15 slices, volumetric MR images used 15 instances. Each instance included T1w, T2w, FLAIR, GdT1w, and APTw images with the matrix shape of 5 (sequences) × 256 (pixels) × 256 (pixels). Instances were the input of proposed slice-level feature extractor CNN.

2.3. Data preprocessing

Data preprocessing steps, including co-registration (Lowekamp et al., 2013), skull-stripping (Lipkova et al., 2019), N4-bias field correction (Tustison et al., 2010), and MRI standardization (Nyúl et al., 2000), were performed sequentially. Notably, based on our experience during the image preprocessing, in order to preserve the distinguishing radiographic patterns on APTw-MRI, MRI scale standardization was not performed on APTw-MRI. We used a rigid-body registration for the co-registration across the APTw images and the anatomical MR images (T1w, T2w, FLAIR, and GdT1w) (Zhang et al., 2016). This was performed through the saturated images at 3.5 ppm to reconstruct the APTw images, which share the same spatial information with the APTw images. Preprocessing was performed by a medical imaging engineer and supervised by a radiologist who also verified the image preprocessing outputs. Lesions of post-treatment malignant glioma that cover the regions of abnormal intensities on multiparameter MR images were segmented on the co-registered FLAIR MR images. Then, the pre-surgery and all clinical follow-up MR images, together with all clinical reports in the electronic medical record system, were reviewed serially for annotation.

2.4. Deep-learning classification pipeline

The classification framework consisted of two main stages: slice-level classification and scan-level classification (Fig. 1a). During the slice-level classification, a CNN using three concatenated residual learning blocks (ResNet-18) (He et al., 2016) served as a feature extractor and the backbone model in a previous study. This backbone architecture without any modification was denoted as the standard model. As explained below, we introduced a learnable subtraction module (LS) and a hierarchical classification (HC) paradigm as modifications to the standard ResNet-18 architecture. To aggregate predictions across all slices and obtain a scan-level prediction, a long short-term memory (LSTM) module was added that sequentially processed all embedded feature representations of all slices. The detailed descriptions of each module are presented in the following sub-sections.

2.5. Slice-level feature extractor CNN

The classification consists of three hierarchical binary-classification sub-tasks (Fig. 1b). The ResNet-18 architecture was modified according to this hierarchy by inserting classification branches for three binary classification tasks above at increasing depths into the network. This procedure increases gradient flow as additional gradients are injected during back-propagation, and promotes generalizable learning since the extracted features must inform several related tasks rather than only one. Binary cross entropy loss was adopted for each branch and is defined as follows:

\[
L = -\sum_{i=1}^{N} y_i \log(p_i) + (1 - y_i) \log(1 - p_i)
\]
The number in parentheses is the number of instances reported for the entire dataset.

\[
L_{BCE}(x, y) = -(y \log(x) + (1 - y) \log(1 - x))
\]

where \(x\) is the predicted probability and \(y\) is the binary indicator (0 or 1) for the target label. The loss function of a CNN is a weighted summation of all branches, \(L_{BCE}\), and is defined as follows:

\[
L_{CNN} = \sum_{d} \alpha_{d} L_{BCE}^{d}(x, y)
\]

where \(D\) is the number of branches, \(L_{BCE}^{d}\) is the binary cross entropy loss of the corresponding branch, and the weight \(\alpha_{d}\) is used to control the relative importance of each branch.

The most accurate slice-level CNN classification framework was used to perform the sanity check.

### 2.6. Learnable subtraction module and long short-term memory

Informed by the radiologic reading workflow for such images, we proposed an LS module on top of the CNN to perform a learnable-parameter-adjusted image subtraction (Fig. 2a). The LS module was calculated between GdT\(_{1}\)w and T\(_{1}\)w, as well as between T\(_{2}\)w and FLAIR images for image comparison. An LSTM was proposed to obtain all extracted, slice-based features from the same scan as the sequential input for scan-level classification.

### 2.7. Long short-term memory

The proposed LSTM takes all extracted slice-level features from a scan as a sequential input to perform scan-level diagnosis. For each element of the input sequence, LSTM was computed as follows:

\[
i_t = \sigma(W_{wx}x_t + b_w + W_{wh}h_{t-1} + b_h)\\
f_t = \sigma(W_{xf}x_t + b_f + W_{wh}h_{t-1} + b_h)\\
g_t = \tanh(W_{wg}x_t + b_g + W_{wh}h_{t-1} + b_w)\\
o_t = \sigma(W_{wo}x_t + b_o + W_{wh}h_{t-1} + b_w)\\
c_t = f_t \odot c_{t-1} + i_t \odot g_t\\
h_t = o_t \odot \tanh(c_t)
\]

where \(x_t\), \(c_t\), and \(h_t\) are the input, cell state, and hidden state for slice \(t\), respectively. \(\sigma\) is the sigmoid function and \(\tanh\) is the hyperbolic tangent function. \(\odot\) is the Hadamard product. \(i_t, f_t, g_t,\) and \(o_t\) are the input, forget, cell, and output gates, respectively. We denoted the output features of the last layer of the LSTM as \(H\). Then, \(H\) was passed through a fully connected layer to produce the final scan-level prediction based on all extracted slice-level features of a single scan from the slice-level CNN. Thus, the dependence between input instances and scan-level prediction in our task was modeled.

### 2.8. Training and implementation details

We adopted the binary cross entropy loss and the Adam (Kingma and Ba, 2015) optimizer to minimize the loss function, with an exponential decay rate \(\beta = (0.9, 0.999)\) for both CNN and LSTM. The initial learning rate was set to \(10^{-4}\) and \(10^{-2}\) for training CNN and LSTM, respectively. The learning rate of first five epochs was an initial learning rate \(\times 0.1\) and constant in the first 50 epochs, and then linearly decayed to zero in the last 50 epochs, with a warm-start, learning-rate scheduler. The batch size was set to 15 and 8 for training CNN and LSTM, respectively. The importance parameter \(\alpha_d\) was set to 1 for all branches. The inference time for the proposed CNN and LSTM was 0.018 s per instance and 0.001 s per scan, respectively. We implemented the proposed approach on an Ubuntu 18.04 computer using an NVIDIA 2080Ti GPU and PyTorch. We utilized the class activation map (CAM) to expose the attention of the CNN on the input slices, which highlights the most informative image regions relevant to the predicted class (Zhou et al., 2016). High-response regions with a salience score of higher than the 95th percentile value were further generated, which were denoted by CAM*.

### 2.9. Statistical analysis and model evaluation

Scan-based data was split into 70% training, 10% validation, and 20% testing according to the chronological order of the MRI scan date (Table 1). The rationale of chronological order data split is to simulate the real-world clinical practice. Our data split reflects the attempt of attaining unbiased performance estimates in prospective deployment. Notably, for the scan-level analysis, all instances from the same scan were treated as an individual sample to prevent data sharing across the split datasets, and the scans from the same patient were not shared.
between training and testing datasets. The diagnostic performances of the proposed methods, including sensitivity, specificity, and the area under the receiver-operating-characteristic (ROC) curve (AUC), were measured on the testing dataset. Notably, to explicitly evaluate the incremental diagnostic impact of APTw MRI, the ROC curves were compared between inputs with and without APTw (DeLong et al., 1988). We ran all experiments three times and reported the mean metrics. Data analysis was performed with the Python scikit-learn package (Pedregosa et al., 2011).

3. Results

3.1. Patient demographic information

A total of 145 scans obtained from 98 patients between April 2010 and February 2018 were included in this study. Patient demographic information and basic lesion characteristics are shown in Table 1. Based on the integrated clinical pathologic results, 86 scans were classified as “progression”, while the remaining 59 scans were classified as “response”. Based on the slice-level features, this led to 2175 instances in total, 742 of which were grouped as “progressive disease”.

3.2. Slice-level diagnostic performance

The slice-level classification performance was first evaluated in the standard CNN model, where the contribution of using GdT1w and APTw as a part of the input was investigated (Table 2). In the slice-level backbone CNN model using T1w, T2w, and FLAIR MRI data as the baseline input, the AUC for distinguishing progressive disease from non-progression was 0.77 (CI, 0.70–0.81). Adding GdT1w or APTw MRI alone to the baseline input, the AUCs were increased to 0.82 (CI, 0.79–0.87) or 0.84 (CI, 0.81–0.89), respectively. Adding GdT1w and APTw MRI jointly achieved the highest AUC, with 0.85 (CI, 0.82–0.89).

Fig. 2. Visual illustration of the concept of the learnable subtraction module. The schematics of the learnable subtraction in a pixel-wise dot product (a) and an example of visualization of image subtraction (b). * denotes the dot product. The learnable parameters for this pixel-wise operation were implemented by a 1×1 convolutional layer with a kernel size of 1 using predefined initializing parameters. Notably, in order to avoid trivial solutions during training, instead of initializing with zero, we set those zero values to a small number (i.e., 1e-3).
Comparisons of performances with different MRI sequence data as input for the slice-level classification in the backbone model.

| Data Input | Diagnostic Performances |
|------------|-------------------------|
|            | AUC (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) | P value* |
| APTw       | 0.77 (0.70, 0.81) | 0.61 (0.53, 0.69) | 0.81 (0.77, 0.85) | - |
| GdT\textsubscript{w} | 0.82 (0.79, 0.87) | 0.82 (0.75, 0.88) | 0.75 (0.70, 0.80) | P < 0.001 |
| T\textsubscript{1w} | 0.84 (0.81, 0.89) | 0.87 (0.81, 0.92) | 0.68 (0.62, 0.73) | P < 0.001 |
| T\textsubscript{2w} | 0.85 (0.82, 0.89) | 0.92 (0.87, 0.96) | 0.63 (0.58, 0.69) | P < 0.001 |
| FLAIR      | 0.82 (0.79, 0.87) | 0.82 (0.75, 0.88) | 0.75 (0.70, 0.80) | - |

Note. √ indicates that the input instance contains the corresponding MR sequence.

* P value for the ROC curve comparison with respect to the input of T\textsubscript{1w}, T\textsubscript{2w} and FLAIR data.
estimates in prospective deployment. However, with this data split method, the clinical and demographic characteristics could barely be controlled to evenly distribute among data splits, such as gender ratio or first progression v.s. in subsequent therapy this study.

In this study, pairs of models using data with and without APTw MRI were compared to explore the incremental value of APTw MRI to structural MRIs. Our data show that APTw MRI improves the diagnostic performance for both slice-level and scan-level classifications. This is consistent with previous studies that demonstrated the value of APTw MRI for post-treatment malignant gliomas (Zhou et al., 2019). Furthermore, CAMs were calculated from the slice-level CNN. As a visualization and attribution method with which to elucidate CNNs, CAM explanations correspond to the gradient of the class score (logit) with respect to the feature map of the last convolutional unit of a CNN (Selvaraju et al., 2016). This mapping uses the gradients of any target concept flowing into the final convolutional layer to produce a coarse localization map, that highlights the important regions in the image for predicting the concept. CAM is capable of generating images that highlight salient regions, arguably acting as a localizer for important regions that contain highly discriminative information, with great promise for clinical translation (Fong and Vedaldi, 2017; Sundararajan et al., 2017). The CAMs, especially the CAM*s of our study, substantially localized the majority of regions of abnormal MRI signals (Fig. 4). The results indicate that the proposed methods can successfully extract truly and biologically relevant distinguishable information from multi-modality MRIs.

There are several limitations to this study. First, the proposed models were trained and tested on single-center data. Our next work will incorporate data from multiple external institutions to create a generalizable algorithm. Second, for the scan-level prediction, the proposed method was developed to use the embedded 2D-CNN-features. An alternative is to directly use volumetric data to build 3D CNN algorithms. However, there was a discrepancy in resolution along the z-axis between APTw images and the structural MR sequences (4.4 mm vs. 1.1 mm). Consequently, a dramatically compromised fidelity for APTw images due to resampling for an isotropic 3D volume, impedes a 3D CNN. Third, several critical genetic markers, such as the status of isocitrate dehydrogenase (IDH) mutation, 1p19q codeletion and (O-6-methylguanine-DNA methyltransferase (MGMT) methylation, were not included in the CNN due to limited, assessable genetic data. Only with these additional data, the added value of APTw imaging in these definite glioma patient subsets can be established. We will collect and analyze genetic profiles in our ongoing prospective study. Forth, perfusion MRI, an advanced MR imaging with great diagnostic value for post-treatment glioma patients, was not included in the analysis of this study. The lack of any analysis with perfusion and APTw images left an unexplored question for the further study. Last but not least, malignant gliomas infiltrated throughout the whole brain. However, APTw MRI only covered up to 66 mm in the z direction (4.4 mm × 15 slices) due to the technique limitation. This led to the fact that a whole brain comparison was not performed in this study.

5. Conclusion

We propose a deep learning-based pipeline to identify true tumor progression versus treatment effects after radiation for patients with malignant glioma utilizing multiparameter MRIs. The proposed learnable subtraction layer shows promise, which indicates the improvement
on the data usage plays an important role in increasing automated analysis outcome. The boosted performance achieved by supervising the natural hierarchy of general-to-specific order under targets class demonstrates that the proposed hierarchical classification paradigm can provide prior for deeper layers as a good guide during the training. The AUCs of our best-performing models (0.90 for both slice-level and scan-level models) verifies our motivation that complementing structural with functional APTw MRI can further improve the diagnostic performance. Based on this performance, the proposed method could be a highly efficient solution that could help clinical experts to make precise diagnoses for patients with post-treatment malignant gliomas.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported in part by grants from the National Institutes of Health (R37CA248077, R01CA228188, and P4103177). The authors thank Ms. Lindsay Blair for patient coordination and Ms. Mary McAllister for editorial assistance.

References

Bacchi, S., Zerner, T., Dongas, J., Asahina, A.T., Abou-Hamden, A., Otto, S., Oakden-ardenas-Rodríguez, J., Pagel, M.D., 2019. Machine learning improves treatment response in glioblastoma patients. J Neuro-Oncol 141, 1325-1336.

Choi, H., Chei, S.H., 2019. Prediction of IDH1 genotype in gliomas with dynamic susceptibility contrast perfusion MRI imaging using an explainable recurrent neural network. Neuro-oncology 21, 1297-1308.

Cohen, O., Huang, S., Rosen, M., Wiederhold, C., 2020. Robust Deep Learning-based Segmentation of Glioblastoma on Routine Clinical MRI Scans Using Sparser Training. Radiol. Res. 2 (5), e190133.

Eiselt, S.C., Wu, Y., Lee, E.O., 2016. Assessment of brain tumor response: RANO and its offspring. Curr. Treat. Options Oncol. 17, 35.

Fong, R.C., Vedaldi, A., 2017. Interpretable explanations of black boxes by meaningful perturbation. Proceedings of the 7th International Conference on Computer Vision, pp. 3429-3437.

Gläng, F., Deshmane, A., Prokudin, S., Martin, F., Herz, K., Lindig, T., Bender, B., Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput 9 (8), 1123-1133.

Hamel, P.C., de Munck, J.C., Witte, M.G., 2020. Robust Deep Learning-based Approach for Predicting Tumor Treatment Response. Biometrics 44, 837-849.

Kiesel, B., Widhalm, G., Furtner, J., Robe, P.A.J.T., Mandonnet, E., De Witt Longueville, A., Hervey-Jumper, S., Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825-2830.

Ray, K.J., Simard, M.A., Larkin, J.R., Coates, J., Kincsh, P., Smart, S.C., Higgins, G.S., Chappell, M.A., Sibson, N.R., 2019. Tumor pH and protein concentration contribute to the signal of amide proton transfer magnetic resonance imaging. Cancer Research 79, 143-152.

Selvaraju, R.R., Das, A., Vedantam, R., Cogswell, M., Parikh, D., Batra, D., 2016. Grad-CAM: Why did you say that? arXiv preprint arXiv:1611.07450.

Sotirios, B., Demetriou, E., Topricic, C.C., Zakrzewska, Z., 2020. The role of APT imaging in gliomas grading: A systematic review and meta-analysis. Eur. J. Radiol. 133, 109353.

Stupp, R., Mason, W.P., van den Bent, M.J., Weller, M., Fisher, B., Taphoorn, M.J.B., Belanger, K., Brandes, A.W., Bogdahn, U., De完好, E., De Witt Longueville, A., Hervey-Jumper, S., Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825-2830.

Wolf, R.L., 2019. MRI of Recurrent Glioblastoma: Reliability and Reality. Radiology 290, 133-141.

Wen, P.Y., Macdonald, D.R., Van den Bent, M.J., Weller, M., Fisher, B., Taphoorn, M.J.B., Belanger, K., Brandes, A.W., Bogdahn, U., De完好, E., De Witt Longueville, A., Hervey-Jumper, S., Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12, 2825-2830.

Ward, K.M., Aletas, A.H., Baldan, R.S., 2000. A new class of contrast agents for MRI based on proton chemical exchange dependent saturation transfer (CEST). J. Magn. Reson. 143 (1), 79-87.

Wen, P.Y., Macdonald, D.R., Reardon, D.A., Cloughesy, T.F., Sorensen, A.G., Galanis, E., DeGroot, J., Wick, W., Gilbert, M.R., Lassman, A.B., Tsien, C., Mikkelsen, T., DeHart, J.C., Yin, B., de Munck, J.C., Witte, M.G., 2020. Tumor Habitat-derived Radiomic Features at Pretreatment MRI That Are Prognostic for Progression-free Survival in Glioblastoma Are Associated with Key Morphologic Attributes at Histopathologic Examination: A Feasibility Study. Radiol. Artif Intell 2 (6), e190168.

Vogelbaum, M.A., Jost, S., Aghi, M.K., Heimberger, A.B., Sampson, J.H., Wen, P.Y., Macdonald, D.R., Van den Bent, M.J., Chang, S.M., 2012. Application of novel response/progression measures for surgically delivered therapies for gliomas: Response assessment in neuro-oncology (RANO) working group. Neurosurgery 70, 1320-1328.
origins in 9L gliosarcoma in the rat brain using proteomic analysis. Mol. Imaging Biol. 17, 479–487.
Zhang, Y.I., Heo, H.-Y., Lee, D.-H., Zhao, X., Jiang, S., Zhang, K., Li, H., Zhou, J., 2016. Selecting the reference image for registration of CEST series. Journal of Magnetic Resonance Imaging 43 (3), 756–761.
Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A. 2016. Learning deep features for discriminative localization. Proceedings of the IEEE conference on computer vision and pattern recognition. 2921-2929.
Zhou, J., Heo, H.-Y., Knutsson, L., van Zijl, P.C.M., Jiang, S., 2019. APT-weighted MRI: Techniques, current neuro applications, and challenging issues. J. Magn. Reson. Imaging 50 (2), 347–364.
Zhou, J., Zaiss, M., Knutsson, L., Sun, P.Z., Ahn, S.S., Aime, S., Bachert, P., Blakeley, J.O., Cai, K., Chappell, M.A., Chen, M., Gochberg, D.F., Goerke, S., Heo, H.-Y., Jiang, S., Jin, T., Kim, S.-G., Laterza, J., Pesch, D., Pogel, M.D., Park, J.E., Reddy, B., Sakata, A., Sartoretti-Schefer, S., Sherry, A.D., Smith, S.A., Stanisz, G.J., Sundgren, P. C., Togao, O., Vandsburger, M., Wen, Z., Wu, Y., Zhang, Y., Zhu, W., Zu, Z., van Zijl, P.C.M., 2022. Review and consensus recommendations on clinical APT-weighted imaging approaches at 3T: Application to brain tumors. Magn. Reson. Med. https://doi.org/10.1002/mrm.29241.
Zhu, X., Bain, M., 2017. B-CNN: branch convolutional neural network for hierarchical classification. arXiv preprint arXiv:1709.09890.