Low-parameter supervised learning models can discriminate pseudoprogression and true progression in non-perfusion-based MRI

Elisa Warner[1], Joonsang Lee[2], Santhoshi Krishnan[1,3], Nicholas Wang[1], Shariq Mohammed[4], Ashok Srinivasan[5], Jayapalli Bapuraj[5], and Arvind Rao[3]

***Abstract*— Discrimination of pseudoprogression and true progression is one challenge to the treatment of malignant gliomas. Although some techniques such as circulating tumor DNA (ctDNA) and perfusion-weighted imaging (PWI) demonstrate promise in distinguishing PsP from TP, we investigate robust and replicable alternatives to distinguish the two entities based on more widely-available media. In this study, we use low-parametric supervised learning techniques based on geographically-weighted regression (GWR) to investigate the utility of both conventional MRI sequences as well as a diffusion-weighted sequence (apparent diffusion coefficient or ADC) in the discrimination of PsP v TP. GWR applied to MRI modality pairs is a unique approach for small sample sizes and is a novel approach in this arena. From our analysis, all modality pairs involving ADC maps, and those involving post-contrast T1 regressed onto T2 showed potential promise. This work on ADC data adds to a growing body of research suggesting the predictive benefits of ADC, and suggests further research on the relationships between post-contrast T1 and T2.**

***Clinical relevance*— Few studies have investigated predictive potential of conventional MRI and ADC to detect PsP. Our study adds to the growing research on the topic and presents a new perspective to research by exploiting the utility of ADC in PsP v TP distinction. In addition, our GWR methodology for low-parametric supervised computer vision models demonstrates a unique approach for image processing of small sample sizes.**

I. INTRODUCTION

Glioma is a life-threatening condition characterized by neoplastic tumor growth in the brain. Malignant gliomas are primarily diagnosed based on the imaging features on conventional MRI sequences. The primary approach to treatment of malignant gliomas is surgical resection within the limits of patients' safety, followed by radiation and temezolomide for six weeks. This treatment cycle is then followed by adjuvant temozolomide therapy for six more weeks. Because of the infiltrative nature of the tumors, surgical treatment is usually never curative since the tumors can extend well beyond the margins demonstrated by MR imaging. Therefore, follow-up imaging often shows tumor recurrence. True tumor recurrence needs to be differentiated from the condition of pseudoprogression which is defined as the appearance of a new lesion or increase in constrast-enhancing areas, but with changes that gradually fade or stabilize while treatment stays the same [1]. Since PsP resembles true progression (TP), treatment needs to be instituted at the earliest for the latter while waiting and watching is the approach to the former. Additionally, because of the similarity of these two conditions, patients are at risk of being diagnosed with alternative treatments prematurely or erroneously withdrawn from treatment altogether [2], [3]. Thus, there is a need to distinguish PsP from TP.

Previous studies have investigated possible methods for distinguishing PsP from TP using bloodstream biomarkers or medical imaging. In one method, evaluation of chromosomal instability using circulating tumor DNA (ctDNA) reported promising results in PsP v TP distinction in a small trial group [4]. However, further analysis on larger cohorts is still needed, and ctDNA extraction may not be available to all clinicians. Another common marker is perfusion weighted imaging (PWI), which is current practice in some clinics [3]. However, this advanced imaging technique requires use of advanced software that are not widely available for all radiologists.

In the absence of perfusion imaging, some attempts have been made to exploit classical sequences of magnetic resonance imaging (MRI) in PsP studies using visual characteristics of the images determined by subject matter experts [5], [6], often finding these sequences to confer little help. A supervised machine learning approach using convolutional neural networks applied to conventional MRI sequences was able to exploit multimodal relationships between these MRI sequences for PsP distinction [7]. However, there are limitations to large multi-parametric models with small sample sizes because they carry potential to overfit the training data. A lower-complexity multimodal supervised model using geographically-weighted regression (GWR) for MRI by Mohammed et al [8] identified IDH-mutant 1p19q non-codeleted patients who exhibited a phenomenon called T2-FLAIR mismatch. This method leveraged relationships between the T2 modality and the FLAIR modality of MRI to successfully distinguish patients with T2-FLAIR mismatch

[1]Elisa Warner, Santhoshi Krishnan and Nicholas Wang are with the Department of Computational Medicine and Bioinformatics, University of Michigan Ann Arbor, Ann Arbor, MI 48109, USA elisawa@umich.edu, ncwang@umich.edu

[2]Joonsang Lee and Arvind Rao are with the Faculty of the Department of Computational Medicine and Bioinformatics, University of Michigan Ann Arbor, Ann Arbor, MI 48109, USA leejoons@umich.edu, ukarvind@umich.edu

[3]Santhoshi Krishnan is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005, USA sankris@umich.edu

[4]Shariq Mohammed is with the Faculty of Biostatistics, Boston University, Boston, MA 02215, USA shariqm@bu.edu

[5]Ashok Srinivasan and Jayapalli Bapuraj are with the Faculty of Radiology, Michigan Medicine, Ann Arbor, MI 48109, USA ashoks@umich.edu, jrajiv@umich.edu

vs those without.

In this study, we investigate the use of GWR to determine PsP presence in patients with increased tumor size after adjuvant therapy. We leverage this lower-complexity statistical approach followed by logistic regression and naive Bayes to explore multimodal relationships between pairs of conventional MRI sequences as a potential tool to discriminate PsP from TP. We also explore the potential role of an alternative advanced imaging technique to PWI called diffusion weighted imaging (DWI) as a possible discriminator of PsP, given some preliminary interest in the modality [9], [10], [3]. Although conventional MRI is known to be of little help in PsP diagnosis [5], [6], we hypothesize that some modality pairs may show differential patterns which can discriminate PsP from TP better than random. Our use of GWR attempts to identify whether class differences exist in the relationships between multimodal pairs and additionally uses fewer features than is expected in deep learning strategies. This methodology is novel in the context of PsP and TP discrimination in treatment-related changes and provides an alternative image processing approach to deep learning methods for small sample sizes.

II. METHODS

A. Data Acquisition

Data were collected in accordance with relevant guidelines and regulations and approved by the Institutional Review Board at the University of Michigan (IRBMED, HUM00145517). Data were analyzed retrospectively and retrieved from the Electronic Medical Record Search Engine (EMERSE) and DataDirect databases [11]. In this study, five MRI sequences from fifty patients with high grade (WHO grade III or IV) diffuse infiltrating glioma (astrocytoma or oligodendroglioma) were recorded from 2009 to 2018. All patients received an adjuvant therapy followed by two to three recorded follow-up visits. At the last follow-up, histopathological analysis of the tumor site was conducted. MRI sequence data included in this study come from the follow-up directly prior to histological analysis of the tumor site. A label of "pseudoprogression" (PsP) or "true progression" (TP) was determined based on the results of the histopathological analysis.

Characteristic	PsP	TP	Total	p-val
Total	13	29	42	
Age (mean)	53.34	45.69		0.100
Sex				0.524
Male	9	17	26	
Female	4	12	16	
IDH status				0.627
wildtype	6	14	20	
mutant	5	8	13	
missing	2	7	9	
1p19q status				0.217
wildtype	0	3	3	
mutant	3	4	7	
missing	10	22	32	
EGFR status				N/A
wildtype	0	0	0	
mutant	4	5	9	
missing	9	24	33	
P53 status				0.590
wildtype	5	4	9	
mutant	12	6	18	
missing	12	3	15	

TABLE I

PATIENT DEMOGRAPHIC TABLE

B. Dataset and Preprocessing

The dataset contained 50 patient MRI scans with masks (PSP=13, TP=37) for the pre-contrast T1-weighted (T1pre), post-contrast T1-weighted (T1post), T2-weighted (T2), and fluid attenuated inversion recovery (FLAIR) modalities. However, after removing patients with zero-signal modalities, 42 patients (PSP=13, TP=29) were remaining. Zero-signal modality primarily meant patients with zero FLAIR signal. An additional subtraction map modality was created by subtracting T1pre from T1post (T1postpre). Lastly, a DWI modality called apparent diffusion coefficient (ADC) was also utilized from each patient.

Patient MRI were registered and sized such that all modalities were the same size then normalized via whitestripe. No skull extraction was necessary for this procedure because only the area indicated by the tumor mask was analyzed.

In order to mitigate the dataset imbalance, PsP patients were oversampled. The most ideal slice for all patients was first chosen based on the largest mask slice using Matlab R2022b. Two additional slices above and below the largest mask slice were extracted from PsP patients, if they existed. Model input values were selected from each modality by overlaying the mask slice over the image and extracting values within the mask bounds. After oversampling PsP patients, 63 samples were available for all modalities (PSP = 34, TP = 29).

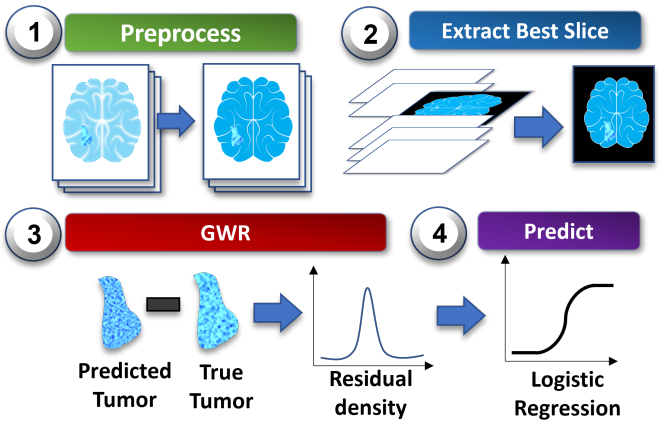

Fig. 1. An illustrated depiction of the methods used in this study. First, (1) pre-processing on the MR images included registration and whitestripe normalization. Then, (2) a single slice is extracted from the MRI volume, and (3) GWR is applied to the tumor region of the extracted slice. Finally, (4) select characteristics of the residual density curve output from GWR are entered into a logistic regression model.

C. Model

We constructed a geographically-weighted regression (GWR) to analyze the relationship between pairs of MRI modalities. Based on the modalities T1pre, T1post, T2, FLAIR, T1postpre and ADC, fifteen modality pairs were constructed. The predictor x of modality X of each pair was regressed onto y pixels from modality Y using the GWModel package in R(4.2.0) [12]. GWR is a form of linear regression that assumes regression weights will differ depending on the spatial locations of $x \in X$ and $y \in Y$. It therefore remains linear within specific localities i for all pixel locations x_i and y_i in a single MRI slice, such that

$$y_i = \beta_0 + \sum_M \beta_m x_{mi} + \varepsilon_i \quad (1)$$

, where β_0 represents the bias at location i, M represents number of features, x_{mi} represents the m^{th} feature at i, and β_m represents the slope for feature m. In this study, the only feature analyzed was the pixel intensity at location i, so $M = 1$.

After fitting a model, residuals were calculated by subtracting the true y_i values from Y with the predicted $\hat{y}_i$ values from the model. Residuals were then flattened and densities were calculated into a density vector of 1×1000. Density curves for all patients were standardized to the same maximum and minimum bounds. Then, features from the curves were extracted as input into a logistic regression model using python 3.8 with Anaconda. In order to minimize the number of inputs into the regression model, we first shrunk the density to 1×500 by sampling every other value from the density curve.

There are two challenges with applying a residual density curve to logistic regression: 1) high dimensionality in a logistic regression model will encourage overfitting and will inappropriately overparameterize a model built on a small sample size, and 2) input expectations for logistic regression require independent and identically-distributed features. Features chosen which are too close together will be too colinear, perhaps biasing the model. In order to mitigate these two challenges, we address the first by sorting the regression vector by standard deviation and sorting in descending order, based on the assumption that areas of the curve with more overall variation will confer more information than those which fluctuate slightly. From these ordered areas of the curve, we select only the top k number of curve positions. In order to mitigate collinearity, we set a distance requirement between positions of the curve to discourage points from belonging to the same monotonic increase or decrease of a curve, and enforced an L2 regularization. Patient-wise leave one out cross validation (LOOCV) was utilized in the implementation and area under the curve (AUC) was assessed for each test for each modality pair. Sensitivity and specificity metrics were also extracted for each test.

Lastly, we conduct a post-hoc test to assess the predictive power of combining multiple modalities, as [6] and [7] suggest that combining multiple predictive indicators may produce better predictability. Using a naive Bayes algorithm, we combined models for all predictive modality pairs (AUC$>$ 0.5) in a late fusion model. Modality pairs were added into the model by their order of predictability in AUC from highest AUC to lowest. As in the above test, patient-wise LOOCV was utilized due to small sample size.

III. RESULTS

Patient demographics can be viewed in Table I. No significant differences in age, sex or any other clinical indicators were found. P-values for EGFR could not be calculated because no variation existed between PsP and TP sample populations.

Results for the logistic regression assessment can be viewed in Table II. We tested k values of 1,2,3,5,10, but found decreasing performance as the number of features increased past $k = 3$, signifying overfitting. All pairs involving ADC (pairs 11-15) demonstrated AUC values above 0.6 in our supervised model in the test set, with ADC regressed on FLAIR as the highest performing pair. Interestingly, two conventional MRI pairs also performed above 0.6 AUC: T1postpre regressed on T2 and T1post regressed on T2. Pairs 7-10 in Table II demonstrated values below 0.50, indicating that the signal for these were too weak given the sample size.

Results from the naive Bayes model are shown in Figure 2. In this analysis, the highest performing model was the result of a combination of the top 5 modality pairs, which included only those modality pairs with ADC as a predictor. Interestingly, after including the contribution of the conventional MRI sequences, the model performance decreases. Models including modality pairs beyond the 5 ADC modality pairs, T1post regressed on T2 and T1postpre regressed on T2, performed substantially worse.

Pair	X	Y	AUC	Sens	Spec
1	T1postpre	FLAIR	0.5968	0.6923	0.4483
2	**T1postpre**	**T2**	**0.6180**	**0.6923**	**0.4828**
3	T1post	FLAIR	0.5889	0.6410	0.4483
4	T1post	T1postpre	0.5084	0.5641	0.4483
5	**T1post**	**T2**	**0.6145**	**0.6923**	**0.4828**
6	T1pre	FLAIR	0.5119	0.4615	0.6897
7	T1pre	T1post	0.4907	0.4615	0.5862
8	T1pre	T1postpre	0.4898	0.4615	0.5862
9	T1pre	T2	0.4598	0.4103	0.5862
10	T2	FLAIR	0.3917	0.1282	0.5862
11	**ADC**	**FLAIR**	**0.6684**	**0.6154**	**0.6552**
12	**ADC**	**T1post**	**0.6251**	**0.5385**	**0.5517**
13	**ADC**	**T1postpre**	**0.6260**	**0.5385**	**0.6207**
14	**ADC**	**T2**	**0.6525**	**0.6667**	**0.6207**
15	**ADC**	**T1pre**	**0.6242**	**0.5385**	**0.5862**

TABLE II

AUC, SENSITIVITY AND SPECIFICITY REPORTED FOR DETECTING PSP FROM TP USING LOGISTIC REGRESSION ANALYSIS[1]

IV. DISCUSSION AND CONCLUSION

Distinguishing pseudoprogression, a result of adjuvant therapy, from true progression, has been a classically difficult

[1]Highlighted rows are for modality pairs where AUC$>$ 0.6

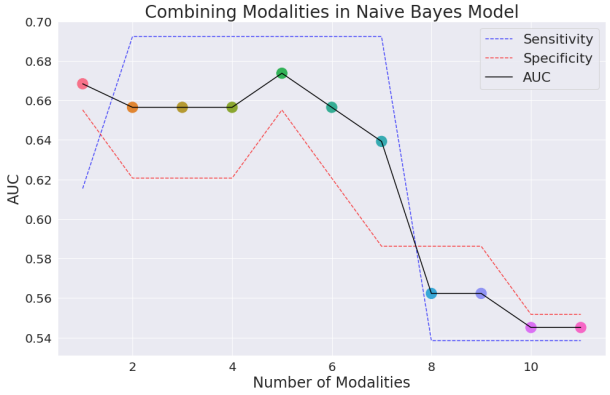

Fig. 2. Results from naive Bayes late fusion model combining multiple modality pairs to distinguish PsP from TP. The best model performs at an AUC of 0.6737 and includes all ADC modality pairs.

challenge, particularly using conventional MRI sequences. However, there are benefits of being able to use conventional MRI to distinguish PsP in HGG patients, because MRI is routinely used for analysis of glial tumors. In this study, we attempted to assess the predictive power of both conventional MRI sequences and one advanced imaging technique (DWI) at distinguishing PsP v TP using low-complexity models for smaller sample sizes and pathologist-confirmed data regarding PsP diagnosis. A set of 15 modality-pairs were investigated using GWR followed by logistic regression to analyze if correlations between imaging modalities could reap discriminative patterns. In the analysis, all modality pairs involving ADC demonstrated marked discriminative ability with AUCs above 0.6, and T1postpre regressed on T2 as well as T1post regressed on T2 also showed some promise in one analysis. Our results on ADC's discriminative ability contribute to a small body of growing research which have also found a utility in diffusion weighted imaging. Two different groups found differences in ADC maps between recurrent and non-recurrent gliomas after radiotherapy using samples of 18 patients and 17 patients, respectively [9], [10], [3]. Thus, our study is the largest to assess ADC as a tool for recurrence (TP) v non-recurrence (PsP) in HGG patients to our knowledge. However, we acknowledge the limitations of our study due to small sample size and believe further studies with larger samples are needed to fully confirm the utility of ADC. Our study also stands as the first to find correlations involving T1post, T1pre and T2 as potentially helpful in distinguishing PsP v TP.

This study carries two main novelties. Firstly, although discrimination of PsP v TP is challenging with conventional MRI, our method found that relationships in some sequences, specifically DWI modalities such as ADC and possibly T1-post or T1postpre regressed on T2, carry some predictive power. Lastly, our method demonstrates the successful use of low-complexity models on conventional MRI sequences as an alternative to deep learning techniques in small sample sizes. The method leverages relationships between multiple modality pairs and can be used for supervised image analysis when deep learning applications are deemed inappropriate.

ACKNOWLEDGMENT

E.W., S.K. and A.R. are supported by NIH Grant R37-CA214955 and American Cancer Society ACS RSG-19-003-01-CCE. A.R. and S.K. were supported by CCSG Bioinformatics Shared Resource 5 P30 CA046592, a gift from Agilent technologies, and a Precision health Investigator award from U-M Precision Health to A.R. S.K. and A.R. were also partially supported by The University of Michigan (U-M) startup institutional research funds. S.K. and A.R. were also supported by a Research Scholar Grant from the American Cancer Society (RSG-16-005-01). N.W. is a founder and equity holder in the startups Applied Morphomics Inc. and mBIOHEALTH.

REFERENCES

[1] Y. Li, Y. Ma, Z. Wu, R. Xie, F. Zeng, H. Cai, S. Lui, B. Song, L. Chen, and M. Wu, "Advanced imaging techniques for differentiating pseudoprogression and tumor recurrence after immunotherapy for glioblastoma," *Frontiers in Immunology*, vol. 12, Nov. 2021.

[2] Y. Ma, Q. Wang, Q. Dong, L. Zhan, and J. Zhang, "How to differntiate pseudoprogression from true progression in cancer patients treated with immunotherapy," *Am J Cancer Res*, vol. 9, Aug. 2019.

[3] T. J. Kruser, M. P. Mehta, and H. I. Robins, "Pseudoprogression after glioma therapy: a comprehensive review," *Expert Review of Neurotherapeutics*, vol. 13, pp. 389–403, Apr. 2013.

[4] N. Guibert, J. Mazieres, M. Delaunay, A. Casanova, M. Farella, L. Keller, G. Favre, and A. Pradines, "Monitoring of KRAS-mutated ctDNA to discriminate pseudo-progression from true progression during anti-PD-1 treatment of lung adenocarcinoma," *Oncotarget*, vol. 8, pp. 38056–38060, June 2017.

[5] R. J. Young, A. Gupta, A. D. Shah, J. J. Graber, Z. Zhang, W. Shi, A. I. Holodny, and A. M. P. Omuro, "Potential utility of conventional MRI signs in diagnosing pseudoprogression in glioblastoma," *Neurology*, vol. 76, pp. 1918–1924, May 2011.

[6] M. Mullins, G. Barest, P. Schaefer, F. Hochberg, R. Gonzalez, and M. Lev, "Radiation necrosis versus glioma recurrence: conventional mr imaging clues to diagnosis," *Neurology*, vol. 26, no. 8, pp. 1967–1972, 2005.

[7] J. Lee, N. Wang, S. Turk, S. Mohammed, R. Lobo, J. Kim, E. Liao, S. Camelo-Piragua, M. Kim, L. Junck, J. Bapuraj, A. Srinivasan, and A. Rao, "Discriminating pseudoprogression and true progression in diffuse infiltrating glioma using multi-parametric MRI data through deep learning," *Scientific Reports*, vol. 10, Nov. 2020.

[8] S. Mohammed, V. Ravikumar, E. Warner, S. Patel, S. Bakas, A. Rao, and R. Jain, "Quantifying t2-FLAIR mismatch using geographically weighted regression and predicting molecular status in lower-grade gliomas," *American Journal of Neuroradiology*, vol. 43, pp. 33–39, Nov. 2021.

[9] C. Asao, Y. Korogi, M. Kitajima, T. Hirai, Y. Baba, K. Makino, M. Kochi, S. Morishita, and Y. Yamashita, "Diffusion-weighted imaging in the follow-up of treated high-grade gliomas: Tumor recurrence versus radiation injury," *American Journal of Neuroradiology*, vol. 26, no. 6, pp. 1455–1460, 2005.

[10] P. A. Hein, C. J. Eskey, J. F. Dunn, and E. B. Hug, "Diffusion-weighted imaging in the follow-up of treated high-grade gliomas: Tumor recurrence versus radiation injuary," *American Journal of Neuroradiology*, vol. 25, pp. 201–209, Feb. 2004.

[11] W. Lin, T. Tong, Q. Gao, D. Guo, X. Du, Y. Yang, G. Guo, M. Xiao, M. Du, X. Qu, and Alzheimer's Disease Neuroimaging Initiative, "Convolutional neural networks-based MRI image analysis for the alzheimer's disease prediction from mild cognitive impairment," *Front. Neurosci.*, vol. 12, p. 777, Nov. 2018.

[12] I. Gollini, B. Lu, M. Charlton, C. Brunsdon, and P. Harris, "bGWmodel/b: Anir/ipackage for exploring spatial heterogeneity using geographically weighted models," *Journal of Statistical Software*, vol. 63, no. 17, 2015.