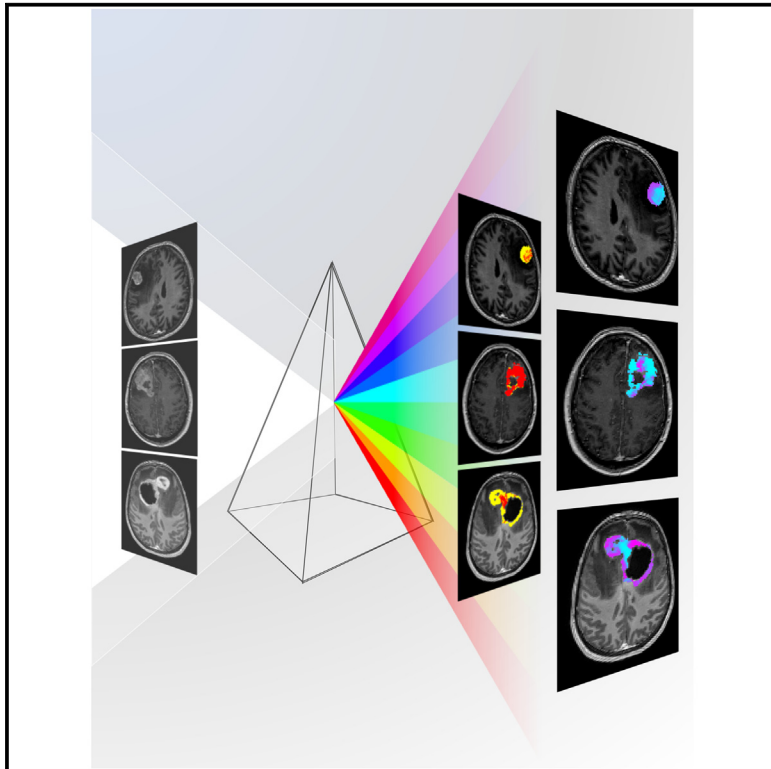**Article**

An accessible deep learning tool for voxel-wise classification of brain malignancies from perfusion MRI

Graphical abstract

Authors

Alonso Garcia-Ruiz, Albert Pons-Escoda, Francesco Grussu, ..., Tyler M. Seibert, Carlos Majos, Raquel Perez-Lopez

Correspondence

rperez@vhio.net

In brief

Diagnosing brain tumors can be challenging, even with multiparametric MRI that includes vascular tumor evaluation through perfusion imaging. In this work, Garcia-Ruiz et al. developed a deep learning-based tool that leverages the full spatial and temporal information of perfusion MRI, outperforming conventional methods.

Highlights

- Voxel-wise approach enables training neural networks with limited patient cohorts

- DISCERN facilitates brain tumor classification to aid medical decisions

- DISCERN is a user-friendly tool designed for use with clinical perfusion MRI

- DISCERN enables accurate brain tumor diagnosis, surpassing conventional metrics

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

An accessible deep learning tool for voxel-wise classification of brain malignancies from perfusion MRI

Alonso Garcia-Ruiz,[1,10] Albert Pons-Escoda,[2,3,10] Francesco Grussu,[1,10] Pablo Naval-Baudin,[2] Camilo Monreal-Aguero,[1] Gretchen Hermann,[4] Roshan Karunamuni,[4] Marta Ligero,[1] Antonio Lopez-Rueda,[5] Laura Oleaga,[5] M. Álvaro Berbís,[6] Alberto Cabrera-Zubizarreta,[7] Teodoro Martin-Noguerol,[7] Antonio Luna,[7] Tyler M. Seibert,[4,8,9] Carlos Majos,[2,3] and Raquel Perez-Lopez[1,11,*]

[1]Radiomics Group, Vall d'Hebron Institute of Oncology (VHIO), 08035 Barcelona, Spain
[2]Radiology Department, Bellvitge University Hospital, 08907 Barcelona, Spain
[3]Neuro-Oncology Unit, Institut d'Investigacio Biomedica de Bellvitge (IDIBELL), 08907 Barcelona, Spain
[4]Radiation Medicine Department and Applied Sciences, University of California, San Diego, La Jolla, CA 92093, USA
[5]Radiology Department, Hospital Clínic de Barcelona, 08036 Barcelona, Spain
[6]Radiology Department, HT Medica, Hospital San Juan de Dios, 14012 Cordoba, Spain
[7]Radiology Department, HT Medica, 23008 Jaen, Spain
[8]Radiology Department, University of California, San Diego, La Jolla, CA 92093, USA
[9]Bioengineering Department, University of California, San Diego, La Jolla, CA 92093, USA
[10]These authors contributed equally
[11]Lead contact
*Correspondence: rperez@vhio.net
https://doi.org/10.1016/j.xcrm.2024.101464

SUMMARY

Noninvasive differential diagnosis of brain tumors is currently based on the assessment of magnetic resonance imaging (MRI) coupled with dynamic susceptibility contrast (DSC). However, a definitive diagnosis often requires neurosurgical interventions that compromise patients' quality of life. We apply deep learning on DSC images from histology-confirmed patients with glioblastoma, metastasis, or lymphoma. The convolutional neural network trained on ~50,000 voxels from 40 patients provides intratumor probability maps that yield clinical-grade diagnosis. Performance is tested in 400 additional cases and an external validation cohort of 128 patients. The tool reaches a three-way accuracy of 0.78, superior to the conventional MRI metrics cerebral blood volume (0.55) and percentage of signal recovery (0.59), showing high value as a support diagnostic tool. Our open-access software, Diagnosis In Susceptibility Contrast Enhancing Regions for Neuro-oncology (DISCERN), demonstrates its potential in aiding medical decisions for brain tumor diagnosis using standard-of-care MRI.

INTRODUCTION

Differential diagnosis between the most common brain malignancies (i.e., glioblastoma multiforme [GBM], brain metastasis from solid tumors, and primary CNS lymphoma [PCNSL]) represents a clinical unmet need because each of these entities requires a distinct therapeutic approach.[1–3] Although pathology evaluation of tumor samples remains the gold standard for diagnosis, it requires invasive neurosurgical procedures, with a significant risk of complications, and eventually can be confounded by the use of prior medication, such as steroids.[4,5]

To overcome the need for surgery, magnetic resonance imaging (MRI) with intravenous contrast injection is being used as a noninvasive support system for differential diagnosis of brain malignancies. GBM, brain metastasis, and PCNSL represent up to 70% of all malignant brain tumors and more than 80% of contrast-enhancing tumors within the brain.[6] Nevertheless, the enhancing patterns on imaging exhibit a high degree of similarity across these tumor types, making differential diagnosis challenging even for experienced neuroradiologists.[7–9]

The noninvasive characterization of brain tumors on MRI has been an active subject of study for years,[10,11] gaining renewed interest with the application of recent machine learning techniques to imaging data. Among the existing literature, some studies have focused on differentiating GBM from solitary brain metastasis, either with anatomical[12–17] or functional imaging,[10,18–23] whereas other works have concentrated on the identification of PCNSL.[24–27] Of particular significance is dynamic susceptibility contrast (DSC) perfusion MRI, which enables the visualization of vascular characteristics, including vascular density and permeability, and is proving to be valuable for brain tumor diagnosis.[9,10,18,20,22,27–37]

DSC is a quantitative MRI technique that consists of a temporal T2*-weighted acquisition during the administration of a vascular

contrast bolus. The contrast agent causes an initial decrease in the T2*-weighted signal intensity, followed by the signal recovery during washout. In DSC, every voxel in the image yields a unique dynamic curve that describes the temporal evolution of the T2*-weighted signal intensity and reflects local tissue vascular properties. The standard approach to analyze DSC is to derive metrics such as the relative cerebral blood volume (rCBV) and the percentage of signal recovery (PSR), both of which simplify the dynamic signal. The rCBV relates to the tumor vascular density with respect to normal tissue, and the PSR reflects the vascular permeability.[38] Both parameters remain the main focus of DSC analyses for tasks such as tumor type differentiation, grade stratification, and treatment response assessment.[28,39,40] However, among diverse clinically used DSC protocols, the performance of these parameters differs greatly,[38,41,42] which limits its use in routine clinical practice. Although recent multidisciplinary efforts have been made in the community to agree on a common procedure,[43–45] a global standardized DSC workflow is still lacking, regarding variations in the contrast preload settings, imaging parameters, and processing methods, all of which pose additional challenges to the generalizability of the technique and the establishment of reference rCBV/PSR values.

Regarding the use of DSC in aiding brain tumor diagnosis, it is worth noting that, with a few exceptions,[16,21,46] most studies to date have been designed to discriminate between two tumor types or pairs of malignancies. Furthermore, even fewer analyses can be found over large populations or validating external data, thus limiting the generalizability and clinical utility of the presented approaches.

DSC curve normalization and voxel-by-voxel analyses of the full dynamic range can overcome the limitations of conventional metrics and unlock the potential of DSC as a tool for differential diagnosis among the most common brain malignancies. Moreover, the application of deep learning techniques may enable new strategies of analysis and inference for dynamic data. In one of the first works exploring deep learning in DSC data,[47] the authors described an end-to-end pipeline to obtain model-free perfusion metrics from the raw data. They used one-dimensional (1D) convolutional neural networks (CNNs) to characterize the dynamic DSC data of individual voxels. Park et al.[34] developed an autoencoder and a clustering strategy to distinguish different brain areas, including pathologies, from the 1D dynamic data.

In this work, we describe the development and validation of an innovative, comprehensive framework for differential diagnosis of GBM, brain metastasis, and PCNSL, taking advantage of all of the time points of the normalized DSC (nDSC) data. The proposed Diagnosis In Susceptibility Contrast Enhancing Regions for Neuro-oncology (DISCERN) app provides voxel-by-voxel signatures of tumor type and is based on training 1D deep CNNs, with only a small number of pilot scans for a given DSC protocol. In the present study, we demonstrate the feasibility and accuracy of the method and show its superior performance compared to classifiers based on conventional DSC metrics. In addition, our method exhibits on par or higher diagnostic capabilities in comparison to those of expert neuroradiologists. The potential of DISCERN is to aid radiologists in interpreting brain MRI data, thereby enhancing the diagnostic capacity of experienced neuroradiologists and allowing less experienced radiologists to achieve a higher level of proficiency. To facilitate ease of use, we have designed a user-friendly interface with the ultimate goal of minimizing the need for invasive brain biopsies and guiding the selection of optimal treatment strategies in clinical practice.

RESULTS

Population demographics

In this multicenter, retrospective study, we analyzed MRI data from a total of 568 patients with biopsy-confirmed GBM, brain metastasis, or PCNSL. The classification model was developed and tested with 440 patients from a single center, and additional independent cohorts with varying imaging protocols were processed for external validation (Figure 1A). Further information about the study cohorts and classification results can be found in the STAR Methods and in Table S1. No statistically significant differences (p > 0.05) in terms of age and gender were observed between the three tumor types.

Development of a CNN for brain tumor classification

We trained our CNN classifier on a development cohort in which patients were randomly split into training and test sets. For the training set, we included 20 patients with PCNSL and 20 non-PCNSL patients (10 with GBM and 10 with metastasis). This provides a comparable number of voxels for each tumor type and each binary classification (i.e., PCSNL vs. non-PCNSL; GBM vs. metastasis for the non-PCNSL cases). The test set consisted of 25 patients with PCNSL, 85 with metastasis, and 290 with GBM (Figure 1). Approximately 50,000 nDSC curves from voxels of the enhancing region in the training group were used to train the classifier. Each nDSC curve corresponds to a specific spatial voxel of the enhancing tumor.

DISCERN outperforms standard classifiers for brain tumor diagnosis

Following a hierarchical classification approach, our CNN method successfully achieved three-way tumoral classification, outperforming the traditional perfusion metrics (i.e., rCBV and PSR) and standing out from simpler binary classifiers. Specifically, for the task of PCNSL diagnosis, DISCERN achieved superior performance, with an accuracy of 0.94 (95% confidence interval [CI]: 0.93–0.94), and mean rCBV and mean PSR classified patients with accuracies of 0.72 (95% CI: 0.70–0.74) and 0.84 (95% CI: 0.83–0.85), respectively. In a second step, patients not classified as PCNSL were categorized as GBM or brain metastasis. DISCERN differentiated GBM from metastases, with an accuracy of 0.81 (95% CI: 0.79–0.82). By contrast, the performance of standard DSC-derived metrics was markedly lower: rCBV classification achieved an accuracy of 0.69 (95% CI: 0.67–0.71) and a mean PSR of 0.65 (95% CI: 0.63–0.67). In Figure 2, the area under the receiver operating characteristic (AUC ROC) curves of the binary classifiers and 3-way average ROC curves are shown for both the DISCERN classifier and conventional rCBV/PSR. For PCNSL vs. non-PCNSL, the CNN provided a significantly higher AUC than rCBV (DeLong test against rCBV: p = 0.0019, against PSR, p = 0.4615). For GBM vs. metastasis, the CNN resulted in a higher AUC than rCBV and also than PSR (against rCBV: p = 0.0049, against PSR, p < 0.001).

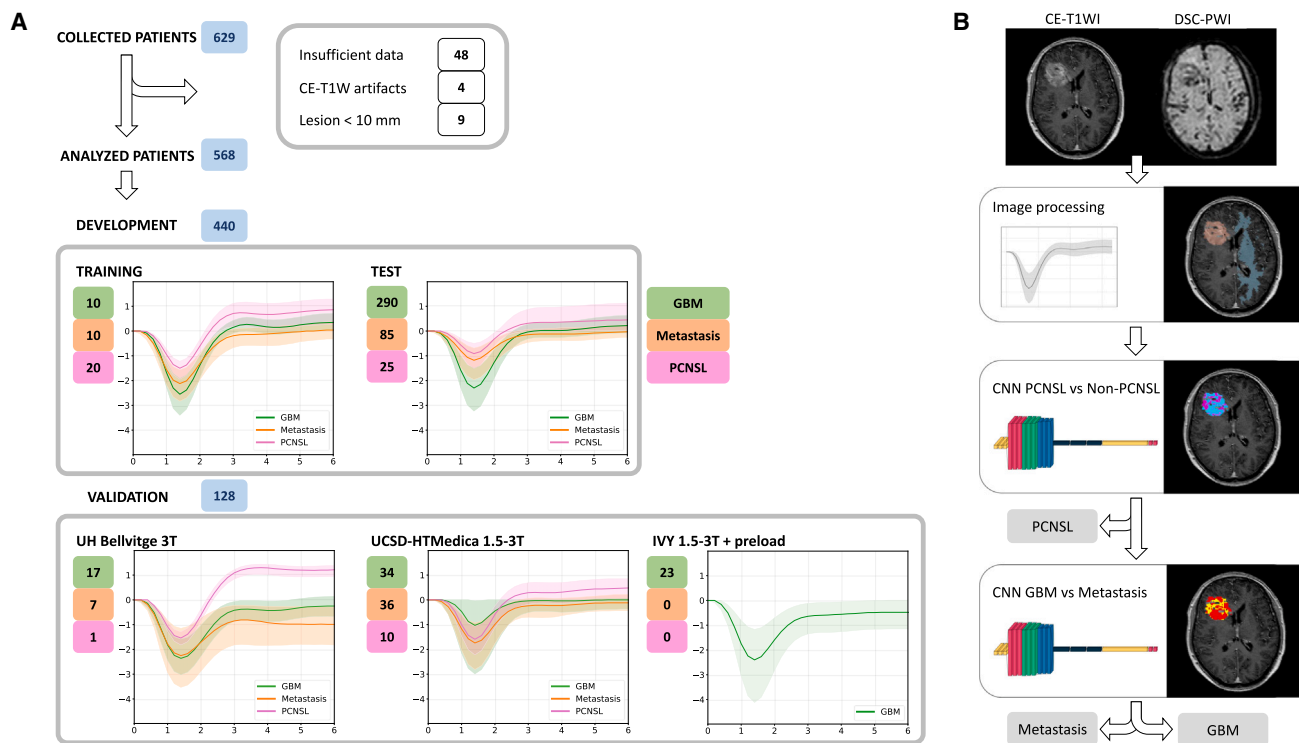# Cell Reports Medicine
Article

Figure 1. Summary of the population and study design

(A) Collected, excluded, and included data for analysis and further split into the development cohort and external validation cohorts. The number of patients for each tumor type and the respective nDSC perfusion distribution are shown for each cohort.

(B) Processing pipeline of the DISCERN app for 3-way tumor classification of DSC data. At the top, the input images of CE-T1WI for automated region of interest selection and DSC for classification are provided. DSC curves are then extracted voxel-wise from the enhancing tumor and normal white matter in the contralateral hemisphere. The dynamic DSC signals from the enhancing tumor are normalized (nDSC) to the white matter. Every nDSC is classified by 2 sequential CNNs, obtaining a probability map and an overall tumor classification. CE-T1WI, contrast-enhanced T1-weighted imaging.

Lastly, we mimicked a real-world clinical scenario in which our diagnostic support system is confronted with brain lesions comprising the three most common malignancies, in this case represented by the internal test dataset (unseen in the training). In this setting, DISCERN achieved an accuracy of 0.78 (95% CI: 0.76–0.79), which is substantially better than the three-way accuracy achieved using mean rCBV, 0.59 (95% CI: 0.57–0.60) and mean PSR, 0.55 (95% CI: 0.53–0.56). Furthermore, the combination of rCBV and PSR into a logistic regression model also yielded poor performance (Table S2). When validating DISCERN, the tool obtained a three-way classification accuracy of 0.71 evaluating 80 scans of patients from external centers, 0.72 on 25 cases from a 3T scanner (same center as the 1.5T development data), and 0.78 on 23 patients with GBM from the Ivy GAP public dataset (Table S3). These data underscore the potential of DISCERN for differentiating among the three most common clinical diagnostic challenges in patients with enhancing brain lesions.

Voxel-wise explainable representation of the CNN decision process

DISCERN provides spatial probability maps of tumor classification, which are then used to obtain a voxel proportion and a patient classification label. In Figure 2A, we present three examples

per tumor type of the voxel-wise probability maps according to the DISCERN classifier. The probability maps are shown overlaid onto the CE-T1W MRI for anatomical references. Overall, the tumor type probability maps are smooth and identify the tumor type with high confidence in most voxels, even when intratumor signal heterogeneity is seen in the contrast-enhanced T1W scan. Voxels exhibiting a high probability of belonging to the incorrect tumor class tend to be located either in the boundary of the enhancing area or around necrotic intratumoral spots. This potentially reflects partial volume (i.e., inclusion of signal from tumor and nontumor areas within a voxel) or intratumoral heterogeneity.

Visual interpretation of the CNN classification

We further sought to implement Class Activation Mapping (CAM) to provide a visual explanation of the DISCERN classification network. The ScoreCAM[48] method yields a normalized score of the contribution of every input to the final classification of a CNN. This allows us to identify the most discriminative time points for nDSC differentiation. ScoreCAM spatial maps were obtained for each binary classification (Figure 3A). The CNN focuses mostly on the bolus passage to classify the central tumor region (middle row for PCNSL vs. non-PCNSL and lower row for GBM vs. metastasis
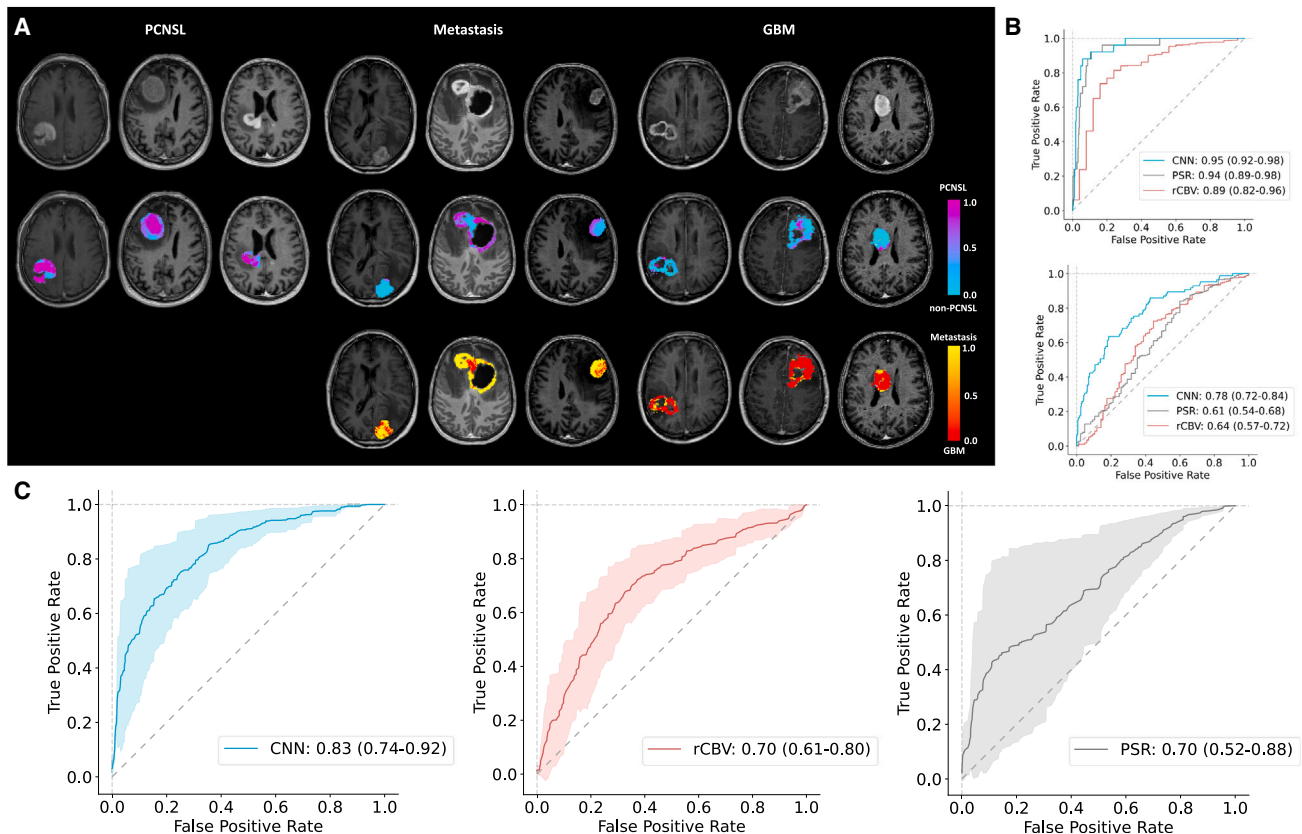
Figure 2. Probability maps and diagnostic performance of DISCERN

(A) Nine cases (3 per tumor type) correctly classified by DISCERN are shown, from left to right: PCNSL, metastasis, and GBM. In the upper row, a representative 2D slice of the CE-T1WI registered to the DSC with overlaid probability maps for PCNSL vs. non-PCNSL (center row) and for GBM vs. metastasis (lower row) of non-PCNSL cases.

(B) ROC curves for binary classifiers PCNSL vs. non-PCNSL (top) and GBM vs. metastasis (bottom) for the proposed CNN, rCBV, and PSR. The CNN provided significantly higher AUC than rCBV for PCNSL vs. non-PCNSL and higher than rCBV and PSR for GBM vs. metastasis.

(C) Three-class ROC curves showing mean and SD of 2-class combinations, from left to right: the proposed CNN, rCBV, and PSR.

differentiation in Figure 3A). In contrast, the bolus passage seems less important for some voxels in surrounding regions. This suggests that the CNN effectively considered the bolus passage as a discerning characteristic, but also that it provides additional tissue perfusion differences compared to the raw perfusion signal (top row in Figure 3A).

The average ScoreCAM values per tumor type and per CNN classifier can be found in Figure 3B (upper row for PCNSL vs. non-PCNSL and lower row for GBM vs. metastasis differentiation). Overall, the sharper signal changes of the nDSC (i.e., steep slopes during contrast arrival and washout) have a higher contribution score. This is especially true for GBM, with greater differences in these time points with respect to the other two tumor types (average nDSC shown in black in Figure 3B). For PCNSL and metastasis, the last part of the signal is also considered important, which can be expected given the overall higher signal magnitude reached in these cases. Importantly, applying 1D CNNs over nDSC signals allowed analysis of the local changes of the signal over time. In this regard, methods that only consider the signal magnitude of specific time points, such as PSR, or a derived measurement such as rCBV, may overlook local nDSC

changes occurring over time that reflect specific physiological traits of the tumor.

A user-friendly app to aid brain tumor diagnosis

The DISCERN app was successfully implemented at the participating institutions for validating the tool in external cohorts, as illustrated in Figure 4 and Table S3. The tool requires approximately 2 min to process a new case and provides a classification outcome, in the form of (1) voxel-wise tumor type probability maps and (2) patient-wise tumor type. In addition, it shows the average nDSC for the enhancing tumor and white matter, as well as a visualization of the segmentation for the user to safely check the process. The mask can be automatically segmented from the enhancing tumor by DISCERN or it can be provided by the user. The DISCERN app provides a classification label with balanced sensitivity and specificity (Youden's index) by default, but a given clinical scenario may require a different classification threshold. To that end, sensitivities and specificities for every threshold are displayed, and the default settings can be changed.

In the benchmark study assessing the diagnostic efficacy of our tool in comparison to two neuroradiologists, notable
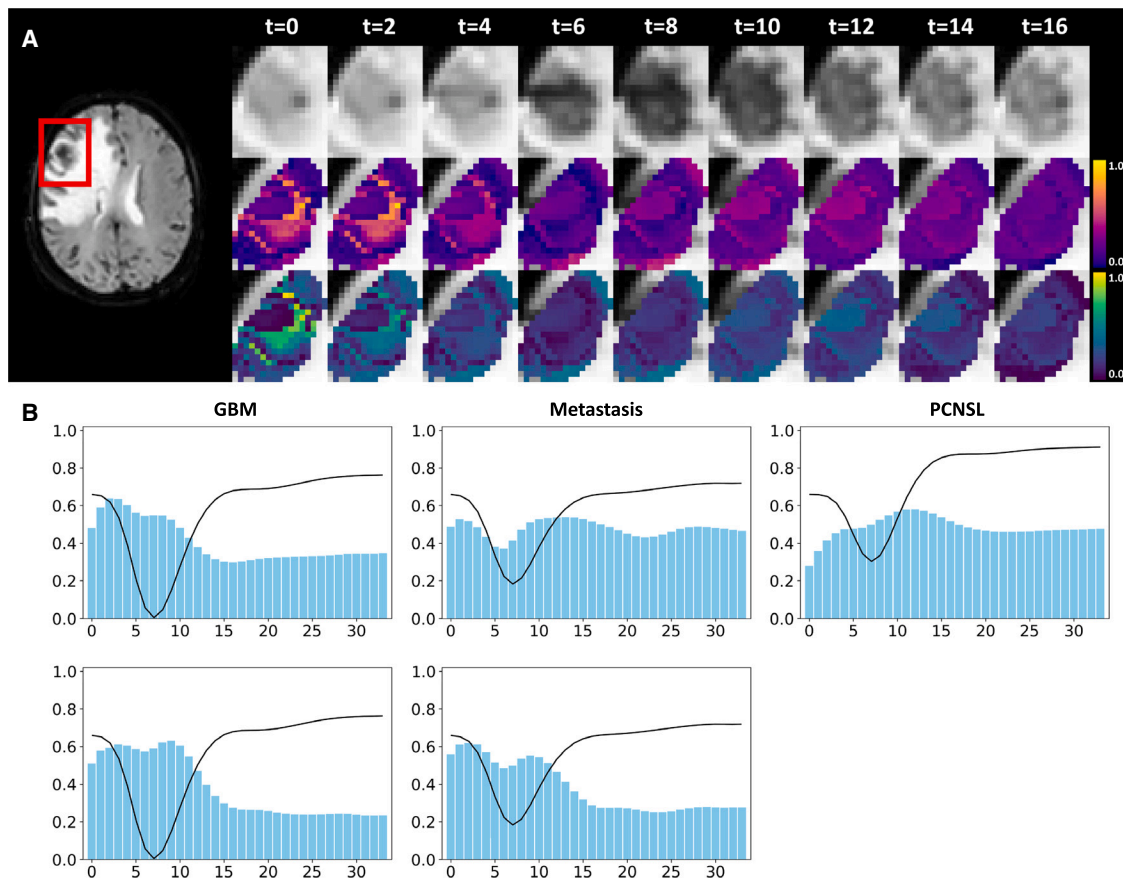
Cell Reports Medicine
Article

Figure 3. Visual interpretation of the CNN classification

(A) ScoreCAM spatial maps to further understand what the most discriminative nDSC time points for classification per voxel are. We show here a representative case of a metastasis in a 2D slice of the DSC (red box at leftmost). In the upper row, consecutive DSC dynamic time points, zoomed in on the lesion. In the center row, spatial importance score maps obtained with ScoreCAM for PCNSL vs. non-PCNSL and for GBM vs. metastasis in the lower row; the score was scaled to sum 1 over all time points in each voxel to observe relative importance in space.

(B) The average importance of each time point obtained from ScoreCAM that contributes to the tumor classification of nDSC curves, for PCNSL vs. non-PCNSL (upper row) and GBM vs. metastasis (lower row) differentiation; average tumor type nDSC in the training set is overlaid as a black solid line.

distinctions were observed between the senior and junior radiologists, achieving accuracies of 0.80 and 0.40, respectively (it is worth noting that a random chance accuracy is 0.33 in the three-way classification scenario). Within this subset, our tool demonstrated a commendable performance with an accuracy of 0.73, effectively identifying all of the instances of PCNSL by relying solely on perfusion-based information. Furthermore, in cases in which the radiologists exhibited elevated levels of uncertainty (16 out of 30 cases), our tool accurately diagnosed 11 out of these 16 instances (Figure S1). These results underscore the potential of the tool to enhance diagnostic accuracy and reliability, particularly in scenarios characterized by increased diagnostic complexity.

DISCUSSION

We present a voxel-wise method for analyzing perfusion scans with CNNs and improve brain cancer diagnosis, built upon prior DSC signal normalization.[49] By applying this method, we were

able to surpass the performance of previous models for noninvasive differential diagnosis of the most frequent malignant brain tumors (i.e., GBM, metastasis, and PCNSL, representing up to 70% of all malignant tumors in the brain[6]), which is critical to define an optimal treatment approach. Notably, our method showcases superior diagnostic capabilities compared to those of neuroradiologists. The potential of DISCERN is to assist radiologists in interpreting brain MRI data, amplifying the diagnostic proficiency of expert neuroradiologists and enabling less experienced counterparts to attain a heightened level of expertise.

Our deep learning framework takes advantage of the large amount of information provided by the thousands of voxel-wise nDSC signals available in each individual DSC MRI scan,[47,50] and achieves optimal performance through training with a limited number of scans from a few patients at fixed DSC protocol (on the order of 30–40 cases). Our approach is particularly appealing for medical imaging applications, in which the design of robust deep learning methods is challenged by the limited number of scans available. In addition, our method
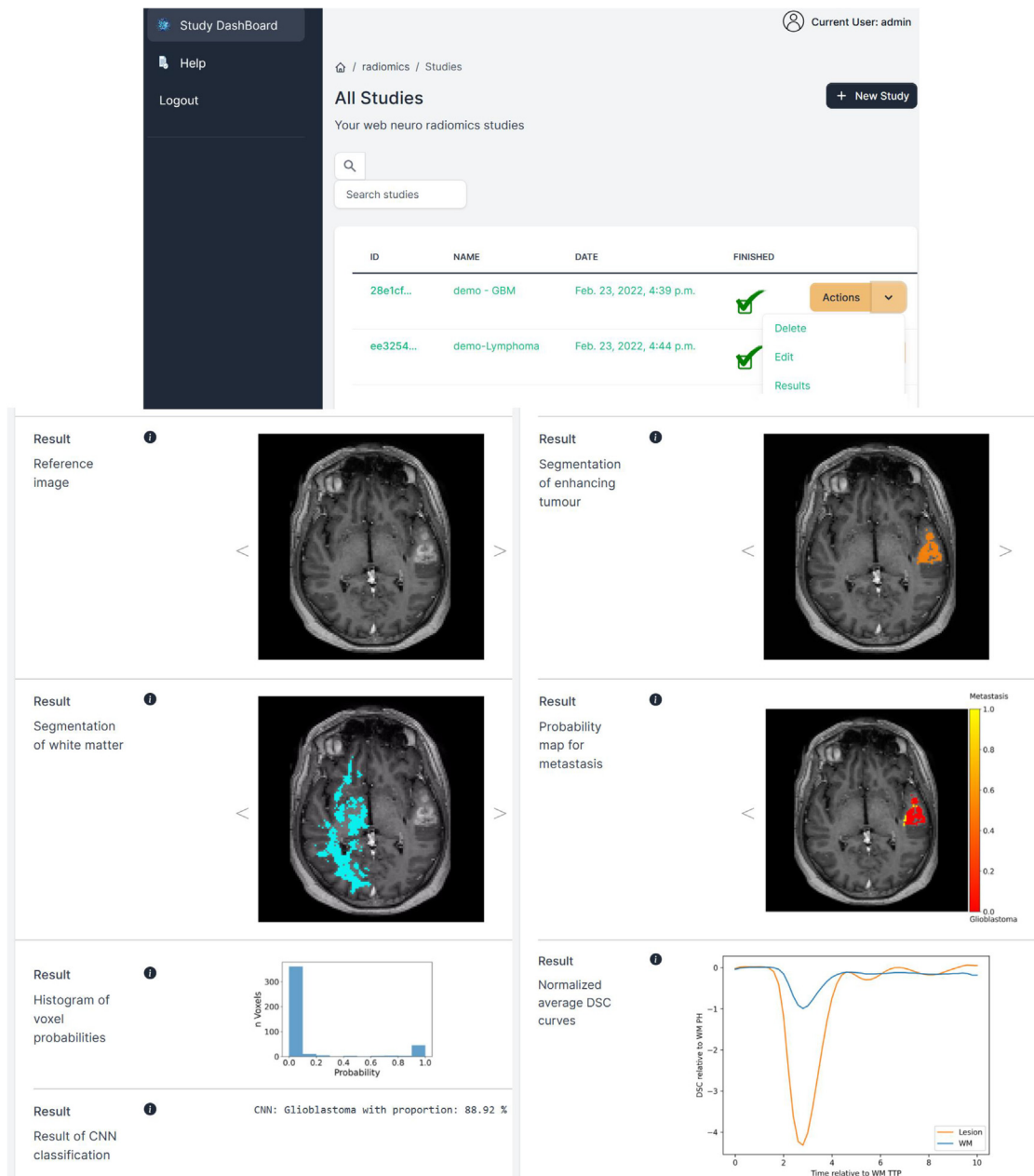
Figure 4. Implementation of the app with an easy-to-use interface

The image illustrates the web interface of the DISCERN app Docker, used to infer results for external validation cohorts. In the uppermost tile, the user dashboard shows the ongoing and finished studies in which to run the pipeline. The next tiles show the results, namely the reference image used for segmentation and respective enhancing tumor and white matter regions, the average nDSC curves of those regions, the tumor probability map and distribution, and a final classification result.

distinguishes between tumor types in a three-way classification task. This can be of particular relevance as a support tool for differential diagnosis in clinical practice, and is a considerable step forward as compared to the current literature, which is dominated by binary classification studies.[11–14,28]

Although three other published works[16,21,46] have included different kinds of three-way classifications among the most common brain tumors (GBM, metastases, and lymphoma), they show some limitations. Liu et al.[16] used the three-way classification as a first step to select the metrics for further pairwise classifications. They considered T1w and T2w anatomical images, with DSC regarded as a future step to improve their results. Tariciotti et al.[46] developed a three-way classification with multiparametric MRI data, using a 2D ResNet model, which would only take advantage of spatial information, but not of dynamic temporal information of DSC, and the reported performance of

Cell Reports Medicine
Article

CellPress
OPEN ACCESS

the model was only moderate. Wang et al.[21] reported a two-step classification scheme similar to ours, and using only conventional DSC metrics, with the main limitation being the lack of validation in external cohorts.

In contrast, we present the largest dataset reported in this context. The voxel-by-voxel approach to the classification of DSC data takes advantage of inherently redundant information. Furthermore, the 1D CNN takes into account the changes in the temporal profile of the bolus passage, which other methods ignore or simplify. It also produces probability maps, facilitating the visual inspection of the spatial distribution of the classification. Finally, we have validated the results in external cohorts, demonstrating the generalizability and potential utility of our findings.

A noninvasive diagnostic support tool is especially relevant when considering PCNSL among potential diagnoses. Corticosteroids are a reasonable therapeutic option for mitigating neurological symptoms secondary to edema in patients with malignant brain tumors. However, early stereotactic biopsy before corticosteroid administration is mandatory when a brain PCNSL is suspected, because medication with steroids can alter the histological pattern of PCNSL.[5] Moreover, PCNSL is highly sensitive to chemoradiotherapy instead of resection, which is contraindicated, as opposed to GBM or metastasis. Therefore, a reliable characterization of the tumor type by imaging is critical to devise the appropriate management of patients.

The DISCERN app provides voxel-wise tumor type probability maps, which are then used to obtain a voxel proportion and a patient classification label. The default Youden's index (tradeoff between sensitivity and specificity) can be changed to the needs of different clinical scenarios. For instance, some medical cases may require a very high specificity for suspected GBM and metastases with respect to PCNSL and, if all evidence supports it, then an additional intervention for a biopsy could be prevented. Therefore, the voxel proportion can be adjusted in the app to match the user's needs.

The presented method successfully achieved three-way tumoral classification, outperforming the traditional perfusion metrics, and standing out from simpler binary classifiers. When tested, our method performed with accuracies of 0.94 for PCNSL identification, 0.81 for differentiation of GBM from metastasis, and 0.78 for three-way classification.

Of note, the DSC protocol used for model development did not include contrast preload. Contrast preload is a common approach described in the literature to achieve a better estimate of the rCBV.[42] However, preload can be undesirable for a number of reasons. First, it delivers a higher contrast dose to the patient. Second, it can introduce potential variability sources, affecting the nDSC signal morphology. As countermeasures, leakage correction and acquisition parameters that minimize the T1 effect, such as low flip angle, have been shown to effectively yield reliable rCBV estimates without preload.[42,44] In this study, we only had access to retrospective data with high flip angle; consequently, rCBV was estimated with leakage correction, obtaining results comparable to those of PSR. The combination of both rCBV and PSR in logistic regression was explored for completeness, but it did not improve the results of the individual parameters. It is noteworthy that even though the DSC proto-

col used a nonoptimal flip angle for PSR or rCVB quantification, DISCERN achieved a good level of discriminatory performance. This was further confirmed by validating the results using heterogeneous external validation data, indicating a high level of robustness in the method.

A key feature of our CNN approach is the computation of voxel-wise spatial representations of perfusion curve characteristics in the form of maps describing the probability of a voxel to belong to a specific tumor type. Such spatial probability maps provide an explainable representation of the CNN decision process and may enable further studies of intratumor heterogeneity, making them an appealing tool for integrative multi-omics research and also of potential clinical interest to plan surgical procedures. Few studies have applied deep learning to voxel-wise DSC signals in neuro-oncological applications. A recent study[34] used a deep autoencoder to derive a set of five descriptors of DSC that could differentiate between pairs of tumor types. However, the reconstructed time signals from such a minimal set of descriptors, in contrast to the original signal, produce a smooth perfusion curve morphology, which may omit relevant details for diagnostic applications.

In conclusion, the presented CNN framework for three-class brain tumor classification based on voxel-wise DSC signal analysis is feasible and outperforms classifiers built on conventional rCBV and PSR metrics. The method can be trained using a limited number of scans, which most centers are likely to have available, with notable generalization to external data. In addition, it provides voxel-wise maps of tumor type signatures that could be useful to visualize the CNN classification process and for tumor spatial characterization. As a way to make this tool more accessible and eventually make an impact in clinical practice, the proposed method has been implemented on the user-friendly DISCERN application, which is made freely accessible at http://84.88.64.102:5000/discern-app, to enhance study reproducibility and accelerate its adoption in future clinical studies.

Limitations of the study

The diagnostic tool DISCERN was trained with perfusion MRI data from scans without preload contrast injection, which may limit its performance on preloaded MRI scans. Tests on all eligible external 3T scans with a contrast preload of 23 patients with GBM from the IvyGAP[51] dataset yielded 18 cases correctly classified as GBM (0.78 accuracy). Further testing in preload MRI scans should be performed to explore the generalizability of the results in this context.

The study used automatically segmented regions of interest, revised by an experienced neuroradiologist. However, the variability in segmentations among different neuroradiologists has not been explored. The proposed future segmentation methods aim to minimize manual input, but this aspect requires further development.

The training data were obtained from only 40 patients, which may introduce a bias toward the inherent characteristics of this subpopulation. To account for patient and scanner biases, we normalized the signals to those of white matter and we trained with an equal number of patients for every malignancy. As a limitation of the retrospective nature of this study, older diagnostic

standards were in place and isocitrate dehydrogenase mutation status was missing from this study, which would provide a cleaner glioblastoma cohort.

The current algorithm is limited to certain MRI sequences (T1-weighted and perfusion MRI) and does not yet incorporate other potentially useful image data, such as diffusion MRI,[52] which may offer more detailed insights into tumor microstructure and potentially improve the performance of the tool.

DISCERN is user-friendly and can classify three common brain tumors, but its application to other tumor types is still under development. In addition, although the framework shows promise, it requires further clinical qualification and approval for use as a medical diagnostic tool.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cohort and clinical characteristics
- METHOD DETAILS
  - Patient eligibility
  - Imaging acquisition parameters
  - Data pre-processing
  - Conventional DSC-PWI metrics
  - CNN architecture
  - Classification scheme
  - Classification performance
  - CNN interpretation
  - Utility in aiding medical decisions for brain tumor diagnosis
  - Development of the online app DISCERN
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2024.101464.

REFERENCES

1. Young, R.M., Jamshidi, A., Davis, G., and Sherman, J.H. (2015). Current trends in the surgical management and treatment of adult glioblastoma. Ann. Transl. Med. *3*, 121.

2. Hatiboglu, M.A., Wildrick, D.M., and Sawaya, R. (2013). The role of surgical resection in patients with brain metastases. Ecancermedicalscience *7*, 308.

3. Hoang-Xuan, K., Bessell, E., Bromberg, J., Hottinger, A.F., Preusser, M., Rudà, R., Schlegel, U., Siegal, T., Soussain, C., Abacioglu, U., et al. (2015). Diagnosis and treatment of primary CNS lymphoma in immuno-competent patients: guidelines from the European Association for Neuro-Oncology. Lancet Oncol. *16*, e322–e332.

4. Dammers, R., Haitsma, I.K., Schouten, J.W., Kros, J.M., Avezaat, C.J.J., and Vincent, A.J.P.E. (2008). Safety and efficacy of frameless and frame-based intracranial biopsy techniques. Acta Neurochir. *150*, 23–29.

5. Chiavazza, C., Pellerino, A., Ferrio, F., Cistaro, A., Soffietti, R., and Rudà, R. (2018). Primary CNS Lymphomas: Challenges in Diagnosis and Monitoring. BioMed Res. Int. *2018*, 3606970.

6. Miller, K.D., Ostrom, Q.T., Kruchko, C., Patil, N., Tihan, T., Cioffi, G., Fuchs, H.E., Waite, K.A., Jemal, A., Siegel, R.L., and Barnholtz-Sloan, J.S. (2021). Brain and other central nervous system tumor statistics. CA A Cancer J. Clin. *71*, 381–406.

7. Leung, D., Han, X., Mikkelsen, T., and Nabors, L.B. (2014). Role of MRI in primary brain tumor evaluation. J. Natl. Compr. Cancer Netw. *12*, 1561–1568.

8. Arita, K., Miwa, M., Bohara, M., Moinuddin, F.M., Kamimura, K., and Yoshimoto, K. (2020). Precision of preoperative diagnosis in patients with brain tumor - A prospective study based on "top three list" of differential diagnosis for 1061 patients. Surg. Neurol. Int. *11*, 55.

9. Chakravorty, A., Steel, T., and Chaganti, J. (2015). Accuracy of percentage of signal intensity recovery and relative cerebral blood volume derived from dynamic susceptibility-weighted, contrast-enhanced MRI in the preoperative diagnosis of cerebral tumours. NeuroRadiol. J. *28*, 574–583.

10. Law, M., Cha, S., Knopp, E.A., Johnson, G., Arnett, J., and Litt, A.W. (2002). High-grade gliomas and solitary metastases: differentiation by using perfusion and proton spectroscopic MR imaging. Radiology *222*, 715–721.

11. Fordham, A.J., Hacherl, C.C., Patel, N., Jones, K., Myers, B., Abraham, M., and Gendreau, J. (2021). Differentiating Glioblastomas from Solitary Brain Metastases: An Update on the Current Literature of Advanced Imaging Modalities. Cancers *13*, 2960.

Cell Reports Medicine
Article

12. Artzi, M., Bressler, I., and Ben Bashat, D. (2019). Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. J. Magn. Reson. Imag. *50*, 519–528.

13. Bae, S., An, C., Ahn, S.S., Kim, H., Han, K., Kim, S.W., Park, J.E., Kim, H.S., and Lee, S.K. (2020). Robust performance of deep learning for distinguishing glioblastoma from single brain metastasis using radiomic features: model development and validation. Sci. Rep. *10*, 12110.

14. Qian, Z., Li, Y., Wang, Y., Li, L., Li, R., Wang, K., Li, S., Tang, K., Zhang, C., Fan, X., et al. (2019). Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers. Cancer Lett. *451*, 128–135.

15. Maurer, M.H., Synowitz, M., Badakshi, H., Lohkamp, L.N., Wüstefeld, J., Schäfer, M.L., and Wiener, E. (2013). Glioblastoma multiforme versus solitary supratentorial brain metastasis: differentiation based on morphology and magnetic resonance signal characteristics. Röfo *185*, 235–240.

16. Liu, D., Chen, J., Ge, H., Hu, X., Yang, K., Liu, Y., Hu, G., Luo, B., Yan, Z., Song, K., et al. (2022). Differentiation of malignant brain tumor types using intratumoral and peritumoral radiomic features. Front. Oncol. *12*, 848846.

17. Liu, Y., Xu, X., Yin, L., Zhang, X., Li, L., and Lu, H. (2017). Relationship between Glioblastoma Heterogeneity and Survival Time: An MR Imaging Texture Analysis. AJNR. Am. J. Neuroradiol. *38*, 1695–1701.

18. Lee, E.J., Ahn, K.J., Lee, E.K., Lee, Y.S., and Kim, D.B. (2013). Potential role of advanced MRI techniques for the peritumoural region in differentiating glioblastoma multiforme and solitary metastatic lesions. Clin. Radiol. *68*, e689–e697.

19. Yang, G., Jones, T.L., Barrick, T.R., and Howe, F.A. (2014). Discrimination between glioblastoma multiforme and solitary metastasis using morphological features derived from the p:q tensor decomposition of diffusion tensor imaging. NMR Biomed. *27*, 1103–1111.

20. Bauer, A.H., Erly, W., Moser, F.G., Maya, M., and Nael, K. (2015). Differentiation of solitary brain metastasis from glioblastoma multiforme: a predictive multiparametric approach using combined MR diffusion and perfusion. Neuroradiology *57*, 697–703.

21. Wang, S., Kim, S., Chawla, S., Wolf, R.L., Knipp, D.E., Vossough, A., O'Rourke, D.M., Judy, K.D., Poptani, H., and Melhem, E.R. (2011). Differentiation between glioblastomas, solitary brain metastases, and primary cerebral lymphomas using diffusion tensor and dynamic susceptibility contrast-enhanced MR imaging. AJNR. Am. J. Neuroradiol. *32*, 507–514.

22. Neska-Matuszewska, M., Bladowska, J., Sąsiadek, M., and Zimny, A. (2018). Differentiation of glioblastoma multiforme, metastases and primary central nervous system lymphomas using multiparametric perfusion and diffusion MR imaging of a tumor core and a peritumoral zone-Searching for a practical approach. PLoS One *13*, e0191341.

23. Swinburne, N.C., Schefflein, J., Sakai, Y., Oermann, E.K., Titano, J.J., Chen, I., Tadayon, S., Aggarwal, A., Doshi, A., and Nael, K. (2019). Machine learning for semi-automated classification of glioblastoma, brain metastasis and central nervous system lymphoma using magnetic resonance advanced imaging. Ann. Transl. Med. *7*, 232.

24. Priya, S., Ward, C., Locke, T., Soni, N., Maheshwarappa, R.P., Monga, V., Agarwal, A., and Bathla, G. (2021). Glioblastoma and primary central nervous system lymphoma: differentiation using MRI derived first-order texture analysis - a machine learning study. NeuroRadiol. J. *34*, 320–328.

25. Alcaide-Leon, P., Dufort, P., Geraldo, A.F., Alshafai, L., Maralani, P.J., Spears, J., and Bharatha, A. (2017). Differentiation of Enhancing Glioma and Primary Central Nervous System Lymphoma by Texture-Based Machine Learning. AJNR. Am. J. Neuroradiol. *38*, 1145–1150.

26. Ahn, S.J., Shin, H.J., Chang, J.H., and Lee, S.K. (2014). Differentiation between primary cerebral lymphoma and glioblastoma using the apparent diffusion coefficient: comparison of three different ROI methods. PLoS One *9*, e112948.

27. Toh, C.H., Wei, K.C., Chang, C.N., Ng, S.H., and Wong, H.F. (2013). Differentiation of primary central nervous system lymphomas and glioblastomas: comparisons of diagnostic performance of dynamic susceptibility contrast-enhanced perfusion MR imaging without and with contrast-leakage correction. AJNR. Am. J. Neuroradiol. *34*, 1145–1149.

28. Cha, S., Lupo, J.M., Chen, M.H., Lamborn, K.R., McDermott, M.W., Berger, M.S., Nelson, S.J., and Dillon, W.P. (2007). Differentiation of glioblastoma multiforme and single brain metastasis by peak height and percentage of signal intensity recovery derived from dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging. AJNR. Am. J. Neuroradiol. *28*, 1078–1084.

29. Lee, M.D., Baird, G.L., Bell, L.C., Quarles, C.C., and Boxerman, J.L. (2019). Utility of Percentage Signal Recovery and Baseline Signal in DSC-MRI Optimized for Relative CBV Measurement for Differentiating Glioblastoma, Lymphoma, Metastasis, and Meningioma. AJNR. Am. J. Neuroradiol. *40*, 1445–1450.

30. Barajas, R.F., Jr., Chang, J.S., Segal, M.R., Parsa, A.T., McDermott, M.W., Berger, M.S., and 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*, 486–496.

31. Hu, L.S., Baxter, L.C., Pinnaduwage, D.S., Paine, T.L., Karis, J.P., Feuerstein, B.G., Schmainda, K.M., Dueck, A.C., Debbins, J., Smith, K.A., et al. (2010). Optimized preload leakage-correction methods to improve the diagnostic accuracy of dynamic susceptibility-weighted contrast-enhanced perfusion MR imaging in posttreatment gliomas. AJNR. Am. J. Neuroradiol. *31*, 40–48.

32. Kim, Y.E., Choi, S.H., Lee, S.T., Kim, T.M., Park, C.K., Park, S.H., and Kim, I.H. (2017). Differentiation between Glioblastoma and Primary Central Nervous System Lymphoma Using Dynamic Susceptibility Contrast-Enhanced Perfusion MR Imaging: Comparison Study of the Manual versus Semiautomatic Segmentation Method. Investig. Magn. Reson. Imaging *21*, 9.

33. Mangla, R., Kolar, B., Zhu, T., Zhong, J., Almast, J., and Ekholm, S. (2011). Percentage signal recovery derived from MR dynamic susceptibility contrast imaging is useful to differentiate common enhancing malignant lesions of the brain. AJNR. Am. J. Neuroradiol. *32*, 1004–1010.

34. Park, J.E., Kim, H.S., Lee, J., Cheong, E.N., Shin, I., Ahn, S.S., and Shim, W.H. (2020). Deep-learned time-signal intensity pattern analysis using an autoencoder captures magnetic resonance perfusion heterogeneity for brain tumor differentiation. Sci. Rep. *10*, 21485.

35. Pons-Escoda, A., Garcia-Ruiz, A., Naval-Baudin, P., Grussu, F., Fernandez, J.J.S., Simo, A.C., Sarro, N.V., Fernandez-Coello, A., Bruna, J., Cos, M., et al. (2022). Voxel-level analysis of normalized DSC-PWI time-intensity curves: a potential generalizable approach and its proof of concept in discriminating glioblastoma and metastasis. Eur. Radiol. *32*, 3705–3715.

36. Surendra, K.L., Patwari, S., Agrawal, S., Chadaga, H., and Nagadi, A. (2020). Percentage signal intensity recovery: A step ahead of rCBV in DSC MR perfusion imaging for the differentiation of common neoplasms of brain. Indian J. Cancer *57*, 36–43.

37. Zhang, J., Liu, H., and Tong, H. (2017). Clinical Applications of Contrast-Enhanced Perfusion MRI Techniques in Gliomas: Recent Advances and Current Challenges. Contrast Media Mol. Imaging *2017*, 7064120.

38. Bell, L.C., Hu, L.S., Stokes, A.M., McGee, S.C., Baxter, L.C., and Quarles, C.C. (2017). Characterizing the Influence of Preload Dosing on Percent Signal Recovery (PSR) and Cerebral Blood Volume (CBV) Measurements in a Patient Population With High-Grade Glioma Using Dynamic Susceptibility Contrast MRI. Tomography *3*, 89–95.

39. Bell, L.C., Semmineh, N., An, H., Eldeniz, C., Wahl, R., Schmainda, K.M., Prah, M.A., Erickson, B.J., Korfiatis, P., Wu, C., et al. (2020). Evaluating the Use of rCBV as a Tumor Grade and Treatment Response Classifier Across NCI Quantitative Imaging Network Sites: Part II of the DSC-MRI Digital Reference Object (DRO) Challenge. Tomography *6*, 203–208.

40. Fu, R., Szidonya, L., Barajas, R.F., Jr., Ambady, P., Varallyay, C., and Neuwelt, E.A. (2022). Diagnostic performance of DSC perfusion MRI to

distinguish tumor progression and treatment-related changes: a systematic review and meta-analysis. Neurooncol. Adv. *4*, vdac027.

41. Boxerman, J.L., Paulson, E.S., Prah, M.A., and Schmainda, K.M. (2013). The effect of pulse sequence parameters and contrast agent dose on percentage signal recovery in DSC-MRI: implications for clinical applications. AJNR. Am. J. Neuroradiol. *34*, 1364–1369.

42. Paulson, E.S., and Schmainda, K.M. (2008). Comparison of dynamic susceptibility-weighted contrast-enhanced MR methods: recommendations for measuring relative cerebral blood volume in brain tumors. Radiology *249*, 601–613.

43. Boxerman, J.L., Quarles, C.C., Hu, L.S., Erickson, B.J., Gerstner, E.R., Smits, M., Kaufmann, T.J., Barboriak, D.P., Huang, R.H., Wick, W., et al. (2020). Consensus recommendations for a dynamic susceptibility contrast MRI protocol for use in high-grade gliomas. Neuro Oncol. *22*, 1262–1275.

44. Schmainda, K.M., Prah, M.A., Hu, L.S., Quarles, C.C., Semmineh, N., Rand, S.D., Connelly, J.M., Anderies, B., Zhou, Y., Liu, Y., et al. (2019). Moving Toward a Consensus DSC-MRI Protocol: Validation of a Low-Flip Angle Single-Dose Option as a Reference Standard for Brain Tumors. AJNR. Am. J. Neuroradiol. *40*, 626–633.

45. Open Science Initiative for Perfusion Imaging (OSIPI). (2023). OSIPI/osipi.github.io. GitHub. https://github.com/OSIPI.

46. Taricotti, L., Ferlito, D., Caccavella, V.M., Di Cristofori, A., Fiore, G., Remore, L.G., Giordano, M., Remoli, G., Bertani, G., Borsa, S., et al. (2022). A Deep Learning Model for Preoperative Differentiation of Glioblastoma, Brain Metastasis, and Primary Central Nervous System Lymphoma: An External Validation Study. NeuroSci *4*, 18–30.

47. Hess A, Meier R, Kaesmacher J Jung, S., Scalzo, F., Liebeskind, D., Wiest, R. and McKinley, R. Synthetic Perfusion Maps: Imaging Perfusion Deficits in DSC-MRI with Deep Learning. InBrainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4 2019 (pp. 447-455). Springer International Publishing.

48. Wang, H., Wang, Z., Du, M., Yang, F., Zhang, Z., Ding, S., Mardziel, P., and Hu, X. (2020). Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks. In Paper presented at: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

49. Pons-Escoda, A., Garcia-Ruiz, A., Naval-Baudin, P., Cos, M., Vidal, N., Plans, G., Bruna, J., Perez-Lopez, R., and Majos, C. (2020). Presurgical Identification of Primary Central Nervous System Lymphoma with Normalized Time-Intensity Curve: A Pilot Study of a New Method to Analyze DSC-PWI. AJNR. Am. J. Neuroradiol. *41*, 1816–1824.

50. Grussu, F., Blumberg, S.B., Battiston, M., Kakkar, L.S., Lin, H., Ianuş, A., Schneider, T., Singh, S., Bourne, R., Punwani, S., et al. (2021). Feasibility of Data-Driven, Model-Free Quantitative MRI Protocol Design: Application to Brain and Prostate Diffusion-Relaxation Imaging. Front. Physiol. *9*.

51. Shah, N., Feng, X., Lankerovich, M., Puchalski, R.B., and Keogh, B. (2016). Data from Ivy Glioblastoma Atlas Project (IvyGAP): The Cancer Imaging Archive.

52. Nilsson, M., Englund, E., Szczepankiewicz, F., van Westen, D., and Sundgren, P.C. (2018). Imaging brain tumour microstructure. Neuroimage *182*, 232–250.

53. Puchalski, R.B., Shah, N., Miller, J., Dalley, R., Nomura, S.R., Yoon, J.G., Smith, K.A., Lankerovich, M., Bertagnolli, D., Bickley, K., et al. (2018). An anatomic transcriptional atlas of human glioblastoma. Science *360*, 660–663.

54. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J. Digit. Imag. *26*, 1045–1057.

55. Data from the Multi-Institutional Paired Expert Segmentations and Radiomic Features of the Ivy GAP Dataset. The Cancer Imaging Archive (TCIA), (2020). https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70222827.

56. Pati, S., Verma, R., Akbari, H., Bilello, M., Hill, V.B., Sako, C., Correa, R., Beig, N., Venet, L., Thakur, S., et al. (2020). Reproducibility analysis of multi-institutional paired expert annotations and radiomic features of the Ivy Glioblastoma Atlas Project (Ivy GAP) dataset. Med. Phys. *47*, 6039–6052.

57. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn. Reson. Imaging *30*, 1323–1341.

58. Louis, D.N., Ohgaki, H., Wiestler, O.D., Cavenee, W.K., Burger, P.C., Jouvet, A., Scheithauer, B.W., and Kleihues, P. (2007). The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol. *114*, 97–109.

59. van Gelderen, P., Duyn, J.H., Ramsey, N.F., Liu, G., and Moonen, C.T.W. (2012). The PRESTO technique for fMRI. Neuroimage *62*, 676–681.

60. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., and Gee, J.C. (2010). N4ITK: improved N3 bias correction. IEEE Trans. Med. Imag. *29*, 1310–1320.

61. Pohl, K.M., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R.W., Kikinis, R., Grimson, W.E.L., Shenton, M.E., and Wells, W.M. (2007). A hierarchical algorithm for MR brain image parcellation. IEEE Trans. Med. Imag. *26*, 1201–1212.

Cell Reports Medicine
Article

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Ivy GAP imaging dataset	Shah et al.[51,53,54]	https://doi.org/10.7937/K9/TCIA.2016.XLwaN6nL
Segmentations for the Ivy GAP dataset	Pati et al.[55,56]	https://doi.org/10.7937/9j41-7d44
Trained models for DISCERN ("models" folder)	This paper; Zenodo; GitHub	https://doi.org/10.5281/zenodo.8301141 https://github.com/radiomicsgroup/discern-app
Software and algorithms		
Code for DISCERN	This paper; Zenodo; GitHub	https://doi.org/10.5281/zenodo.8301141 https://github.com/radiomicsgroup/discern-app
Dcm2niix	GitHub	https://github.com/rordenlab/dcm2niix
3D Slicer	Fedorov et al.[57]; website	https://doi.org/10.1016/j.mri.2012.05.001 https://slicer.org/
Python v3	Python Software Foundation	https://www.python.org

RESOURCE AVAILABILITY

Lead contact
All inquiries for further information regarding this work should be directed to and will be fulfilled by the lead contact, Raquel Perez-Lopez, 115–117 Natzaret, 08035 Barcelona, Spain (rperez@vhio.net).

Materials availability
No materials such as reagents or other products were generated in this study.

Data and code availability
- The datasets used in this study from Bellvitge Hospital, UC San Diego, and HT Medica health centers are not publicly available due to patient privacy concerns. To request access to the data, please contact the lead contact, who will connect you with the responsible researcher at each of the centers. Data will be accessible only if the Ethics Committee of each center where the data were collected grants permission. Therefore, the requester must describe the project for which data access is requested, detailing the objectives and data management plan. Data access will be considered for research purposes and non-commercial use only. To ensure patient privacy, access to personally identifiable information or sensitive clinical information (including medical histories) will not be provided, and requests for data access must rigorously adhere to the consent agreements established with study participants. Additional terms and conditions for accessing data by collaborating institutions may apply, as defined by the institutional Ethics Committee. The Ivy GAP dataset including MR DSC and T1w scans is publicly available and can be accessed through The Cancer Imaging Archive (https://doi.org/10.7937/K9/TCIA.2016.XLwaN6nL), as well as the annotated brain regions for the same dataset (https://doi.org/10.7937/9j41-7d44). The references and links for the datasets are listed in the key resources table.
- All the original codes and the trained models generated in this study are available under the folder "models" both in a repository at Zenodo (https://doi.org/10.5281/zenodo.8301141) and in a mirroring repository at GitHub (https://github.com/radiomicsgroup/discern-app), both publicly accessible as of the date of publication. The DOIs and links are listed in the key resources table. Instructions for use can be found on the GitHub readme file.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cohort and clinical characteristics
Data was retrospectively collected from multiple centers from patients with biopsy-confirmed GBM, brain metastasis or PCNSL. A total of 568 patients from three institutions (Bellvitge University Hospital, Spain; UC San Diego Health Center, USA; HT Medica Jaen, Spain) were included in the study.

Cell Reports Medicine
Article

Four hundred and forty patients (45 for PCNSL, 95 for metastasis and 300 for GBM) diagnosed from 2007 (2007 WHO criteria[58]) to 2020 at Bellvitge University Hospital (Spain) with MRI available were included in the development cohort after image quality inspection and exclusion (Figure 1A). Patient demographics and clinical characteristics per groups are shown in Table S1.

Additional independent cohorts were included and processed for external validation: a) 80 patients from UC San Diego Health Center (USA) and HT Medica Jaen (Spain), b) 25 patients from Bellvitge University Hospital (Spain), acquired at magnet strength of 3T and c) 23 patients from the IvyGAP[51] open database of patients with GBM and MRI scan with pre-bolus contrast administration, to account for the effect of different DSC protocols in the tool performance.

The research ethics committee of Bellvitge University Hospital (Barcelona, Spain) approved the study and informed consent was waived. The confidential data from patients were anonymized and protected in accordance with national and European regulations.

METHOD DETAILS

Patient eligibility

Eligibility criteria included: i) histologically confirmed diagnosis of GBM, single brain metastasis or PCNSL, ii) diagnostic MR scan on 1.5T or 3T including DSC and contrast-enhanced T1-weighted imaging (CE-T1WI) acquired prior to any oncological treatment and iii) a minimum of 10 mm of diameter of enhancing tumor in the CE-T1WI. Exclusion criteria included missing imaging data and poor image quality. The number of patients included can be seen in Figure 1A.

Imaging acquisition parameters

For the training and internal testing cohort, acquisitions were performed with 1.5T Philips scanners. The DSC dynamic sequences (gradient-echo echo-planar imaging and PRESTO[59]) salient parameters were: TE of 24–40 ms, and FA of 75° (PRESTO 7°), acquired during a bolus administration of gadolinium contrast agent (gadobutrol 0.1 mmol/kg) without contrast preload.

The MRI scans from the development set were performed in 1.5T Philips scanners (219 on Ingenia and 221 on Intera). The DSC-PWI dynamic sequences (gradient-echo echo-planar imaging and PRESTO[59]) salient parameters were: temporal sampling of 1.26–1.93 s, 40–60 timepoints, TR of 1640 ms (PRESTO 16 ms) TE of 24–40 ms, FA of 75° (PRESTO 7°) and in-plane resolution of 1.7 mm, acquired during a bolus administration of gadolinium contrast agent (gadobutrol 0.1 mmol/kg) without contrast preload.

The CE-T1WI main parameters were gradient-echo with repetition time (TR) of 6.58–25 ms, echo time (TE) of 3.17–5.49 ms, flip angle (FA) of 8°-30°, and spin-echo with TR of 535–625 ms, TE of 10–15 ms, FA of 70°-90°, all with in-plane resolution of 0.9 mm.

The acquisition details of the external datasets were as follows. The DSC-PWI protocol did not include contrast preload, the acquisition parameters were: TR of 1123–3031 ms, TE of 28–60 ms, FA of 40°-90° and in-plane resolution of 0.9 mm. Imaging was performed on Philips Achieva or Intera 1.5T (n = 55), Siemens Essenza (n = 4), Toshiba Vantage Elan 1.5T (n = 2), GE Signa HDxt 1.5T (n = 2), Philips Achieva 3T (n = 9), GE Signa Architect 3T (n = 5), GE Discovery 3T (n = 2) and Philips Panorama HFO 1T (n = 1).

Data pre-processing

The DSC dynamic sequence was motion-corrected by rigid registration of all DSC volumes. Bias field correction[60] was applied to the CE-T1WI, which was rigidly registered to the DSC. Brain region masking was obtained on the CE-T1WI using a hierarchical approach,[61] used for normal-appearing white matter segmentation and to split the brain into hemispheres. The white matter in the hemisphere contralateral to the tumor was selected once the tumor segmentation was available. Segmentations of the enhancing tumor were first obtained automatically by intensity thresholding on the image histogram. Specifically, an intensity threshold was set on the histogram right tail (voxels with high intensities) at the intensity where the histogram density falls below the mean histogram density. This result was followed by morphological operations to remove spurious voxels and small objects such as vessels. The segmentations were afterward revised by an experienced neuroradiologist.

The DSC signal reflecting the bolus passage in every voxel of the MRI sequence were extracted for the aforementioned enhancing tumor and white matter regions. According to a previously presented normalisation method,[23] the intensity and timescale of the DSC curves from the enhancing tumor were normalised (nDSC) to the white matter, interpolating to the same time resolution, providing comparable data among voxels of the same patient and between patients. The minimum peak point of the nDSC signals was retrieved, the curves were aligned to this point and signals with points within the average plus and minus standard deviation were used for training the CNNs, to avoid training on low signal or high noise data. Slicer[57] (www.slicer.org) and Python 3.8 were used for segmentation, processing, training, inference and statistical tests.

Conventional DSC-PWI metrics

Standard metrics of DSC analyses were obtained to compare against our voxel-trained CNN. The T2*-weighted signal S from a DSC-PWI sequence can be expressed as:

$$S \propto e^{-\frac{TE}{T2^*}} ; \frac{1}{T2^*} = R2^*$$

Cell Reports Medicine
Article

With R2* representing the relaxation rate. In a dynamic acquisition, a signal S(t) is acquired at n discrete time points t = [0,1,2, …,n] during the contrast bolus passage. The change in relaxation rate ΔR2*(t) is proportional to the contrast concentration C(t), and can be obtained as follows:

$$C_t \propto \Delta R2_t^* = R2_t^* - R2_0^* = -\frac{1}{TE} \ln \frac{S_t}{S_0}$$

The relative cerebral blood volume (rCBV) was obtained as the area under the ΔR2* curve with leakage correction using Slicer.[57] The percentage of signal recovery (PSR) was defined as the ratio of the recovered DSC-PWI signal over the baseline signal, as follows:

$$PSR = \frac{(S_r - S_{min})}{(S_b - S_{min})} \cdot 100$$

Where Smin is the minimum peak point reached by the perfusion curve, Sb is the baseline perfusion signal before contrast bolus arrival and Sr is the recovered signal after bolus passage.

CNN architecture

A 1D convolutional neural network (CNN) allows to take advantage of the convolution operation over the dynamic nDSC curves, considering the intensity change in a window of neighboring timepoints of size equal to the kernel size.

To take into account the information over different time windows, a CNN was designed with three 1D convolutional layers with kernel sizes [3,5,7]. A small size would mean looking at the closest neighbor timepoints and considering immediate changes in the dynamic signal, while bigger kernel sizes would consider further neighbor timepoints, therefore capturing the signal change over a longer time period. By including three possible kernel sizes, we expected to make the model more flexible toward heterogeneous nDSC input signals.

The convolutional layers were followed by a 10% dropout layer and max pooling layer (pool size 2). Because in the previous layers we extracted many features, these layers act as a filter to retain only relevant information: the dropout layer discards data from the previous layer and the max pooling layer down-samples the information into a smaller size. By reducing the number of features, these steps are meant to prevent overfitting and improve generalizability.

Finally, the output was concatenated into a dense layer with 100 nodes of rectified linear units and a final binary output layer with softmax activation, cross-entropy loss function and Adam optimizer. The internal parameters of the CNN were self-optimised during training and internal validation, and hyper-parameters were set experimentally as: batch size of 5000, maximum number of epochs of 3000 with a stopping condition of 50 iterations without validation loss improvement. The CNN was built using Tensorflow v2 with Keras frontend. The CNN classifier receives a given nDSC as input and outputs a binary probability.

Classification scheme

By applying the CNN classifier voxel-by-voxel, a probability map is obtained over the enhancing tumor region. These voxel-wise probabilities are then converted into a patient-wise classification as:

$$voxel\ proportion = \frac{n_{confident\ type\ 1}}{n_{confident\ type\ 1} + n_{confident\ type\ 2}} . \qquad \text{(Equation 1)}$$

Above, $n_{confident\ type\ 1}$ is the number of voxels with a probability higher than 0.9 for one tumor type and $n_{confident\ type\ 2}$ is the number of voxels with a probability higher than 0.9 for the second tumor type on the binary classifier. The tumor type of each patient was inferred by applying Youden's index (highest sum of specificity and sensitivity in the training set) to the voxel proportion above. In practice, a 3-class classifier differentiating between PCNSL, GMB and metastases was implemented by concatenating two 2-class classifiers. The first classifier distinguishes PCNSL from non-PCNSL cases, the least frequent of the three malignancies, while the second classifier differentiates the non-PCNSL cases into GBM or metastasis (Figure 1). This allows obtaining binary classification performance to compare with other binary methods and to apply ROC curves.

Classification performance

Classification accuracy, sensitivity and specificity were obtained for 100 groups of 25 randomly-selected patients from the test cohort in order to obtain average classification performance and 95% confidence intervals (CI) for all classifiers, as reported in Table S2.

In addition, the area under the ROC curve was obtained for binary classifications, which shows the trade-off between sensitivity and specificity of different classification thresholds (Figure 2B). The thresholds were set by Youden's index as described above, but the app allows users to change them to meet different clinical needs, as discussed. Average ROC curves were obtained for the 3-class problem (Figure 2C).

Classification metrics and ROC curves were obtained for PSR and rCBV applying the same classification structure used for the CNN-based approach.

To compare the proposed approach with the performance of different machine learning algorithms, we built Random Forests and Support Vector Machine models from the rCBV and PSR values, following the classification scheme described for the CNN, i.e., applying binary classifications consecutively. Training was performed on the 40 subjects of the training dataset and tested on the test set of 400 subjects. A Random Forests hyperparameter search was performed with 5-fold cross-validation and the best parameters applied were maximum depth of 1 and 500 trees, and an overall 3-way accuracy of 0.61. A Support Vector Machine classification model with radial kernel was also trained with 5-fold cross-validation and the best hyperparameters found were gamma = 1 and C = 10, with a three-way accuracy of 0.37 (Table S4).

CNN interpretation

The CNN provides a tumour-type probability value from each nDSC signal found in each voxel. The map inherently informs about the decision process of the CNN classifier toward one tumor type or another and, more importantly, about the confidence of the classification in spatial regions.

To further explore the features that the CNN associated with each tumor type, down to the individual timepoints of the nDSC curve, we applied a score-weighted visual explanation for CNNs (ScoreCAM[48]). On Figure 3A, the importance score was scaled to sum 1 over all timepoints in every voxel, in order to see the spatial relative importance. On Figure 3B, the average importance score is shown in each timepoint for each tumor type, with the average tumor nDSC from the training data overlaid in black, in order to see the temporal differences.

Utility in aiding medical decisions for brain tumor diagnosis

A benchmark diagnostic study was conducted to assess and compare the performance of DISCERN with that of expert neuroradiologists. The study also aimed to evaluate the potential improvement in diagnostic accuracy when DISCERN was utilised to support medical decision-making. For this purpose, two expert neuroradiologists with 10 and more than 20 years of experience in the field were selected. They were blinded to each other's assessments and also to the histopathological diagnosis results. The participants evaluated 30 cases, corresponding to 10 per each tumor type randomly distributed. The neuroradiologists had access to the MRI scans, given all imaging sequences (including T1w pre and post-contrast and the DSC maps), age and gender. To facilitate the evaluation process, each case was labeled by the neuroradiologists across four categories, ranging from "very unlikely" to "very likely," based on the likelihood and confidence of their diagnosis for each tumor type. A final diagnosis per radiologist was reached by considering the malignancy for which each radiologist estimated the highest likelihood, and then compared with the results from the tool. To measure the confidence in diagnosing, accounting for multiple answers for a given patient, we assigned scores (0.75 to "very likely", 0.25 to "likely" and "unlikely", 0 to "very unlikely") and applied:

$$Confidence = max\{A\} * \left(1 + max\{A\} - \sum_{a \in A} a \right)$$

Where 'A' is the set of answers 'a' for each patient. This function gives a maximum confidence score when assigning "very likely" to one malignancy and "very unlikely" to the others. The results of this subanalysis are shown in Figure S1.

Development of the online app DISCERN

The processing and classification pipeline was bundled into a Docker image which can run as a standalone application in any system (Figure 4). The user can input their anonymized DSC and CE-T1WI scans in raw DICOM, Nifti or NRRD formats and, optionally, their own segmentations. When the study is processed, the tool shows the average nDSC curve, the result of the classification and the spatial probability map. For demonstrative purposes, the DISCERN app can be accessed at http://84.88.64.102:5000/discern-app for research purposes.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical tests were performed to compare: (i) patient age distribution between training and test sets of each tumor type (Welch's t-test), as well as among tumor types (one-way ANOVA); (ii) patient gender distribution between training and test sets of each tumor type (Fisher's exact test), as well as among tumor types (Chi-square test). Relations are shown in Table S1. The DeLong test was used to compare differences in AUC between ROC curves. Statistical significance was considered for $p < 0.05$. Variable names and abbreviations are explained in each figure or table legend, numbers in tables are explained in the heading of each column.