Neuroscience Informatics 2 (2022) 100088

Contents lists available at ScienceDirect

Neuroscience Informatics

www.elsevier.com/locate/neuri

Original article

Comparative study of radiologists vs machine learning in differentiating biopsy-proven pseudoprogression and true progression in diffuse gliomas

Sevcan Turk [a,c,*,1], Nicholas C. Wang [b], Omer Kitis [c], Shariq Mohammed [b,d], Tianwen Ma [d], Remy Lobo [a], John Kim [a], Sandra Camelo-Piragua [e], Timothy D. Johnson [d], Michelle M. Kim [g], Larry Junck [f], Toshio Moritani [a], Ashok Srinivasan [a], Arvind Rao [b], Jayapalli R. Bapuraj [a]

[a] Department of Radiology, Division of Neuroradiology Michigan Medicine Ann Arbor Michigan, United States of America
[b] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, United States of America
[c] Department of Radiology, Ege University Izmir, Turkey
[d] Department of Biostatistics, University of Michigan, Ann Arbor Michigan, United States of America
[e] Department of Pathology, Division of Neuropathology, University of Michigan, Ann Arbor Michigan, United States of America
[f] Department of Neurology, University of Michigan Ann Arbor Michigan, United States of America
[g] Department of Radiation Oncology, University of Michigan Ann Arbor Michigan, United States of America

ARTICLE INFO

Article history:
Received 28 February 2022
Received in revised form 1 June 2022
Accepted 8 June 2022

Keywords:
Pseudoprogression
Tumor progression
Machine learning
Radiomics
Glial tumor

ABSTRACT

Background and Purpose: MRI features of tumor progression and pseudoprogression may be indistinguishable especially without enhancing portion of the diffuse gliomas. Our aim is to discriminate these two conditions using radiomics and machine learning algorithm and to compare them with human observations.

Materials and Methods: Three consecutive MRI studies before a definitive biopsy in 43 diffuse glioma patients (7 pseudoprogression and 36 true progression cases) who underwent treatment were evaluated. Two neuroradiologists reviewed pre- and post-contrast T1, T2, FLAIR, ADC, rCBV, rCBF, K2, and MTT maps. Patterns of enhancement, ADC maps, rCBV, rCBF, MTT, K2 values, and perilesional FLAIR signal intensity changes were recorded. Odds ratios (OR) for each descriptor, raters' success in predicting true and pseudoprogression, and inter-observer reliability were calculated using the R statistics software. Unpaired Student's *t*-test and receiver operating characteristic (ROC) analysis were applied to compare the texture parameters and histogram analysis of pseudo- and true progression groups. All first-order and second-order image texture features and shape features were used to train and test the Random Forest classifier (RFC). Observers' success and RFC were compared.

Results: Observers could not identify true progression in the first visit. However, accuracy of the RFC model was 81%. For the second and third visits, the rater's success of prediction was between 62% and 72%. The accuracy for the second and last visit with RFC was 75% and 81%.

Conclusions: Random Forest classifier was more successful than human observations in predicting pseudoprogression using MRI.

1. Introduction

After chemoradiation 20–36% of glial tumor patients could represent with increased contrast agent uptake and apparent enlargement of the lesion with or without new lesions which mimicking tumor progression. This phenomenon is referred to as pseudoprogression which could be seen 36% of the high-grade glioma and 20% of low-grade glioma patients after chemoradiation [22, 19]. Pseudoprogression has become a significant challenge in the

* Corresponding author at: 1500E. Medical Center Drive, University Hospital, B2A209B SPC5030, Ann Arbor MI 48109-5030, USA.
E-mail addresses: sevcanturk.ege@hotmail.com, sevcant@med.umich.edu (S. Turk).

[1] Grant Support: TÜBİTAK research fellowship grant, University of Michigan seed grant. Presentation: ASNR 2019, ECR 2019.

https://doi.org/10.1016/j.neuri.2022.100088

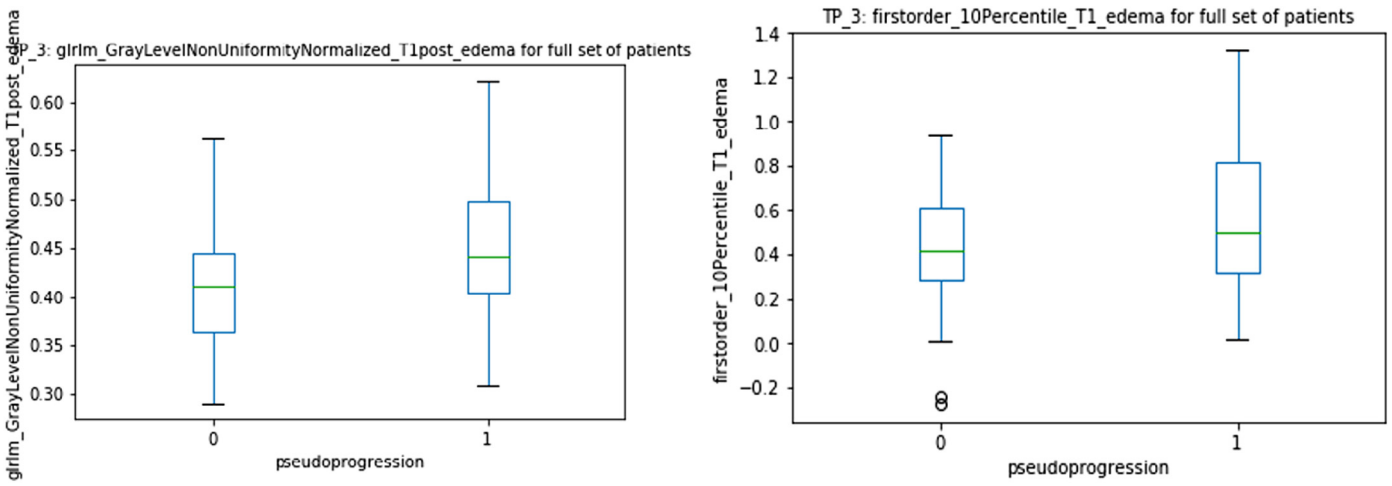

Fig. 1. Texture features for the first visit after chemo-radiotherapy. The first graph (right) shows the statistically significant difference between true tumor progression (0) and pseudoprogression (1) groups on post-contrast T1W images. The second order statistics texture feature, gray level non uniformity, for the area labeled edema shows higher heterogeneity in pseudoprogression group. Second graph (left) shows the statistically significant difference between true tumor progression (0) and pseudoprogression (1) groups on pre-contrast T1W images. The first order statistics analysis, 10^{th} percentile mean intensity value on pre-contrast T1W image for the area labeled edema was higher in pseudoprogression group.

follow-up of diffuse glioma undergoing treatment, as only repeat biopsy is conclusive in differentiating pseudoprogression or true recurrence. Radiological follow up have been the mainstay in resolving the issue in day-to-day practice and may result in treatment delay for true progressive disease [1]. 10–30% of patients treated with chemoradiation for high-grade glioma may exhibit confounding MRI findings [2]. To overcome these limitations, the Response Assessment in Neuro-Oncology (RANO) Working Group formulated updated response assessment criteria for high-grade gliomas. The RANO criteria consider additional clinical and radiographic factors such as a change in clinical status, appearance of new lesions, and influence of corticosteroid use, changes in the area of edema on T2-weighted imaging (T2WI)/fluid attenuation inversion recovery (FLAIR) imaging, and the area of enhancement of tumors [3]. In addition to the conventional MRI sequences, diffusion-weighted and perfusion MRI are widely used to diagnose true- and pseudo-progression and have high reported diagnostic accuracy [2,3]. The objective of this retrospective study was two-fold. The first was to assess the morphological and functional MRI features to differentiate true progression and pseudoprogression in diffuse glioma cases. The second objective was to compare radiomics features and machine learning algorithms with human observers to discriminate between these two groups of patients.

2. Material and methods

In this IRB-approved, HIPAA compliant study, 43 diffuse glioma patients (13 oligodendroglioma, 7 diffuse astrocytoma, 23 glioblastoma) with 129 MRI studies who underwent total tumor resection and adjuvant chemoradiation, and had multiple follow-up MRI studies between 2010–2018, were evaluated. The patients were randomly selected. 60% of image acquisitions were obtained on a 1.5 Tesla magnet. The rest of the studies (38%) were performed on a 3 Tesla magnet, and three studies (2%) were done on a 1 Tesla magnet. 1, 1.5, and 3T Philips, 3T Siemens, and 1.5T GE MRI machines were used for image acquisition.

2.1. Enrollment criteria

A written, informed consent was waived for this retrospective study. We used an existing MRI data set for this novel and unique analysis of the imaging studies [4]. All cases were selected from the existing Electronic Medical Research Search Engine (EMERSE) [5] and the Data Direct database with search terms for “pseudoprogression”, “treatment related changes,” “tumor recurrence,” “recurrent tumor”, “tumor progression” and “high grade gliomas”. All histopathological diagnoses were explicitly reviewed for this study and were compliant with the 2016 revision of the WHO Classification of Central Nervous System Tumors [6].

All participants of the study were blinded to the final pathological diagnosis for the study. An initial selection of cases was based upon the evaluation of the charts which had shown two surgical procedures in patients presenting with the initial diagnosis of glioma. All histopathological studies were reviewed at the commencement of this project which were the best possible biopsy sites despite the known tumor heterogeneity. Seven cases of pseudoprogression were included in this study, together with 36 cases of true progression. All cases were proven by histopathological evaluations following repeat open biopsy or repeat surgery following an initial resection or biopsy. True progression cases included grade 2 ($n = 17$), grade 3 ($n = 2$) and grade 4 ($n = 23$), diffuse astrocytoma NOS ($n = 1$). Seven cases of pseudoprogression consisted of grade 2 ($n = 2$), and grade 4 ($n = 5$). Radiation necrosis was not included in the study.

2.2. Image analysis by human observers and statistical analyses of the data

Imaging evaluation was carried out as per the following protocol. The MRI study before revision surgery or second biopsy with a known pathology result was chosen as the final time point of study where all the patients were undergone maximal resection. This study and immediate two prior follow-up studies were evaluated with a set of three MRI evaluations per patient. The sequences included in the study evaluation were: pre- and post-contrast axial T1, axial T2, axial FLAIR, and axial ADC maps. Additionally, on DSC perfusion evaluation the corrected rCBV, rCBF, K2, and MTT maps, with leakage correction, were reviewed. Two board-certified Neuroradiologists at least 4 years’ experience in this academic hospital, reviewed each case blindly to the pathology report and rated the image features as shown in Figs. 1 and 2. Raters were aware of the previous MRI when they were evaluating second and third follow-up images. Linear or nodular enhancement, one or multiple enhancing lesions, high or low ADC, corrected rCBV, rCBF, MTT, K2 values (compared to the contralateral normal-appearing white matter), FLAIR signal changes within time were noted, in-

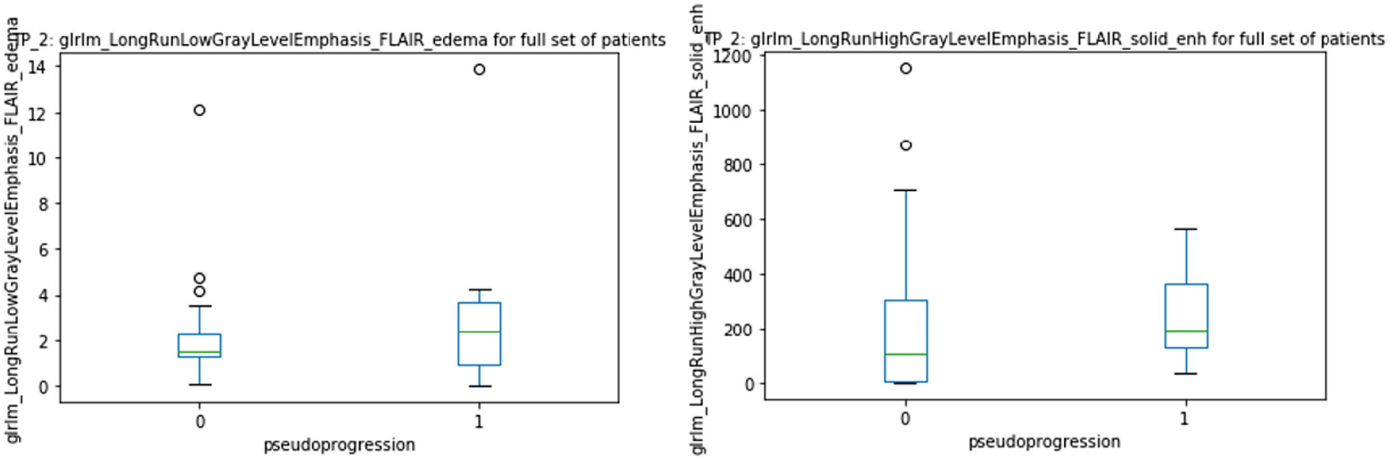

Fig. 2. Texture features for the second visit after chemo-radiotherapy. The first graph (right) shows the statistically significant difference between true tumor progression (0) and pseudoprogression (1) groups on FLAIR images. The second order statistics texture feature, long run low gray level emphasis, for the area labeled edema shows higher intensity values in pseudoprogression group. Second graph (left) shows the statistically significant difference between true tumor progression (0) and pseudoprogression (1) groups on FLAIR images. The second order statistics analysis, long run high gray level emphasis, for the area labeled solid enhancing part shows higher mean intensity values in pseudoprogression group than true tumor progression.

Table 1
Tabulated assessment criteria of the human observers which was used by human observers to document impressions of evaluation of MRI studies.

DATE 1	DATE 2	DATE 3
1-Band-like or linear enhancement	1-Band-like or linear enhancement	1-Band-like or linear enhancement
2-Nodular enhancement	2-Nodular enhancement	2-Nodular enhancement
3-One enhancing lesion	3-One enhancing lesion	3-One enhancing lesion
4-Multiple enhancing lesion	4-Multiple enhancing lesion	4-Multiple enhancing lesion
5-Nodular nonenhancing tumor part	5-Nodular nonenhancing tumor part	5-Nodular nonenhancing tumor part
6-ADC lower than contralateral white matter	6-ADC lower than contralateral white matter	6-ADC lower than contralateral white matter
7-ADC higher than contralateral white matter	7-ADC higher than contralateral white matter	7-ADC higher than contralateral white matter
8-rCBV lower than contralateral white matter	8-rCBV lower than contralateral white matter	8-rCBV lower than contralateral white matter
9-rCBV higher than contralateral white matter	9-rCBV higher than contralateral white matter	9-rCBV higher than contralateral white matter
10-CBF lower than contralateral white matter	10-CBF lower than contralateral white matter	10-CBF lower than contralateral white matter
11-CBF higher than contralateral white matter	11-CBF higher than contralateral white matter	11-CBF higher than contralateral white matter
12-MTT lower than contralateral white matter	12-MTT lower than contralateral white matter	12-MTT lower than contralateral white matter
13-MTT higher than contralateral white matter	13-MTT higher than contralateral white matter	13-MTT higher than contralateral white matter
14-K2 lower than contralateral white matter	14-K2 lower than contralateral white matter	14-K2 lower than contralateral white matter
15-K2 higher than contralateral white matter	15-K2 higher than contralateral white matter	15-K2 higher than contralateral white matter
16-Increase in enhancing part size more than 25%	16-Increase in enhancing part size more than 25%	16-Increase in enhancing part size more than 25%
17-Decrease in enhancing part size more than 25%	17-Decrease in enhancing part size more than 25%	17-Decrease in enhancing part size more than 25%
18-Stable size in enhancing/nonenhancing nodular lesion	18-Stable size in enhancing/nonenhancing nodular lesion	18-Stable size in enhancing/nonenhancing nodular lesion
19-New enhancing/nonenhancing nodular lesion	19-New enhancing/nonenhancing nodular lesion	19-New enhancing/nonenhancing nodular lesion
20-Increase in FLAIR hyperintense part size more than 25%, less than 50%	20-Increase in FLAIR hyperintense part size more than 25%, less than 50%	20-Increase in FLAIR hyperintense part size more than 25%, less than 50%
21-Increase in FLAIR hyperintense part size more than 50%	21-Increase in FLAIR hyperintense part size more than 50%	21-Increase in FLAIR hyperintense part size more than 50%
22-Decrease in FLAIR hyperintense part size more than 25%	22-Decrease in FLAIR hyperintense part size more than 25%	22-Decrease in FLAIR hyperintense part size more than 25%
23-Increasing/decreasing ADC	23-Increasing/decreasing ADC	23-Increasing/decreasing ADC
24-Increasing/decreasing CBV	24-Increasing/decreasing CBV	24-Increasing/decreasing CBV
25-Increasing/decreasing enhancement	25-Increasing/decreasing enhancement	25-Increasing/decreasing enhancement
Final Impression:	Final Impression:	Final Impression:

cluding interval changes in these parameters. A scoring sheet was created depending on the RECIST, MacDonald, and RANO criteria (Table 1). The neuroradiologists evaluated each statement and marked it with three responses: yes/no/not applicable for each imaging feature. The two neuroradiologists had an independent final impression for each study with the following descriptor: true tumor progression, pseudoprogression, or not-applicable/indeterminate. The last criteria were employed for first assessment since, logically, no conclusion could be drawn on a single assessment, especially without any enhancing area within the operation site. All evaluations were carried out on standard imaging consoles in the department. Ambient light, monitors, and computer peripherals were optimized for the visualization of images. All images were accessed from the PACS system of the department. All readers had access to all three follow- up images with all planes. They started evaluation from the first study and proceeded with second and third follow-up studies. The neuroradiologists did not access the clinical notes of the patients. Odds ratios (OR) for each statement, raters' success of predicting true and pseudoprogression, and inter-observer reliability were calculated. $P < 0.05$ was considered significant, OR < 1 implied that the statement was more strongly associated with pseudoprogression than true progression.

2.3. Radiomics analysis and machine learning

The MRI data were digitally transferred from the PACS workstation to a secured storage drive. All images were segmented with an open-source segmentation tool (ITK-SNAP ver. 3.6) [7] manu-

Table 2
Odds ratios (OR) for each statement, statements' statistical significance of predicting true and pseudoprogression were shown. p-value < 0.05 was considered significant, OR < 1 implied that the statement more strongly associated with pseudoprogression than true progression. Statements 2, 8, 9, 10, and 11 were statistically significant to discriminate the groups.

Statement rater	Visit	S2		S8		S9		S10	
		OR	*P*	OR	*P*	OR	*P*	OR	*P*
Rater 1	1	***0.18***	***0.05***	0.18	0.16	0.71	0.79	0.19	0.17
Rater 1	2	0.19	0.14	***0.11***	***0.03***	1.38	0.74	***0.12***	***0.03***
Rater 1	3	0.31	0.45	0.20	0.12	2.40	0.38	0.20	0.12
Rater 2	1	0.21	0.17	**31.9**	**0.03**	***0.03***	***0.03***	***24.2***	***0.04***
Rater 2	2	0.69	0.75	2.00	0.47	0.50	0.47	1.47	0.69
Rater 2	3	0.92	0.96	3.16	0.46	0.32	0.46	2.54	0.55

Statement rater	Visit	S11		S16		S21	
		OR	*P*	OR	*P*	OR	*P*
Rater 1	1	0.75	0.82	0.20	0.52	0.14	0.44
Rater 1	2	1.47	0.69	0.07	0.07	0.75	0.81
Rater 1	3	2.63	0.34	0.78	0.83	0.23	0.10
Rater 2	1	***0.04***	***0.04***	3.00	0.67	*1.00*	*1.00*
Rater 2	2	0.68	0.69	1.50	0.63	0.60	0.59
Rater 2	3	0.39	0.55	0.583	0.64	0.51	0.43

ally and slice by slice using DICOM data by a 4th-year radiology resident under the supervision of an independent board-certified neuroradiologist who was not involved in the abovementioned assessment. The labels were specified as enhancing solid tumor part, non-enhancing solid tumor part, necrosis, edema, cystic tumor part, and hemorrhage. For normalization purposes, contralateral grey matter, white matter, and CSF were labeled separately with a round 2-dimensional (2-D) ROI. To minimize heterogeneity in image intensity caused by the use of different scanners and acquisition sequences, normalization correction was done according to the data to the mean value of normal-appearing contralateral white matter from a small ROI drawn contralateral above the ventricles [8]. ADC maps were evaluated without prior normalization. Proprietary software (Python; Python Software Foundation, Delaware USA) was used for all machine learning applications. Texture features were obtained from the normalized images by introducing groups of standard first-order and second-order image texture analysis by using PyRadiomics. Calculations were based on the information contained in regions of interest (ROIs) defined by the neuroradiologist and subsequently superimposed on the feature images. Areas of visible necrosis, hemorrhage, calcifications, and cystic components in the tumors were not included in the evaluation. Tissue discrimination was performed by comparing the mean values of the ROI data obtained in each image feature.

Textural information of 2D pixels on images was generated using gray-level co-occurrence matrices (GLCMs) and gray-level run-length matrices (GLRLMs). All the post-processed feature images mentioned above were calculated for each of the modalities on a slice by slice basis. To get a per-patient prediction, though, the features were averaged per scan and fed into the machine learning prediction by Random Forest (RF). For comparisons of the texture parameters of pseudo- and true progression groups, Student's t-test and receiver operating characteristic (ROC) analysis were applied. A one-way analysis of variance with a post-hoc Bonferroni test and ROC analysis was performed to compare each group's features, applying a significance level of 0.05.

Each of the three time-points was considered separately and the patients were split 70–30 randomly into a training and testing set. Two thousand seven hundred texture features were initially measured using first and second-order statistics. To further reduce the feature set, recursive feature elimination and correlation filtering were used to eliminate redundant features [9] resulting in 720 features. After determining the features with the highest diagnostic accuracy in the training dataset through the above analyses, this subset of features was used in training the RF classifier on the training dataset. Classifiers were trained using 5-fold cross-validation of the training cohort. Their predictive performance was evaluated in the cross-validation folds using the area under the ROC curve (AUC). Accuracy, AUC, Youden sensitivity, and specificity were calculated for the held aside testing set data model. The importance of each feature to each random forest model was also calculated and reported.

3. Results

3.1. Descriptives

Nineteen female and 24 male patients were included. The mean age for female patients was 56 years, and the mean age for male patients was 48 years. Mean time interval between biopsy and the MRI immediately before biopsy (third visit) was 51 (SD 143) days. Other consecutive MRIs (second and first visit) were taken approximately 161 (SD 207) and 331 (SD 460) days before the resection. From the first visit, none of the radiologists could identify true progression ($p = 0.87$). From the first visit, nodular enhancement (compared to a band-like) pattern more likely suggested pseudoprogression (for the first visit OR:0.18, $p = 0.05$) (Table 2). Since our dataset consisted of biopsy-proven cases, pseudoprogression imaging features were closer to tumor progression than in what is encountered in a routine clinical scenario. Lower rCBV values compared to contralateral white matter were most likely associated with pseudoprogression for the second visit with one of the human observers' data (OR:0.11, $p = 0.03$). Other radiologist's data support lower rCBV values association with tumor progression for the first visit (OR:32, $P = 0.03$). Higher rCBV values associated with pseudoprogression for the first visit (OR:0.03, $p = 0.03$).

Lower rCBF values around the operation cavity margins were more suggestive for pseudoprogression for second visit (OR:0.12, $p = 0.03$) and more associated with tumor progression for the first visit (OR:24, $p = 0.04$). Higher rCBF values were related with pseudoprogression for first visit (OR:0.04, $p = 0.04$) (Table 3).

A 25% or more increase in the enhancing area or corresponding increase in the surrounding FLAIR hyperintensity by more than 50% of the area was not statistically significant to differentiate groups ($p = 0.6$). Non-enhancing portions of the tumor, one or multiple enhancing lesions, ADC, MTT, and K2 values were not found to be significant discriminators between groups on the visual assessment by radiologists (Table 3). Interobserver kappa correlation was low between radiologists (Table 3).

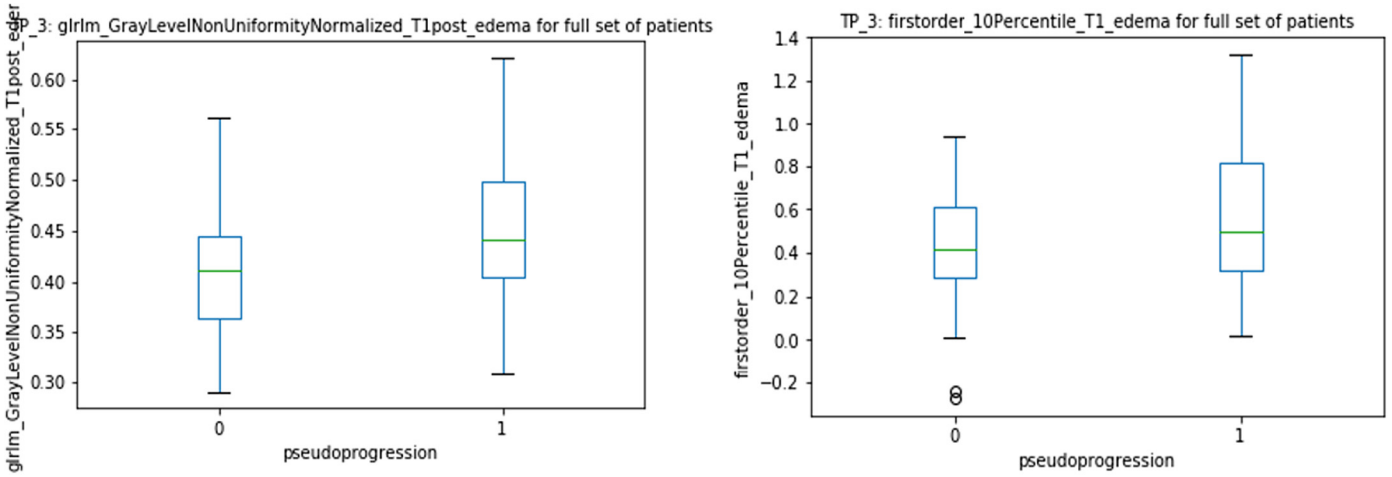

Fig. 3. Texture features for the third visit after chemo-radiotherapy. The first graph (right) shows the statistically significant difference between true tumor progression (0) and pseudoprogression (1) groups on post contrast T1W images. The second order statistics texture feature, gray level non uniformity, for the area labeled edema shows higher values in pseudoprogression group, which means pseudoprogression cases contain more heterogeneous areas than true tumor progression. Second graph (left) shows the statistically significant difference between true tumor progression (0) and pseudoprogression (1) groups on pre-contrast T1W images. The first order statistics analysis, 10^{th} percentile mean intensity value on pre-contrast T1W image for the area labeled edema was higher in pseudoprogression group.

Table 3
Inter-observer kappa correlation for the statements (S) from 1 to 24, and for the final impression (SFI) were low.

Statement number	**Visit 1** Kappa between readers for MR	**Visit 2** Kappa between readers for MR	**Visit 3** Kappa between readers for MR
S1	0.35	0.018	0.071
S2	0.37	0.48	0.46
S3	0.13	0.15	0.18
S4	0.076	0.21	0.14
S5	−0.038	−0.049	−0.087
S6	0.17	0.098	0.11
S7	0.17	0.13	0.1
S8	0.093	0.22	0.14
S9	0.55	0.29	0.051
S10	0.034	0.31	0.2
S11	0.47	0.31	0.11
S12	0	0	0
S13	−0.11	0.086	0.25
S14	0.047	0.038	0.26
S15	0.14	−0.027	0.15
S16	0	0.49	0.38
S17	0	0.48	0
S18	0	0.2	0
S19	0	0.57	0.3
S20	0	0.1	−0.14
S21	0	0.33	0.45
S22	0	−0.033	0.66
S23	0	−0.012	−0.13
S24	0	0.02	−0.13
S25	0	0.28	0.55
Final Impression	0	0.15	−0.006

3.2. Relevant features for RF model

First and second-order statistics were applied for the segmented area on each conventional MRI sequence for every visit separately. More than 2700 texture features were measured, and 700 were statistically significant to discriminate groups when mean values were compared between groups with student *t*-test. A group of graphs among texture features were shown in Figs. 1–3.

3.3. Time to correct diagnosis humans vs RF model

The RF algorithm accuracy was 81.9% for the last visit before revision surgery. The accuracy was 75% for the second visit and 81% for the first visit after chemo-radiotherapy; this may be related to the heterogeneity of time points in this retrospectively collected data. We compared the expert radiologists' success, which was lower than the machine learning method's accuracy for all visits (Table 4).

4. Discussion

This retrospective study compared trained neuroradiologists' interpretations with machine learning algorithms to differentiate true tumor progression from pseudoprogression for glioma patients. The diagnostic performance of the algorithm used in this study reached 0.70–0.90 AUC, which outperformed that of the previous study using the same image data set with multiparametric MRI sequences applying convolutional neural networks to previously segmented images (0.64–0.81 AUC) [4].

A brief review of the relevant studies regarding the differentiation of pseudoprogression and true progression by machine learning techniques is as follows: Chen et al. showed that the most significant GLCM texture maps' AUROC curve was 88% to discriminate groups by using conventional images in a cohort of 22 patients [10]. Jang, Bum-Sup et al. included clinical factors to their CNN and machine learning paradigms and reached an AUROC of 83% by only using post-contrast T1W images for their algorithms training and testing [11]. Moran et al. used only DCE perfusion data to discriminate tumor progression from pseudoprogression for their SVM algorithm and reached 97% sensitivity and specificity, with a 20 patients high-grade glioma population [12]. Bahrami N. et al. found a correlation between true tumor progression and FLAIR signal intensity around the operation cavity by using the Cox proportional hazards model for 33 high-grade glioma patients after chemo-radiotherapy [13]. Xintao Hu et al. trained SVM with pre-post-contrast T1W, T2W, FLAIR, PD, ADC, rCBV, rCBF, MTT maps, and reached to AUROC 89% in their small cohort of 31 glioblastoma patients [14]. Park Ji-Eun et al. showed that a volume weighted voxel-based multiparametric clustering (VBMC) technique was more successful for the differentiation when they used both ADC and rCBV maps in their cohort of patients [15]. Yoon Ra Gyoung et al. reached to 95% sensitivity for discrimination the groups with VBMC by using ADC and CBV maps in their cohort of 76 glioblastoma patients [16]. Random forest (RF) algorithm was found to be one of the best machine learning tool for regression based prediction in these patient groups' texture fea-

Table 4
Comparison of the success between human observers, texture features and RF algorithm.

Visit after treatment	Human Observers' Success for correct diagnosis	Single Texture Feature AUROC	RF Algorithm Accuracy	RF Algorithm AUROC (95% CI)
First Visit	Under %50	67%	81%	57% (44%–70%)
Second Visit	Under %50	65%	75%	62% (50%–73%)
Third Visit	62%–72%	68%	81%	75% (66%–84%)

tures [17]. In our study perfusion images did not add significant value. This may be related with early invasive changes which did not include enhancing portion or significant change in perfusion parameters. However, according to the current literature, DSC and DCE perfusion MRI demonstrated higher diagnostic accuracy than conventional technics with 87–92% sensitivity [21]. To reveal those unusual, non perfusion interpretable cases, our dataset was supposed to be more suitable due to predetermined selection bias related with analyzing only biopsy needed-proven cases [21].

All of the previous studies included only images for a single time point before revision surgery with a focus on the enhancing portion of the treated tumor and the surrounding FLAIR hyperintense area. However, our research included evaluations of three previous follow up images before revision surgery and focused on the predominantly non-enhancing portions of the infiltrating tumor, the FLAIR hyperintense signals and the enhancing tumor. FLAIR hyperintense area was more useful to detect tumor invasion than post contrast T1 images in our series especially before progression to enhancement around the operation cavity as a definite marker for disease progression. There was no statistically significant difference between groups for texture features for enhancing areas in our data. Chang PD et al., in their hallmark study, also concluded that with the logistic regression model for FLAIR images there was a statistically significant difference on FLAIR hyperintense area to discriminate tumor invasion in their 26 patients' data set [18]. We had 17 cases of grade 2 gliomas two of whom showed pseudoprogression (12%). Literature among low grade glioma pseudoprogression is very limited since these patients' treatment protocol does not include chemoradiation. Review of the relevant literature of pseudoprogression in low grade gliomas showed on one comprehensive study by van West et al. who have described low grade glioma patients' post-treatment changes with 20% incidence of pseudoprogression in a series of 71 patients [19]. The one characteristic feature they demonstrated was that enhancing area was smaller in pseudoprogression than tumor progression. Since our dataset only consisted of biopsy proven cases, lower prevalence could represent only a portion of the patients that we see in the clinic.

Non blinded human observations for presumed enhancement of tumor recurrence and radiation necrosis was described in a recent study by Gao et al. [20]. The authors carried out human assessments to select slices which could then be processed for automatic segmentation to reduce observer bias and to provide the ground truth in this cohort. This was unlike the procedure we followed where the ground truth rested with the histopathological assessment. We included 43 glioma patients with heterogenous tumors (oligodendroglioma, diffuse astrocytoma and glioblastoma). It is known that these tumors are different on MRI perfusion and therefore differences in interpretations for differentiating tumor progression and pseudoprogression are expected.

4.1. Limitations

The limitations are a relatively small number of patients, and data from a single institution. Increasing the sample size, using standardized time points and image acquisition, including an independent validation data set and deep learning application, may help refine delineation of pseudoprogression from true tumor progression by machine learning paradigms. Future studies will aim stricter time points and standard MRI acquisitions. Adding clinical evaluations to the decision algorithms also expected to rise the success. The future advances in AI could facilitate integration of machine learning techniques for an online computation of the images to yield probable diagnosis of true progression or pseudoprogression as add on feature on PACS systems which is now common for example in the assessment of perfusion MRI scans. Field strength of magnets may have an influence in the interpretation of certain sequences. While we are acutely aware of this aspect of data accrual; it may be challenging to obtain images on a single vendor platform and may itself introduce a bias in the final observations by either humans or by machine learning. It is apparent at this time that the heterogeneous data base was helpful for us to make robust evaluation. Manual segmentation of the images is another perceived limitation, however, newer federated tumor segmentation pipelines will afford a more uniform processing of image for machine learning analyses. Lastly since there is no consensus in the pathology literature for what actually constitutes pseudoprogression, evaluation of serial studies in the time period early in the course of treatment of infiltrative gliomas may prove to be the bedrock in the assessment of true and pseudoprogression.

5. Conclusion

Random Forest algorithm and texture features alone are more successful than human observers in discriminating tumor progression even without using perfusion data. This suggests that human observers are more prone to misevaluate non enhancing FLAIR hyperintensity around the operation cavity. The intrinsic differences in the imaging characteristics between groups are not visible to human eye and post contrast images and perfusion data are not helpful in these non-enhancing early invasion. All these subtle nuances make texture features more reliable than human eyes alone. None of the MRI features which radiologists use for assessment including perfusion parameters, could reach to high accuracy to differentiate pseudoprogression from true tumor progression by human observers; thereby, qualitative evaluation of MRI features with machine learning paradigms appear to be more successful for this distinction than human observers within the limitations of this study design.

Human and animal rights

The authors declare that the work described has been carried out in accordance with the Declaration of Helsinki of the World Medical Association revised in 2013 for experiments involving humans as well as in accordance with the EU Directive 2010/63/EU for animal experiments.

Informed consent and patient details

The authors declare that this report does not contain any personal information that could lead to the identification of the patient(s) and/or volunteers.

Declaration of competing interest

The authors declare that they have no known competing financial or personal relationships that could be viewed as influencing the work reported in this paper.

Funding

This work has been supported by:
TÜBİTAK research fellowship grant, University of Michigan seed grant.

Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.neuri.2022.100088.

References

[1] S.C. Thust, M.J. van den Bent, M. Smits, Pseudoprogression of brain tumors, J. Magn. Reson. Imaging 48 (3) (2018 May 7) 571–589.
[2] I. Arrillga-Romany, M. Monje, P.Y. Wen, Neurologic complications of oncologic therapy, in: Herbert B. Newton (Ed.), Handbook of Neuro-Oncology Neuroimaging, 2nd ed., San Diego Elsevier Science & Technology, 2016, pp. 125–142.
[3] P.Y. Wen, D.R. Macdonald, D.A. Reardon, et al., Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group, J. Clin. Oncol. 28 (2010) 1963–1972.
[4] J. Lee, N. Wang, S. Turk, et al., Discriminating pseudoprogression and true progression in diffuse infiltrating glioma using multi-parametric MRI data through deep learning, Sci. Rep. 10 (1) (2020 Nov 23) 20331.
[5] D.A. Hanauer, Q. Mei, J. Law, R. Khanna, K. Zheng, Supporting information retrieval from electronic health records: a report of University of Michigan's nine-year experience in developing and using the Electronic Medical Record Search Engine (EMERSE), J. Biomed. Inform. 55 (2015 Jun) 290–300.
[6] D.N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W.K. Cavenee, H. Ohgaki, O.D. Wiestler, P. Kleihues, D.W. Ellison, The 2016 world health organization classification of tumors of the central nervous system: a summary, Acta Neuropathol. 131 (6) (2016 Jun) 803–820.
[7] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, Guido Gerig, User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability, NeuroImage 31 (3) (2006 Jul 1) 1116–1128.
[8] H. Pang, Y. Ren, X. Dang, X. Feng, Z. Yao, J. Wu, C. Yao, N. Di, D.C. Ghinda, Y. Zhang, Diffusional kurtosis imaging for differentiating between high-grade glioma and primary central nervous system lymphoma, J. Magn. Reson. Imaging 44 (1) (2016 Jul) 30–40.
[9] M. Moon, K. Nakai, Stable feature selection based on the ensemble L_1-norm support vector machine for biomarker discovery, BMC Genomics 17 (Suppl 13) (2016 Dec 22) 1026.
[10] T.T. Tang, J.A. Zawaski, K.N. Francis, et al., Image-based classification of tumor type and growth rate using machine learning: a preclinical study, Sci. Rep. 9 (2019) 12529.
[11] X. Chen, X. Wei, Z. Zhang, R. Yang, Y. Zhu, X. Jiang, Differentiation of true-progression from pseudoprogression in glioblastoma treated with radiation therapy and concomitant temozolomide by GLCM texture analysis of conventional MRI, Clin. Imaging 39 (2015) 775–780.
[12] B.S. Jang, S.H. Jeon, I.H. Kim, et al., Prediction of pseudoprogression versus progression using machine learning algorithm in glioblastoma, Sci. Rep. 8 (2018) 12516.
[13] M. Artzi, G. Liberman, G. Nadav, et al., Differentiation between treatment-related changes and progressive disease in patients with high grade brain tumors using support vector machine classification based on DCE MRI, J. Neurooncology 127 (2016) 515–524.
[14] N. Bahrami, D. Piccioni, R. Karunamuni, Y.H. Chang, N. White, R. Delfanti, T.M. Seibert, J.A. Hattangadi-Gluth, A. Dale, N. Farid, C.R. McDonald, Edge contrast of the FLAIR hyperintense region predicts survival in patients with high-grade gliomas following treatment with bevacizumab, Am. J. Neuroradiol. 39 (6) (2018 Jun) 1017–1024.
[15] X. Hu, K.K. Wong, G.S. Young, L. Guo, S.T. Wong, Support vector machine multiparametric MRI identification of pseudoprogression from tumor recurrence in patients with resected glioblastoma, J. Magn. Reson. Imaging 33 (2) (2011 Feb) 296–305.
[16] J.E. Park, H.S. Kim, M.J. Goh, S.J. Kim, J.H. Kim, Pseudoprogression in patients with glioblastoma: assessment by using volume-weighted voxel-based multiparametric clustering of MR imaging data in an independent test set, Radiology 275 (3) (2015 Jun) 792–802.
[17] Ra Gyoung Yoon, Ho Sung Kim, Myeong Ju Koh, Woo Hyun Shim, Seung Chai Jung, Sang Joon Kim, Jeong Hoon Kim, Radiology 285 (1) (2017) 206–213.
[18] P.D. Chang, D.S. Chow, P.H. Yang, C.G. Filippi, A. Lignelli, Predicting glioblastoma recurrence by early changes in the apparent diffusion coefficient value and signal intensity on FLAIR images, Am. J. Roentgenol. 208 (1) (2017 Jan) 57–65.
[19] E. van West Sophie, G. de Bruin Hein, Bart van de Langerijt, et al., Incidence of pseudoprogression in low-grade gliomas treated with radiotherapy, Neuro-Oncology 19 (5) (2017) 719–725, https://doi.org/10.1093/neuonc/now194.
[20] Y. Gao, X. Xiao, B. Han, G. Li, X. Ning, D. Wang, W. Cai, R. Kikinis, S. Berkovsky, A. Di Ieva, L. Zhang, N. Ji, S. Liu, Deep learning methodology for differentiating glioma recurrence from radiation necrosis using multimodal magnetic resonance imaging: algorithm development and validation, Med. Inform. 8 (11) (2020 Nov 17) e19805.
[21] B.R.J. van Dijken, P.J. van Laar, G.A. Holtman, A. van der Hoorn, Diagnostic accuracy of magnetic resonance imaging techniques for treatment response evaluation in patients with high-grade glioma, a systematic review and meta-analysis, Eur. Radiol. 27 (10) (2017 Oct) 4129–4144, https://doi.org/10.1007/s00330-017-4789-9. Epub 2017 Mar 22. PMID: 28332014; PMCID: PMC5579204.
[22] A.W. Abbasi, H.E. Westerlaan, G.A. Holtman, K.M. Aden, P.J. van Laar, A. van der Hoorn, Incidence of tumour progression and pseudoprogression in high-grade gliomas: a systematic review and meta-analysis, Clin. Neuroradiol. 28 (3) (2018 Sep) 401–411, https://doi.org/10.1007/s00062-017-0584-x. Epub 2017 May 2. PMID: 28466127; PMCID: PMC6105173.