



Longitudinal structural and perfusion MRI enhanced by machine learning outperforms standalone modalities and radiological expertise in high-grade glioma surveillance

Loizos Siakallis¹ · Carole H. Sudre^{2,3} · Paul Mulholland⁴ · Naomi Fersht⁴ · Jeremy Rees^{5,6} · Laurens Topff⁷ · Steffi Thust¹ · Rolf Jager^{1,5} · M. Jorge Cardoso² · Jasmina Panovska-Griffiths^{8,9} · Sotirios Bisdas^{1,5}

Received: 1 December 2020 / Accepted: 12 April 2021
© The Author(s) 2021

Abstract

Purpose Surveillance of patients with high-grade glioma (HGG) and identification of disease progression remain a major challenge in neurooncology. This study aimed to develop a support vector machine (SVM) classifier, employing combined longitudinal structural and perfusion MRI studies, to classify between stable disease, pseudoprogression and progressive disease (3-class problem).

Methods Study participants were separated into two groups: group I (total cohort: 64 patients) with a single DSC time point and group II (19 patients) with longitudinal DSC time points (2-3). We retrospectively analysed 269 structural MRI and 92 dynamic susceptibility contrast perfusion (DSC) MRI scans. The SVM classifier was trained using all available MRI studies for each group. Classification accuracy was assessed for different feature dataset and time point combinations and compared to radiologists' classifications.

Results SVM classification based on combined perfusion and structural features outperformed radiologists' classification across all groups. For the identification of progressive disease, use of combined features and longitudinal DSC time points improved classification performance (lowest error rate 1.6%). Optimal performance was observed in group II (multiple time points) with SVM sensitivity/specificity/accuracy of 100/91.67/94.7% (first time point analysis) and 85.71/100/94.7% (longitudinal analysis), compared to 60/78/68% and 70/90/84.2% for the respective radiologist classifications. In group I (single time point), the SVM classifier also outperformed radiologists' classifications with sensitivity/specificity/accuracy of 86.49/75.00/81.53% (SVM) compared to 75.7/68.9/73.84% (radiologists).

Conclusion Our results indicate that utilisation of a machine learning (SVM) classifier based on analysis of longitudinal perfusion time points and combined structural and perfusion features significantly enhances classification outcome (p value= 0.0001).

Keywords Machine learning · Glioblastoma (GB) · Glioma · Radiomics · Perfusion

✉ Loizos Siakallis
Loizos.siakallis@nhs.net

¹ Lysholm Department of Neuroradiology, National Hospital for Neurology and Neurosurgery, Queen Square, London WC1N 3BG, UK

² Translational Imaging Group, Centre for Medical Image Computing, University College London, London, UK

³ Department of Medical Physics, University College London, London, UK

⁴ Department of Oncology, University College London Hospitals NHS Foundation Trust, London, UK

⁵ Department of Brain Repair and Rehabilitation, UCL Institute of Neurology, London, UK

⁶ Department of Neurooncology, National Hospital for Neurology and Neurosurgery, London, UK

⁷ Department of Radiology, The Netherlands Cancer Institute, Amsterdam, Netherlands

⁸ Institute for Global Health, University College London, London, UK

⁹ The Queen's College, University of Oxford, Oxford, UK

Introduction

High-grade gliomas (HGGs) are the most common type of malignant primary brain tumours, representing 80% of newly diagnosed cases, the majority of which (>50%) correspond to glioblastoma (GB) [1, 2]. Current reference standard treatment includes maximal safe resection, radiation therapy and concurrent temozolomide (TMZ) [2, 3]. Poor prognosis and heterogeneous response to treatment warrant imaging surveillance for these patients [4].

Differentiation of progressive disease (PD) from pseudoprogression (PsP) remains critical for patient management [5]. Structural MRI, even under evolving diagnostic criteria, has been inefficient to reliably differentiate PD from PsP [2, 5–7]. Perfusion MRI has been previously shown to improve this classification [8–10]. Studies aiming to differentiate PD from PsP using dynamic susceptibility contrast (DSC) MR perfusion with a single perfusion time point produced initially promising, albeit conflicting results [8, 10]. A recent meta-analysis on the differentiation between PD and PsP by DSC MR perfusion indicates a pooled sensitivity and specificity of 90% and 88% respectively [11].

The combination of perfusion and structural metrics with the use of multiparametric MRI has been successfully applied to differentiate between PD and PsP in treated GBs [12], to predict the location of recurrence postoperatively [13] and to enhance prediction of overall survival [14]. The addition of perfusion time points combined with multiparametric histogram analysis has been shown to enhance prediction of tumour progression and survival [9]. Similarly, longitudinal DSC perfusion assessment for glioma surveillance has been previously shown to identify malignant transformation of low-grade gliomas [15] including oligodendrogliomas [16, 17]. Studies have applied longitudinal MR perfusion to differentiate between PD and PsP, demonstrating that DSC perfusion parameters such as relative cerebral blood volume (rCBV) are potentially superior to conventional MRI parameters such as trends in enhancing tumour volume [18, 19].

The analysis of multiparametric MRI datasets has been further enhanced by the use of machine learning. Such techniques have been successfully employed for the differentiation of PsP from tumour recurrence in patients with resected GB [20, 21], glioma classification by grade and mutation status [22], prediction of overall survival, molecular subtyping of GB [23, 24] and differentiation of tumour from non-tumour components in HGG [25]. Support vector machine (SVM) classifiers have been successfully applied on multiparametric MRI datasets for the classification between PD and PsP based on perfusion MRI [26]. More recent studies demonstrate improved classification accuracy by including perfusion features in multiparametric MRI datasets for radiomic model analysis [27, 28].

Our aim was to comparatively assess the performance of an SVM classifier trained to differentiate PD from PsP for

different radiomic feature combinations: (I) combined perfusion and structural radiomic features compared to standalone features and (II) longitudinal compared to single time point radiomic features. Furthermore, we aimed to compare the performance of the SVM classifier against radiologists' interpretation.

To the best of our knowledge, this is the first study to examine the potential advantage of applying combined structural and perfusion MRI datasets both on single and longitudinal time points, for SVM enhanced classification between PD and PsP in patients with HGG.

Materials and methods

Study participants

We retrospectively analysed imaging studies of patients with HGG, investigated in our department with perfusion MRI between 2012 and 2018. Institutional research ethics review approval was obtained. The requirement for informed consent was waived by the UCL research ethics committee. We included patients with histologically confirmed primary diagnosis of HGG, availability of DSC MR perfusion on at least one time point following any treatment and of structural MRI before and after each perfusion time point. We excluded patients who did not have histological confirmation of HGG, had inadequate imaging studies prior and after DSC MR perfusion, or underwent surgery and/or antiangiogenic agent treatment (e.g. bevacizumab) between perfusion time points. Two groups were created for analysis: “group I” consisting of all patients with a single DSC MR perfusion time point (total cohort 64 patients) and “group II” consisting of patients with multiple (2 or 3) DSC MR perfusion time points (19 patients).

MRI protocol

Structural MRI studies were acquired on a clinical 3T MRI system (Magnetom Prisma Siemens Healthcare, Erlangen, Germany) including the following sequences: T2 fluid-attenuated inversion recovery (FLAIR) [TR/TE/IR 6500/88/2130 ms, slice/gap 5/6.5 mm, field of view (FOV) 165 × 220 mm²]; T2WI [TR/TE 4610/99.5, slice/gap 5/1.5 mm, FOV 210 × 210 mm²]; T1WI [TR/TE 415/20 ms, slice/gap 5/1.5 mm, FOV 210 × 210 mm²]; DWI [TR/TE 3700/55 ms, slice/gap 4/5 mm, acquisition matrix 192×192, FOV 220×220 mm²]; T1WI post-contrast (Dotarem, Guerbet, Villepinte, France) [TR/TE 6.8/450 ms, slice/gap 5/6.5 mm, acquisition matrix 256×256, FOV 217×240 mm²].

Perfusion-weighted MRI studies were acquired on the same system using a standard dynamic susceptibility weighted contrast perfusion MR imaging protocol consisting of a gradient echo-EPI sequence, TR/TE 1370/30, slice/gap 4/5.2

mm, field of view (FOV) 220x220, acquisition matrix 128x128, flip angle 65, scan time: 2 min 20 s. EPI data were acquired following injection of a 0.1-mmol/kg body weight bolus of gadoterate meglumine (Dotarem, Guerbet, Villepinte, France) followed by a 20-ml bolus of saline, both at a constant rate of 5 ml/s. Pre-load with half-dose gadolinium was applied. All external MRI scans included in the study were based on similar protocols.

DSC perfusion analysis

Postprocessing of DSC PWI studies was performed using Olea Sphere 3.0 (Olea Sphere®

3.0, Olea Medical®) following correction for patient motion using the built-in software feature. The arterial input function was selected automatically using a cluster analysis algorithm [29] and manually corrected in cases of discrepancy with the anatomical images. Deconvolution-based perfusion parameters were calculated using Bayesian probabilistic methods, following contrast leakage correction with the built-in software feature [30]. Relative cerebral blood volume (rCBV) and relative cerebral blood flow (rCBF) maps were calculated on DSC MR perfusion studies. Normalised values (z-scores) of both rCBV and rCBF were calculated respective to the ipsilateral basal ganglia. Perfusion normalisation with similar methodology has been previously shown to decrease variability of rCBV measurements. Based on previous studies, the basal ganglia have been used for normalisation to allow automated segmentation and reduce variability related to perfusion variations within the white matter and user-dependent selection of regions of interest [31–33].

Treatment response assessment

Treatment response assessment included the following categories: progressive disease (PD), pseudoprogression (PsP), stable disease (SD), partial response (PR) and complete response (CR). Lesion classification was based on histology following repeat surgery or biopsy where available (13 of a total cohort of 64 cases: 20%). In the remaining cases, classification was based on prolonged radiological and clinical surveillance. Radiological surveillance was based on the course of enhancing lesions on prolonged longitudinal MRI. Clinical surveillance was based on the final outcome of the local neuro-oncology multidisciplinary team meeting which assessed all available clinical and radiological data at the latest available time point. This classification was considered as expert consensus ground truth. No complete response cases were encountered in our cohort. The few preliminary partial response cases encountered (4), converted to other categories during surveillance. Therefore, training of the SVM classifier was based on three categories: PD, PsP and SD (3-class problem).

Image co-registration, segmentation

Our methodology is outlined in Fig. 1. A common 3D space was created for each patient using axial T1, post contrast T1W and FLAIR images of all time points. This was based on a previously reported method for affine registration following log-transformation, normalisation, bias field correction and intensity matching of the skull-stripped images [34]. Images from all available time points were resampled to a common space and subtraction datasets of normalised maps were created.

Segmentation was conducted by a neuroradiologist with 2 years' experience (L.S), and results were supervised independently by a neuroradiologist with more than 10 years' experience in brain tumour imaging (S.B). Areas of enhancing tissue on post-contrast T1W images and non-enhancing T2 hyperintensity were manually segmented for all available time points. Segmentation was based on T2 FLAIR images overlaid on post contrast T1W images, excluding any resection cavities and areas of macroscopic necrosis (www.itksnap.org) [35].

Feature extraction

The segmentation masks were applied to the common multiparametric space and utilised for feature extraction for all patients. Imaging features were derived from multiple sequences including T1 (pre- and post-contrast), T2, FLAIR, rCBV and rCBF at all available time points. Total extracted features were 130 for the single time point group and 110 for the multiple time point group. For the multiple time point group, subtracted features between different time points were calculated (subtractions included 2-1, 3-2, 3-1). These included longitudinal changes in signal intensity on pre- and post-contrast sequences on structural MRI studies, as well as longitudinal differences in rCBV and rCBF values on consecutive DSC MR perfusion studies. Distribution and textural features over the normalised subtraction images were extracted and used as features for automated classification. A least absolute shrinkage and selection operator (LASSO) framework [36] was applied for feature selection preceding the application of the SVM classifier for automated lesion classification (PD, PsP, SD).

SVM classifier

A binary SVM with radial basis function (RBF) kernel was constructed using all selected features. The code used to generate results is freely available at <https://github.com/csudre/LongPerfusion>. The single time point SVM training dataset was based on all available structural and perfusion MRI studies at a single time point. The longitudinal SVM training dataset included all available normalised images form

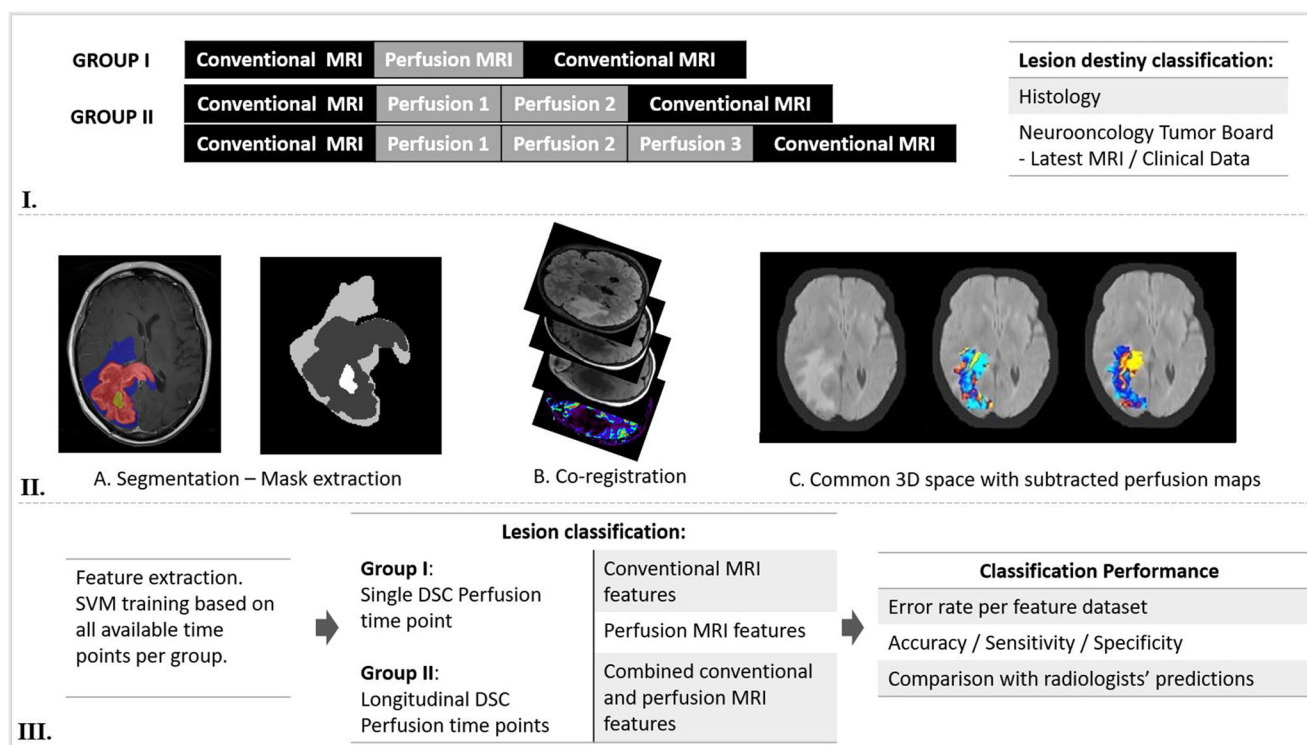


Fig. 1 Summary of methodology and radiomics workflow. I. Study participants were separated into two groups: group I with a single DSC MR perfusion time point and group II with multiple (2 or 3) DSC MR perfusion time points. All patients had structural MRI prior and following each perfusion time point. Each imaging time point was classified as progressive disease (PD), pseudoprogression (PsP) or stable disease (SD). Histology was used as ground truth for lesion classification where available. In cases without histological confirmation, lesion classification was based on the final outcome of the local neuro-oncology multidisciplinary team meeting which assessed all available serial radiological surveillance studies as well as clinical data at the latest available time point. This was considered as the expert consensus ground truth. II. Lesion areas were identified and segmented including hyperintensity on FLAIR (blue),

contrast enhancement (red) and necrosis (yellow—excluded). Segmentation masks were exported. A common 3D space was created for each patient using axial T1, post contrast T1W and FLAIR images from every time point, following log-transformation, normalisation, bias field correction and intensity matching of the skull-stripped images. The perfusion maps corresponding to the extracted masks were co-registered on the common 3D space. III. Feature extraction and SVM training were based on different combinations of feature datasets (structural, perfusion and combined) and perfusion time points (single, longitudinal). Classification performance was assessed by calculating error rates and accuracy/sensitivity/specificity of classification for each feature dataset. SVM classification results were compared to radiologists' predictions

structural and perfusion MRI studies, at every time point (time points 1, 2, and in some cases 3). K-fold cross validation was employed for SVM training and validation, and results were averaged over 250 two-fold cross-validation iterations [37]. Separate classification steps were performed and included “one vs all” dichotomous classifications between: (a) SD vs (PD and PsP) and (b) PD vs (SD and PsP). Final SVM analysis utilised both steps and resulted in a final classification for each case (SD, PD or PsP).

Diagnostic performance

To assess diagnostic performance, error rates and sensitivity/specificity/accuracy were calculated for SVM classification based on different feature datasets (structural, perfusion, combined), different classification steps and different time points (single and multiple perfusion time points). Comparisons of

classification accuracy were performed based on these metrics for both groups.

Consequently, radiological reports were extracted for comparison. These were based on structural and perfusion MRI studies for all patients, reported by a team of three senior neuroradiologists (at least 7 years' experience in neuro-oncology imaging and core members of the multidisciplinary neuro-oncology board). Radiological reports included assessment of disease status for each time point and formed the basis for “radiologist classification”. Radiologists had access to all available structural and perfusion imaging studies at every time point and were blinded to lesion classification at the time of reporting.

Comparison between radiological and SVM classification performance was based on sensitivity, specificity, and accuracy of classification. Statistical significance was assessed using McNemar's statistical test based on methodology employed in previous studies [38, 39].

Results

Study participants - lesion classification

The final analysed cohort included 64 participants (age 48.5 ± 12.8 [mean, SD], 24 female). All time points were selected following completion of initial chemoradiation therapy. From an initial group of 187 patients with DSC MRI exams, we excluded 123 cases due to histological diagnosis other than high-grade glioma, unavailable or inadequate structural imaging studies prior to and after DSC MR perfusion, or inadequate information on lesion histology and patient treatment. All included patients had histologically proven HGG, at least one DSC MR perfusion time point, as well as structural MRI preceding and following each perfusion time point. The final cohort comprised of 64 patients: 45 cases with a single time point and 19 cases with multiple DSC time points. Study participants were separated into two groups: group I (single DSC time point: 64 patients) and group II (multiple DSC time points: 19 patients). A flowchart outlining patient selection is provided (Supplementary Figure 1).

Patient demographics and clinical characteristics are summarised in Table 1. Lesion histology is available in supplementary material (Supplementary Tables 1 and 2). The single time point group (group I: 64 patients) included 43 patients with GB (WHO grade 4), 14 patients with anaplastic astrocytoma (grade 3) and 7 patients with oligodendroglioma (grade 3). The multiple time point group (group II 19 patients) included 14 patients with GB, 3 patients with anaplastic astrocytoma and 2 patients with anaplastic oligodendroglioma.

In total, we included 269 complete structural MRI and 92 DSC MR perfusion studies. The time interval between the initial and final imaging studies during surveillance was (mean, SD [95% CI] days): group I (201, 159 [182–220]), group II (208, 170 [172–243]).

Treatment response assessment and lesion classification are summarised in supplementary material (Supplementary Tables 1 and 2). Lesion classification per group included: group I (single time point): 37 PD, 13 PsP and 14 SD cases and group II (multiple time points): 8 PD, 5 PsP and 6 SD cases.

SVM feature selection—classification error rates

Classification performance of radiomic features was assessed via an integrative analysis. Features with the highest predicting accuracy were similar in both groups. The best performance was identified for subtracted values of these features in the multiple time point group including differences in the 25th percentile (P25) of rCBF (Diff ZrBF First Quartile), rCBV (Diff ZrBV Correlation), T2 kurtosis (Diff T2 Kurtosis) and subtracted signal intensity on T1 post contrast images (Diff T1Gad sum average).

SVM classification error rates were calculated for both groups following multiple iterations, to allow comparison between different datasets. In an exploratory way, classification performance was also assessed for different combinations of lesion status as follows: (PsP/SD vs PD and SD vs PsP/PD).

Classification results for group I (single time point, 64 patients) are provided in supplementary material

Table 1 Patient demographics, tumour histology and lesion classification

Patient population – tumour type	Group II (multiple DSC time points)	Group I (single DSC time point)
Total number of patients	19	64
Sex ratio (M/F)	11/8	40/24
Mean age	45	48.5
Tumour type		
Glioblastoma (WHO grade 4)	13	43
Anaplastic astrocytoma (WHO grade 3)	4	14
Anaplastic oligodendroglioma (WHO grade 3)	2	7
Progressive disease (PD)	8	37
Pseudoprogression (PsP)	5	13
Stable disease (SD)	6	14
Surgical treatment		
Gross total resection, <i>N</i> (%)	13 (68%)	38 (60%)
Sub-total resection, <i>N</i> (%)	2 (11%)	13 (20%)
No surgery, <i>N</i> (%)	4 (21%)	13 (20%)
Follow up interval (days)	208, 170 [172–243]	201, 159 [182–220]
Mean, SD [95% CI]:		
Follow up interval in days	272, 218 [149–395]	216, 176 [61–370]
Mean, SD [95% CI] [PsP cases]		

(Supplementary Figure 2). In this group, the combination of structural and perfusion features outperformed both standalone perfusion and structural feature datasets for the clinically relevant classification of PD versus PsP/SD (median error rates: combined structural and perfusion features 2%, perfusion features 4%, structural features 15.6%. Mean error rates: 23%, 27% and 28% respectively).

Classification in group II (multiple time points, 19 patients) was performed via sampling all combinations of different time points and feature datasets (structural and perfusion) for each time point. When a single perfusion time point was employed (Fig. 2), the combination of perfusion and structural features resulted in lower final classification error rates compared to standalone modalities. Specifically, for the clinically relevant classification of PD vs PSP/SD, the lowest error rate was achieved when the combination of structural and perfusion features was employed (median error rate: 1.6%, mean error rate: 5%).

In group II, classification performance was also assessed for subtracted datasets from different perfusion time point combinations (i.e., first and second time points, first and third time points etc.). Overall, the input of combined perfusion and structural features consistently resulted in lower classification error rates for all time point combinations. Classification based on the first time point resulted in a mean error rate of 10.5%, which was improved to 9.8% when a second time point was introduced (mean interval of 88.2 days). The introduction of an additional, third time point

(mean interval of 287.5 days from the initial time point) further reduced the mean error rate to 9.3%. The lowest achievable error rate for stepwise classifications was observed for combined subtracted features between the third and first perfusion time points (median error rate: 0.4%, mean error rate: 0.7%).

Comparison between radiological reports and SVM-based classification

We calculated sensitivity, specificity and accuracy of lesion classification by radiologists and the SVM classifier. Results are summarised in Table 2. In group II (multiple DSC time points), the SVM outperformed radiologists' classification. The sensitivity/specificity/accuracy for the SVM classification was 100/91.67/94.7% (analysis based on the first perfusion time point) and 85.71/100/94.7% (longitudinal analysis based on multiple time points) compared to 60/78/68% and 70/90/84.2% for the respective radiologist classifications. In group I (single perfusion time point), the SVM also exceeded radiologist classification performance, albeit by a smaller margin and resulted in sensitivity/specificity/accuracy of 86.49/75.00/81.53% (SVM) compared to 75.7/68.9/73.84% (radiologists). Combined perfusion and structural features consistently outperformed standalone datasets for SVM-based classification for all time point combinations.

Statistical assessment using McNemar's statistical test rejected the null hypothesis of equal performance between

Fig. 2 Box plots illustrating the calculated error rate for the clinically relevant SVM classification (SD/PsP vs PD) for group II (multiple DSC perfusion time points), per feature category. The combination of structural and perfusion features outperformed standalone structural or perfusion feature datasets yielding the lowest classification error rate (median error rate: 1.6%, mean error rate: 5%). Error rate differences were statistically significant (Wilcoxon/Kruskal-Wallis test: p value = 0.0001)

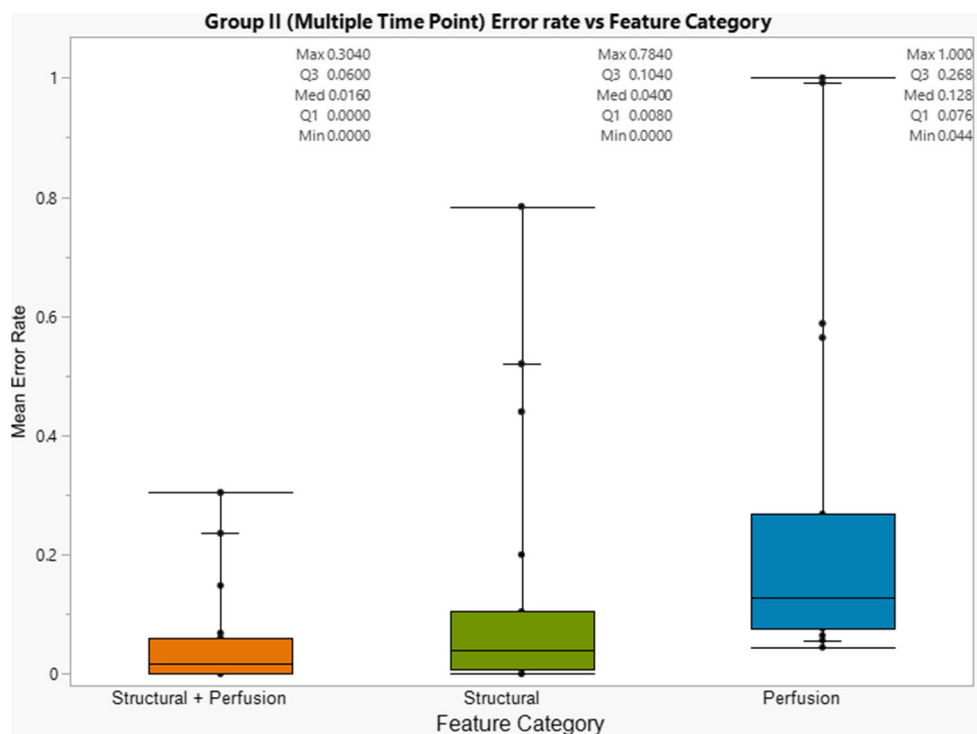


Table 2 SVM and radiologist classification performance assessment

	SVM			Radiologists		
	Group I (single time point)	Group II (first time point analysis)	Group II (longitudinal analysis)	Group I (single time point)	Group II (first time point analysis)	Group II (longitudinal analysis)
Sensitivity (%)	86.49	100	85.71	75.7	60	70
Specificity (%)	75.00	91.67	100	68.9	78	90
Accuracy (%)	81.53	94.7	94.7	73.84	68	84.2

McNemar's test, SVM vs radiologist classifications (p value): group I: p value = 0.041, group II (first time point analysis): p value = 0.034, group II (longitudinal analysis): p value = 0.025

the radiologists and the SVM classifier (p value < 0.05 for the respective comparisons, provided in Table 2).

Discussion

This study comparatively assessed the performance of an SVM classifier for the differentiation between PD, SD and PSP during post-treatment surveillance of patients with high-grade glioma (HGG). Our results demonstrate improved SVM classification performance following the application of combined perfusion and structural MRI features and the introduction of longitudinal perfusion time points. Furthermore, our results indicate that the optimal SVM classifier outperforms radiologists' interpretation in our cohort.

Optimal SVM classification performance was observed when longitudinal perfusion studies were analysed based on combined perfusion and structural features, exceeding radiologists' classification performance in both patient groups. Applying the classifier across two groups (group I and II with either 1 or 2–3 DSC points respectively), we demonstrate that the sensitivity/specificity/accuracy of the SVM were superior to radiologists'. Specifically, for SVM-based classification, these metrics are 100/91.67/94.7% (first perfusion time point analysis) and 85.71/100/94.7% (longitudinal analysis), compared to 60/78/68% and 70/90/84.2% for the respective radiologists' classifications. To the best of our knowledge, this is the first study to comparatively assess SVM classification performance based on single and longitudinal perfusion time points using combinations of structural and perfusion feature datasets, and to compare SVM classification performance to that of expert radiologists.

Our findings highlight the potential of multiparametric MRI for determining disease progression and are in accordance with previous studies. Several studies utilised perfusion MRI parameters for discrimination between disease progression and treatment-related effects predominantly evaluating either mean rCBV or maximum rCBV, however producing inconsistent results [8, 40–42]. A recent meta-analysis reports a pooled specificity and sensitivity of 90% and 88% (95% CI

0.85–0.94; 0.83–0.92) for each study's best-performing DSC MR perfusion parameter [11]. However, in contrast to these studies and similarly to previous studies by Hu et al and Park et al [43], our methodology is not based on predetermined perfusion features or cut-off values during SVM training.

We observed that the combination of perfusion and structural MRI features consistently improved classification performance for both single and longitudinal perfusion time point group analyses. Recent studies examining the identification of pseudoprogression based on a single perfusion time point are consistent with this observation [27, 28, 44].

Extending further than current knowledge, our results indicate that SVM classification based on longitudinal DSC perfusion time points outperforms single time point analysis in predicting lesion destiny, a finding not previously described by similar studies. The longitudinal analysis in our cohort was characterised by improved performance both in terms of classification error rate and sensitivity/specificity/accuracy, although inherent group cofounders may bias direct comparisons.

Specifically, classification performance appeared to be related to the number of the DSC MR perfusion time points. When a single time point was used for SVM training based on combined perfusion and structural features, sensitivity/specificity/accuracy were 86.49/75.00/81.53% respectively (mean error rate 10.5%). The application of a second perfusion time point (mean interval time 88.2 days) resulted in increased classification performance (sensitivity/specificity/accuracy 85.71/100/94.73%, mean error rate 9.8%). The introduction of an additional, third time point (mean interval 287.5 days) resulted in similarly improved classification performance and further reduction in error rate (sensitivity/specificity/accuracy 85.71/100/94.73%, mean error rate 9.3%). The limited number of subjects undergone three longitudinal perfusion scans in our cohort does not allow definite conclusions. However, it is worth noting that the lowest achieved error rate was observed when combined subtracted features of all three time points were applied for SVM classification (median error rate 0.4%, mean error rate 7%).

Another important finding with potential clinical implication is that the SVM outperformed radiologists' classification performance up to 27% in terms of classification accuracy, more profoundly in the multiple time point group. A similar observation is described in a recent study employing radiomic features derived from structural MRI [45]. However, any generalisation of such comparisons would require confirmation by studies based on larger patient cohorts, prospective design and use of external validation.

The identified difference in performance may be attributed to the ability of the SVM algorithm to detect minute composite differences in time and space using simultaneously perfusion and structural imaging parameters, a time-consuming and challenging task for the reporting radiologist. Indeed, the best-performing features included differences in rCBV, rCBF, T2 kurtosis and subtracted signal intensity on post-contrast T1 sequences on longitudinal MRI studies. Similar features have been identified by previous studies as promising for the differentiation of PsP from PD [44–47]. Employing machine learning, Akbari et al. further demonstrated a correlation between such features, namely enhancement on post-contrast T1 and rCBV, with histologically validated tissue characteristics related to PsP and PD [48]. Differences in contrast enhancement and perfusion metrics are routinely employed by radiologists for lesion characterisation using advanced imaging. However, the incorporation of longitudinal changes of multiparametric MRI features in routine clinical reporting poses a time-consuming and challenging task for human readers.

This study has potential limitations. Importantly, the small patient cohort and retrospective design of the study potentially limit generalisability of our results, which should be validated on larger, well-characterised patient cohorts. To accommodate for the small number of included studies and mitigate potential overfitting, K-fold cross validation was employed to derive training and validation datasets. Histological confirmation of lesion destiny was not available for all patients, however, this is rarely available at multiple imaging time points in HGG patients. Therefore, in such cases, we used expert consensus as ground truth. A similar approach has been almost universally employed in similar studies [19, 25–27, 44, 48]. Potential co-occurrence of viable tumour tissue and radiation necrosis within enhancing lesions poses an additional challenge for the characterisation of treatment response. Prolonged serial imaging was employed to account for this and to allow a more objective final lesion characterisation. A limited number of external scans were included, introducing partial heterogeneity of MRI scan protocols. This potential source of bias is well recognised in clinical practice and was addressed by the construction of a common 3D space. In general, accurate comparison of multiple MRI studies of referred patients which are frequently inconsistent in imaging protocol and quality, dictates the creation of a tool to mitigate any bias and allow assessment of disease evolution despite any

technical or quality differences. We believe that our approach is promising to address this need. Overcoming the above limitations is crucial towards clinical application of automated lesion classification based on similar methodologies. Specifically, confirmation of the clinical value of the described approach requires prospective studies incorporating larger patient cohorts, homogeneous scanning parameters, higher percentage of histologically confirmed lesion classifications and external validation of model performance with independent datasets.

In conclusion, our findings indicate that analysis of perfusion and structural MRI data enhanced by machine learning, significantly improves classification between SD, PD and PsP, peaking in performance when multiple perfusion time points are acquired and taken into analysis. This is, to date, the first study designed specifically to allow comparative assessment of classification performance for standalone and combined structural and perfusion MRI features, derived from a single and longitudinal perfusion time points. Automatic classification of lesions by SVM classifiers trained on longitudinal perfusion and structural MRI studies, may also outperform neuroradiological expertise in predicting lesion destiny. Our preliminary findings warrant confirmation by larger, ideally prospective, studies.

Abbreviations PD, Progressive disease; PsP, Pseudoprogession; SD, Stable disease; PR, Partial response; CR, Complete response; rCBV, Relative cerebral blood volume; rCBF, Relative cerebral blood flow; SVM, Support vector machine; HGG, High-grade glioma; GB, Glioblastoma; FLAIR, Fluid-attenuated inversion recovery

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00234-021-02719-6>.

Code availability The code used to generate results is freely available at <https://github.com/csudre/LongPerfusion>.

Author's Contribution Conceptualization: L.S, S.B
Literature search: L.S, S.B
Study design: L.S, C.S, S.B
Methodology: L.S, C.S, S.B
Data Collection: L.S, C.S, S.B, L.T
Formal analysis: C.S, L.S
Data Interpretation: L.S, S.B, C.S
Writing – original draft preparation: L.S, C.S, S.B
Figures: L.S, C.S
Writing – review and editing: L.S, C.S, S.B, S.T, J. P-G, L.T, J.C, N.F, P.M, J.R, R.J.
Supervision/ Project administration: S.B

Funding This study was funded by the National Institute of Health Research (NIHR), Biomedical Research Centre UCL/UCLH.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the UCL research ethics committee.

Informed consent A waiver for informed consent was approved by the UCL research ethics committee.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ostrom QT, Gittleman H, Fulop J, Liu M, Blanda R, Kromer C, Wolinsky Y, Kruchko C, Barnholtz-Sloan JS (2015) CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008-2012. *Neuro-oncology* 17(Suppl 4):iv1
- Stupp R, Mason WP, Van Den Bent MJ, Weller M, Fisher B, Taphoorn MJ, Belanger K, Brandes AA, Marosi C, Bogdahn U (2005) Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* 352(10):987–996
- Young RM, Jamshidi A, Davis G, Sherman JH (2015) Current trends in the surgical management and treatment of adult glioblastoma. *Ann Transl Med* 3(9). <https://doi.org/10.3978/2Fj.issn.2305-5839.2015.05.10>
- Ellingson BM, Wen PY, Cloughesy TF (2017) Modified criteria for radiographic response assessment in glioblastoma clinical trials. *Neurotherapeutics* 14(2):307–20. <https://doi.org/10.1007/s13311-016-0507-6>
- Brandesma D, Stalpers L, Taal W, Sminia P, van den Bent MJ (2008) Clinical features, mechanisms, and management of pseudoprogression in malignant gliomas. *Lancet Oncol* 9(5):453–461
- Pope WB, Lai A, Nghiemphu P, Mischel P, Cloughesy T (2006) MRI in patients with high-grade gliomas treated with bevacizumab and chemotherapy. *Neurology* 66(8):1258–1260
- Ananthnarayan S, Bahng J, Roring J, Nghiemphu P, Lai A, Cloughesy T, Pope WB (2008) Time course of imaging changes of GBM during extended bevacizumab treatment. *J Neuro-Oncol* 88(3):339–347
- Barajas RF Jr, Chang JS, Segal MR, Parsa AT, McDermott MW, Berger MS, Cha S (2009) Differentiation of recurrent glioblastoma multiforme from radiation necrosis after external beam radiation therapy with dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. *Radiology* 253(2):486–496
- Cha J, Kim ST, Kim H-J, B-j K, Kim Y, Lee J, Jeon P, Kim K, Kong D-s, Nam D-H (2014) Differentiation of tumor progression from pseudoprogression in patients with posttreatment glioblastoma using multiparametric histogram analysis. *Am J Neuroradiol* 35(7):1309–1317
- Young RJ, Gupta A, Shah AD, Graber JJ, Chan TA, Zhang Z, Shi W, Beal K, Omuro AM (2013) MRI perfusion in determining pseudoprogression in patients with glioblastoma. *Clin Imaging* 37(1):41–49
- Patel P, Baradaran H, Delgado D, Askin G, Christos P, John Tsiouris A, Gupta A (2016) MR perfusion-weighted imaging in the evaluation of high-grade gliomas after treatment: a systematic review and meta-analysis. *Neuro-oncology* 19(1):118–127
- Nael K, Bauer AH, Hormigo A, Lemole M, Germano IM, Puig J, Stea B (2017) Multiparametric MRI for differentiation of radiation necrosis from recurrent tumor in patients with treated glioblastoma. *Am J Roentgenol* 210(1):18–23. <https://doi.org/10.2214/ajr.17.18003>
- Akbari H, Macyszyn L, Da X, Bilello M, Wolf RL, Martinez-Lage M, Biros G, Alonso-Basanta M, O'Rourke DM, Davatzikos C (2016) Imaging surrogates of infiltration obtained via multiparametric imaging pattern analysis predict subsequent location of recurrence of glioblastoma. *Neurosurgery* 78(4):572–580
- Boonzaier NR, Larkin TJ, Matys T, van der Hoorn A, Yan J-L, Price SJ (2017) Multiparametric MR imaging of diffusion and perfusion in contrast-enhancing and nonenhancing components in patients with glioblastoma. *Radiology* 284(1):180–90. <https://doi.org/10.1148/radiol.2017160150>
- Danchaivijitr N, Waldman AD, Tozer DJ, Benton CE, Brasil Caseiras G, Tofts PS, Rees JH, Jager HR (2008) Low-grade gliomas: do changes in rCBV measurements at longitudinal perfusion-weighted MR imaging predict malignant transformation? *Radiology* 247(1):170–178
- Hlaihel C, Guilloton L, Guyotat J, Streichenberger N, Honnorat J, Cotton F (2010) Predictive value of multimodality MRI using conventional, perfusion, and spectroscopy MR in anaplastic transformation of low-grade oligodendrogliomas. *J Neuro-Oncol* 97(1):73–80
- Law M, Oh S, Johnson G, Babb JS, Zagzag D, Golfinos J, Kelly PJ (2006) Perfusion magnetic resonance imaging predicts patient outcome as an adjunct to histopathology: a second reference standard in the surgical and nonsurgical treatment of low-grade gliomas. *Neurosurgery* 58(6):1009–1107
- Steidl E, Müller M, Müller A, Herrlinger U, Hattungen E (2019) Longitudinal, leakage corrected and uncorrected rCBV during the first-line treatment of glioblastoma: a prospective study. *J Neuro-Oncol* 144(2):409–417
- Boxerman JL, Ellingson BM, Jeyapalan S, Elinzano H, Harris RJ, Rogg JM, Pope WB, Safran H (2017) Longitudinal DSC-MRI for distinguishing tumor recurrence from pseudoprogression in patients with a high-grade glioma. *Am J Clin Oncol* 40(3):228–234
- Hu X, Wong KK, Young GS, Guo L, Wong ST (2011) Support vector machine multiparametric MRI identification of pseudoprogression from tumor recurrence in patients with resected glioblastoma. *J Magn Reson Imaging* 33(2):296–305
- Jang B-S, Jeon SH, Kim IH, Kim IA (2018) Prediction of pseudoprogression versus progression using machine learning algorithm in glioblastoma. *Sci Rep* 8(1):1–9
- Sudre CH, Panovska-Griffiths J, Sanverdi E, Brandner S, Katsaros VK, Stranjalis G, Pizzini FB, Ghimenton C, Surlan-Popovic K, Avsenik J (2020) Machine learning assisted DSC-MRI radiomics as a tool for glioma classification by grade and mutation status. *BMC Med Inf Decis Mak* 20(1):1–14
- Kickingeder P, Bonekamp D, Nowosielski M, Kratz A, Sill M, Burth S, Wick A, Eidel O, Schlemmer H-P, Radbruch A (2016) Radiogenomics of glioblastoma: machine learning-based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. *Radiology* 281(3):907–918
- Macyszyn L, Akbari H, Pisapia JM, Da X, Attiah M, Pigrish V, Bi Y, Pal S, Davuluri RV, Roccograndi L (2015) Imaging patterns

- predict patient survival and molecular subtype in glioblastoma via machine learning techniques. *Neuro-oncology* 18(3):417–425
25. Blumenthal D, Artzi M, Liberman G, Bokstein F, Aizenstein O, Bashat DB (2017) Classification of high-grade glioma into tumor and nontumor components using support vector machine. *Am J Neuroradiol* 38(5):908–914
 26. Artzi M, Liberman G, Nadav G, Blumenthal DT, Bokstein F, Aizenstein O, Bashat DB (2016) 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 Neuro-Oncol* 127(3):515–524
 27. Kim JY, Park JE, Jo Y, Shim WH, Nam SJ, Kim JH, Yoo R-E, Choi SH, Kim HS (2018) Incorporating diffusion-and perfusion-weighted MRI into a radiomics model improves diagnostic performance for pseudoprogression in glioblastoma patients. *Neuro-oncology* 21(3):404–414
 28. Elshafeey N, Kotrotsou A, Hassan A, Elshafei N, Hassan I, Ahmed S, Abrol S, Agarwal A, El Salek K, Bergamaschi S (2019) Multicenter study demonstrates radiomic features derived from magnetic resonance perfusion images identify pseudoprogression in glioblastoma. *Nat Commun* 10(1):3170
 29. Mouridsen K, Christensen S, Gyldensted L, Østergaard L (2006) Automatic selection of arterial input function using cluster analysis. *Magn Reson Med* 55(3):524–531
 30. Boutelier T, Kudo K, Pautot F, Sasaki M (2012) Bayesian hemodynamic parameter estimation by bolus tracking perfusion weighted imaging. *IEEE Trans Med Imaging* 31(7):1381–1395
 31. Ellingson BM, Zaw T, Cloughesy TF, Naeini KM, Lalezari S, Mong S, Lai A, Nghiemphu PL, Pope WB (2012) Comparison between intensity normalization techniques for dynamic susceptibility contrast (DSC)-MRI estimates of cerebral blood volume (CBV) in human gliomas. *J Magn Reson Imaging* 35(6):1472–1477
 32. Sakaie KE, Shin W, Curtin KR, McCarthy RM, Cashen TA, Carroll TJ (2005) Method for improving the accuracy of quantitative cerebral perfusion imaging. *J Magn Reson Imaging* 21(5):512–519
 33. Carroll TJ, Haughton VM, Rowley HA, Cordes D (2002) Confounding effect of large vessels on MR perfusion images analyzed with independent component analysis. *Am J Neuroradiol* 23(6):1007–1012
 34. Sudre CH, Cardoso MJ, Ourselin S, Initiative ADN (2017) Longitudinal segmentation of age-related white matter hyperintensities. *Med Image Anal* 38:50–64
 35. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31(3):1116–1128
 36. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Methodol* 58(1):267–288
 37. Wong T-T (2015) Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recogn* 48(9):2839–2846
 38. Juntu J, Sijbers J, De Backer S, Rajan J, Van Dyck D (2010) Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in T1-MRI images. *J Magn Reson Imaging* 31(3):680–689
 39. Salzberg SL (1997) On comparing classifiers: pitfalls to avoid and a recommended approach. *Data Min Knowl Disc* 1(3):317–328
 40. Sugahara T, Korogi Y, Tomiguchi S, Shigematsu Y, Ikushima I, Kira T, Liang L, Ushio Y, Takahashi M (2000) Posttherapeutic intraaxial brain tumor: the value of perfusion-sensitive contrast-enhanced MR imaging for differentiating tumor recurrence from nonneoplastic contrast-enhancing tissue. *Am J Neuroradiol* 21(5):901–909
 41. Gasparetto EL, Pawlak MA, Patel SH, Huse J, Woo JH, Krejza J, Rosenfeld MR, O'Rourke DM, Lustig R, Melhem ER (2009) Posttreatment recurrence of malignant brain neoplasm: accuracy of relative cerebral blood volume fraction in discriminating low from high malignant histologic volume fraction. *Radiology* 250(3):887–896
 42. Hu LS, Baxter L, Smith K, Feuerstein B, Karis J, Eschbacher J, Coons S, Nakaji P, Yeh R, Debbins J (2009) Relative cerebral blood volume values to differentiate high-grade glioma recurrence from posttreatment radiation effect: direct correlation between image-guided tissue histopathology and localized dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging measurements. *Am J Neuroradiol* 30(3):552–558
 43. Park JE, Kim HS, Goh MJ, Kim SJ, Kim JH (2015) 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):792–802
 44. Akbari H, Rathore S, Bakas S, Nasrallah MP, Shukla G, Mamourian E, Rozycki M, Bagley SJ, Rudie JD, Flanders AE (2020) Histopathology-validated machine learning radiographic biomarker for noninvasive discrimination between true progression and pseudo-progression in glioblastoma. *Cancer* 126(11):2625–2636
 45. Sun Y-Z, Yan L-F, Han Y, Nan H-Y, Xiao G, Tian Q, Pu W-H, Li Z-Y, Wei X-C, Wang W (2021) Differentiation of pseudoprogression from true progression in glioblastoma patients after standard treatment: a machine learning strategy combined with radiomics features from T1-weighted contrast-enhanced imaging. *BMC Med Imaging* 21(1):1–12
 46. Gao X-Y, Wang Y-D, Wu S-M, Rui W-T, Ma D-N, Duan Y, Zhang A-N, Yao Z-W, Yang G, Yu Y-P (2020) Differentiation of treatment-related effects from glioma recurrence using machine learning classifiers based upon pre- and post-contrast T1WI and T2 FLAIR subtraction features: a two-center study. *Cancer Manag Res* 12:3191–3201
 47. Bani-Sadr A, Eker OF, Berner L-P, Ameli R, Hermier M, Baritault M, Meyronet D, Guyotat J, Jouanneau E, Honnorat J (2019) Conventional MRI radiomics in patients with suspected early- or pseudo-progression. *Neuro-oncology Adv* 1(1):vdz019
 48. Jang B-S, Park AJ, Jeon SH, Kim IH, Lim DH, Park S-H, Lee JH, Chang JH, Cho KH, Kim JH (2020) Machine learning model to predict pseudoprogression versus progression in glioblastoma using MRI: a multi-institutional study (KROG 18-07). *Cancers* 12(9):2706

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.